ÉCOLE DOCTORALE SCIENCES ÉCONOMIQUES, JURIDIQUES, POLITIQUES ET DE GESTION
Université Clermont Auvergne

École Doctorale des Sciences Économiques, Juridiques, Politiques et de Gestion
Centre d'Études et de Recherche sur le Développement International (CERDI)

Université Clermont Auvergne, CNRS, IRD, CERDI, F-63000 Clermont-Ferrand, France

Three Essays on the Economics of Migration

Thèse présentée et soutenue publiquement le 11.09.2020

pour l'obtention du titre de Docteur en Sciences Économiques

par

Lucas Guichard

sous la direction de Simone Bertoli et Herbert Brücker

Membres du Jury

Michel Beine	Professeur, Université du Luxembourg	Rapporteur
Timothy Hatton	Professeur, University of Essex	Rapporteur
Panu Poutvaara	Professeur, Ludwig-Maximilians-Universität	Rapporteur
Ekrame Boubtane	Maître de conférences HDR, Université Clermont Auvergne	Suffragante
Simone Bertoli	Professeur, Université Clermont Auvergne	Directeur thèse
Herbert Brücker	Professeur, Humboldt-Universität	Directeur thèse

L'Université Clermont Auvergne et le CERDI n'entendent donner aucune approbation ou improbation aux opinions émises dans cette thèse. Ces opinions doivent être considérées comme propres à leur auteur.

To Christiane and Jacques Guichard.

"To be an immigrant, good or bad, is about straddling two homes, whilst knowing you don't really belong to either."

— Nikesh Shukla, The Good Immigrant (2016).

Acknowledgements

One page. Or two. Maybe I should write more than two. My acknowledgements will remain etched in these few lines I write, in my office, where it all began. I want you to know that these words eventually represent a tiny part of the gratitude I have for all people who have supported and followed me in this marathon. This dissertation is also yours and we are closing the finish line together with this manuscript.

I would first like to thank my two PhD supervisors, Simone Bertoli and Herbert Brücker. Your invaluable help and the precious advice have greatly contributed to the quality of my work. Your support, during times of high and low tides, has been one of the driving forces behind my first years of research. Beyond economic skills, I have learned with you how to become a researcher and the scientific rigour it requires. The excellence of your supervision has been an asset throughout my doctoral studies, and I am grateful to you for the time you have given me and for the patience you have shown.

I would like to express my sincere thanks to Michel Beine, Ekrame Boubtane, Timothy Hatton and Panu Poutvaara for having accepted to be part of the jury for my defense.

I would also like to thank all the researchers I met and exchanged with during my doctoral years. In particular, I have learned and improved many skills during the co-writing of the second chapter of my dissertation with Simone Bertoli and Jesús Fernández-Huertas Moraga. My research has also greatly benefited from discussion with my (former or current) colleagues in the department "Migration and International Labour Studies" of the IAB: Hanna Brenzel, Tanja Fendel, Teresa Freitas Monteiro, Lidwina Gundacker, Andreas Hauptmann, Philipp Jaschke, Sekou Keita, Yuliya Kosyakova, Steffen Sirries, Ignat Stepanok, Parvati Trübswetter, and Ehsan Vallizadeh.

My (almost) four years as PhD candidate were spent in an exceptional working environment, which largely contributed to the quality of my research. For this unique opportunity, I am grateful for the financial and logistical support I received as a member of the joint graduate programme (GrabAB) of the Institute for Employment Research (IAB) and the Faculty of Economics of the Friedrich-Alexander-Universität Erlangen-Nürnberg (FAU). I include to this

experience the other members of the programme, in particular Sabrina Genz and Sandra Huber, with whom I have often shared my daily life (and my questions) as a growing researcher. Finally, a word for the Université Clermont Auvergne and the research centre CERDI to which I have been affiliated for nine years and whose teachings have made my academic career possible.

I end up this part with what is certainly the most important paragraph, the one for which I have often (but not always) been said with a mischievous smile: "That way, Lucas, you can add me to the acknowledgements of your dissertation". So, as implicitly or explicitly promised, I offer you here my heartfelt gratitude. To my family, especially my parents, Pascale and Thierry, and my two sisters, Anne and Margaux, for their unfailing support. To my in-laws, Camille, Denis, Jean-François and Simon, for the moments shared in Puisaye and during my visits to Clermont-Ferrand. To my adoptive family in Nuremberg, Agathe, Anne-Sophie, Guillaume and Lucie, without whom weekends would have been much longer and boring. To my friends, in France, Germany and elsewhere, who have always found the right ways to follow me on my dissertation's journey. Finally, a special mention to Marie-Alice, *Lissou* for relatives, *Farfulle* for kin; the only partner who makes me who I am and contributes to making me better each and every day. To all of you, Thank You.

Nuremberg, September 8, 2020 Lucas Guichard

Remerciements

Une page. Ou deux. Peut-être devrais-je en écrire plus que deux. Mes remerciements resteront gravés dans ces quelques lignes que je rédige, à mon bureau, là où tout a commencé. Sachez que ces mots ne représentent finalement qu'une infime partie de la reconnaissance que j'ai pour toutes les personnes qui m'ont soutenu et accompagné dans ce marathon. Cette thèse est aussi la vôtre et nous clôturons la ligne d'arrivée ensemble avec ce manuscrit.

Je tiens d'abord à remercier mes deux directeurs de thèse, Simone Bertoli et Herbert Brücker. Votre aide précieuse et les nombreux conseils prodigués ont largement contribué à la qualité de mes travaux. Votre soutien, dans les bons et les mauvais moments, aura été un des moteurs de mes premières années de recherche. Au-delà des compétences économiques, j'ai appris avec vous le métier de chercheur et la rigueur scientifique qu'il requiert. L'excellence de votre encadrement aura été un atout tout au long de mon parcours de doctorant, et je suis reconnaissant du temps que vous m'avez accordé et de la patience dont vous avez fait preuve.

J'adresse des remerciements appuyés à Michel Beine, Ekrame Boubtane, Timothy Hatton et Panu Poutvaara pour avoir accepté de faire partie du jury pour ma défense.

Je remercie également l'ensemble des chercheurs que j'ai rencontré et avec qui j'ai échangé lors de mes années de doctorat. En particulier, j'ai appris et amélioré de nombreuses compétences durant la co-écriture du deuxième chapitre de ma thèse avec Simone Bertoli et Jesús Fernández-Huertas Moraga. Mes recherches ont aussi grandement bénéficié de multiples discussions avec mes (anciens ou actuels) collègues de travail au sein du département "Études sur les migrations et le travail international" de l'IAB : Hanna Brenzel, Tanja Fendel, Teresa Freitas Monteiro, Lidwina Gundacker, Andreas Hauptmann, Philipp Jaschke, Sekou Keita, Yuliya Kosyakova, Steffen Sirries, Ignat Stepanok, Parvati Trübswetter et Ehsan Vallizadeh.

Mes (presque) quatre années de thèse ont été réalisées dans un environnement de travail exceptionnel, qui a largement contribué à la qualité de mes recherches. Pour cette unique opportunité, je suis reconnaissant du soutien financier et logistique reçu en tant que membre du programme doctoral (GrabAB) de l'Institut de Recherche sur l'Emploi (IAB) en partenariat avec la Faculté d'Économie de la Friedrich-Alexander-Universität Erlangen-Nürnberg (FAU).

J'associe à cette expérience les autres membres du programme, notamment Sabrina Genz et Sandra Huber, avec qui j'ai souvent partagé mon quotidien (et mes interrogations) de chercheur en devenir. Enfin, un mot pour l'Université Clermont Auvergne et le laboratoire CERDI avec lesquels je suis affilié depuis maintenant neuf années et dont les enseignements ont rendu mon parcours académique possible.

Je conclus cette partie par ce qui est certainement le plus important des paragraphes, celui pour lequel on m'a souvent (mais pas toujours) dit avec un sourire malicieux : "Comme ça, Lucas, tu pourras m'ajouter dans les remerciements de ta thèse". Alors, comme promis implicitement ou explicitement, je vous offre ici ma plus grande reconnaissance. À ma famille, notamment mes parents, Pascale et Thierry, et mes deux sœurs, Anne et Margaux, pour leur soutien indéfectible. À ma belle-famille, Camille, Denis, Jean-François et Simon, pour les moments partagés en Puisaye et lors de mes visites à Clermont-Ferrand. À ma famille d'adoption à Nuremberg, Agathe, Anne-Sophie, Guillaume et Lucie, sans qui les week-ends auraient été bien plus longs et ennuyeux. À mes amis, en France, en Allemagne et ailleurs, qui ont toujours trouvé les bons moyens de m'accompagner dans le périple de ma thèse. Enfin, mention toute particulière à Marie-Alice, *Lissou* pour les proches, *Farfulle* pour les intimes ; l'unique conjointe qui fait de moi ce que je suis et qui contribue à me rendre meilleur de jour en jour. À vous tous, Merci.

Nuremberg, 8 septembre 2020 Lucas Guichard

Abstract

The questions around international migration, determined by economic and/or socio-political motives, regularly appear as an important and divisive topic in the political world, in public opinion or in the media. While attention towards this debate is relatively recent, the analysis of the causes and consequences of migration flows between countries has been addressed by economists and, more generally, by academia for many years. The three articles in this dissertation are in line with the economic literature on migration, while contributing to existing research on similar issues.

In the first chapter, I examine the pattern of selection on education of asylum seekers recently arrived in Germany from five key source countries: Afghanistan, Albania, Iraq, Serbia, and Syria. The analysis relies on original individual-level data collected in Germany combined with surveys conducted at origin. The results reveal a positive pattern of selection on education for asylum seekers who were able to flee Iraq and Syria, and the selection is neutral for individuals seeking asylum from Afghanistan and negative for asylum seekers from Albania and Serbia. I provide an interpretation of these patterns based on differences in the expected length of stay at destination, the migration costs faced by asylum seekers to reach Germany, and the size of migration networks at destination.

In the second chapter, we emphasize that acquiring information about destinations can be costly for migrants. We model information frictions in the rational inattention framework and obtain a closed-form expression for a migration gravity equation that we bring to the data. The model predicts that flows from countries with a higher cost of information or stronger priors are less responsive to variations in economic conditions in the various destinations, as migrants rationally get less information before deciding where to move. The econometric analysis reveals systematic heterogeneity in the pro-cyclical behaviour of migration flow across origins that is consistent with the existence of information frictions.

In the third chapter, I attempt to provide an answer to the following question: does the adoption of a list of safe countries of origin influence the asylum applications lodged in OECD member states? I draw on a structural gravity model to derive an empirical migration

equation that is brought to the data to estimate the direct effect of the list on the bilateral number of asylum claims. This, in turn, allows me to solve the structural model to quantify the externalities arising from a counter-factual experiment about the safe country policy. The empirical analysis reveals that the introduction of a list of safe source countries leads to a decrease of around 30% in the bilateral volume of asylum applications. The simulation exercise under an hypothetical change of the asylum policy suggests the presence of diversion effects on the sheer scale of asylum claims across *both* origin and destination countries.

Résumé

Les questions autour des migrations internationales, déterminées par des motifs économiques et/ou socio-politiques, apparaissent régulièrement comme un sujet important et clivant à la fois dans le monde politique, dans l'opinion publique ou dans les médias. Alors que l'attention sur ce débat est relativement récente, l'analyse des causes et des conséquences des flux migratoires entre pays est abordée par les économistes et, plus généralement, par le monde universitaire depuis de nombreuses années. Les trois articles de cette thèse s'inscrivent dans la continuité de la littérature économique sur les migrations, tout en contribuant aux recherches déjà existantes sur des problématiques similaires.

Dans le premier chapitre, j'examine le schéma de sélection en matière d'éducation des demandeurs d'asile récemment arrivés en Allemagne en provenance de cinq pays d'origine clés : Afghanistan, Albanie, Irak, Serbie et Syrie. L'analyse repose sur des données individuelles uniques collectées en Allemagne, combinées à des enquêtes menées dans les pays d'origine. Les résultats révèlent une sélection positive en matière d'éducation des demandeurs d'asile qui ont pu fuir l'Irak et la Syrie, la sélection est neutre pour les personnes demandant l'asile en provenance d'Afghanistan et est négative pour les demandeurs d'asile venant d'Albanie et de Serbie. Je propose une interprétation de ces résultats fondée sur les différences de durée de séjour attendue à destination, les coûts de migration auxquels sont confrontés les demandeurs d'asile pour atteindre l'Allemagne et la taille des réseaux de migration à destination.

Dans le deuxième chapitre, nous soulignons que l'acquisition d'information sur les pays de destination peut être coûteuse pour les migrants. Nous modélisons les éléments de friction relatifs à l'information dans le modèle d'inattention rationnelle et nous dérivons une solution analytique d'une équation de gravité pour les migrations que nous évaluons avec des données. Le modèle prédit que les flux en provenance de pays où le coût de l'information est plus élevé, i.e. où les a priori sont plus importants, sont moins sensibles aux variations des conditions économiques dans les différentes destinations, car les migrants obtiennent rationnellement moins d'informations avant de décider où aller. L'analyse économétrique révèle une hétérogénéité systématique dans le comportement pro-cyclique des flux migratoires entre les origines,

ce qui est cohérent avec l'existence de frictions en matière d'information.

Dans le troisième chapitre, je tente de répondre à la question suivante : est-ce que l'adoption d'une liste de pays d'origine sûrs influence les demandes d'asile déposées dans les États membres de l'OCDE ? Je m'appuie sur un modèle de gravité structurelle pour dériver une équation empirique de migration qui est évaluée avec des données pour estimer l'effet direct de la liste sur le nombre bilatéral de demandes d'asile. Cela me permet ensuite de résoudre le modèle structurel pour quantifier les externalités provenant d'une expérience contrefactuelle sur la politique des pays sûrs. L'analyse empirique révèle que l'introduction d'une liste de pays d'origine sûrs entraîne une diminution d'environ 30% du nombre de demandes d'asile entre pays. L'exercice de simulation basé sur un changement hypothétique de la politique d'asile suggère la présence d'effets de diversion sur le volume de demandes d'asile entre *à la fois* les pays d'origine et de destination.

Contents

List of Figures

List of Tables

Introduction

"Tout homme a le droit de partir, c'est son pays qui doit le persuader de rester - quoi qu'en disent les politiques grandiloquents. Ne te demande pas ce que ton pays peut faire pour toi, demande-toi ce que tu peux faire pour ton pays. Facile à dire quand tu es milliardaire, et que tu viens d'être élu, à 43 ans, président des États-Unis d'Amérique ! Mais lorsque, dans ton pays, tu ne peux ni travailler, ni te soigner, ni te loger, ni t'instruire, ni voter librement, ni exprimer ton opinion, ni même circuler dans les rues à ta guise, que vaut l'adage de John F. Kennedy ?"

Amin Maalouf, *Les Désorientés* (2012).

Depuis les temps les plus anciens, l'humanité est en mouvement. Cette pensée s'avère être particulièrement appropriée à la période contemporaine. Aujourd'hui, les frontières nationales apparaissent comme étant de plus en plus poreuses. En 2019, le nombre de personnes dans le monde vivant dans un pays autre que celui dans lequel elles sont nées a atteint 270 millions, soit 50 millions de plus qu'en 2010.[1] Les migrants internationaux représentent ainsi 3.5% de la population mondiale, en comparaison avec 2.8% en 2000, attestant que la proportion de migrants dans la population mondiale a également augmenté. Si de nombreux individus émigrent par choix, beaucoup d'autres émigrent par nécessité. Le nombre de personnes déplacées de force dans le monde a dépassé les 70 millions à la fin de 2018, pour la première fois en presque 70 ans d'histoire de l'Agence des Nations Unies pour les réfugiés (UNHCR). Ce nombre inclut 26 millions de réfugiés, 3.5 millions de demandeurs d'asile et plus de 41 millions de personnes déplacées à l'intérieur de leur propre pays.

L'évolution longue et les plus abrupts changements, souvent qualifiés de crises migratoires par les médias et le monde politique, font que les sujets autour des migrations internationales sont plus que jamais d'actualité. Dans le même temps, le nombre de questions autour des problématiques migratoires ne cesse d'augmenter. Le bref état des lieux réalisé ci-dessus oblige déjà à un premier questionnement lié aux causes des déplacements de population :

[1] Toutes les statistiques données dans cette introduction proviennent des Nations Unies.

pourquoi les individus migrent-ils ? Quels sont les déterminants des migrations de personnes entre pays dans le monde ? Cette introduction commence par décrire les principaux éléments de réponse apportés par la littérature sur ces points fondamentaux, ainsi que les contributions du *chapitre II* de cette thèse concernant ce thème.

Les nombres agrégés, comme ceux cités dans le premier paragraphe, cachent en réalité la grande hétérogénéité qui caractérise les populations de migrants. Tout d'abord, les individus n'ont pas tous le même pays d'origine et ils ne se déplacent pas tous au même endroit. En 2019, un tiers des migrants internationaux vient de seulement dix pays, l'Inde étant le principal pays d'origine avec 18 millions de personnes vivant à l'étranger. Les migrants en provenance du Mexique constituaient la deuxième plus grande diaspora (12 millions), suivi par la Chine (11 millions) et la Russie (10 millions). La plupart des migrants internationaux vont dans des pays situés dans la même région du monde. Ainsi, environ 31% de l'ensemble des migrants internationaux résident en Asie, 30% en Europe, 26% sur les continents américains, 10% en Afrique et 3% en Océanie.

Au-delà du pays d'origine et du choix de la destination, les groupes de migrants diffèrent en termes de composition démographique. Les femmes représentent un peu moins de la moitié du total des migrants, avec une proportion dans le nombre de migrants internationaux qui a légèrement diminué passant de 49% en 2000 à 48% en 2019. Au niveau de l'âge, 14% de la population mondiale de migrants a moins de 20 ans et 75% des individus sont en âge de travailler (20 à 64 ans). Ces dernières statistiques conduisent à la seconde série de questions : qui sont les migrants ? Quelles sont les caractéristiques individuelles des personnes qui décident de quitter leur pays de naissance pour s'installer à l'étranger ? La seconde partie de cette introduction présente les avancées fondamentales de la littérature sur ce sujet, en se concentrant sur le niveau d'éducation des migrants, tout en explicitant les contributions du *chapitre I* de la thèse dans ce domaine.

Enfin, les flux migratoires ne sont pas sans conséquences, à la fois pour les pays d'origine et les pays d'accueil des migrants. Un large segment de la littérature économique relative aux migrations internationales analyse les effets sur le marché du travail dans les pays d'origine et de destination, à la suite des changements de l'offre induits par l'immigration, en se focalisant notamment sur l'ajustement des salaires et de l'emploi. Les recherches en la matière convergent généralement vers deux axes principaux, à savoir expliquer comment les migrants s'adaptent à leur nouvel environnement (on parle ici souvent d'assimilation ou d'intégration des immigrés) et déterminer l'amplitude des réponses du marché du travail aux évolutions migratoires entre pays (on cherche souvent à évaluer les effets sur l'emploi et les

salaires des natifs). Ces deux thématiques ne sont pas mutuellement exclusives et l'adaptation des migrants à leur nouvelle situation influence généralement les effets observés dans les différents pays.

Les anticipations sur l'impact des flux migratoires et/ou les évènements post-migration s'accompagnent régulièrement de réactions dans les pays concernés, qui se traduisent en particulier par des modifications en termes de politique migratoire. Les décisions politiques en matière de migration varient dans le temps et en fonction du contexte dans lequel se trouvent les pays, en lien avec les informations disponibles sur les causes des migrations et sur les caractéristiques des migrants. Elles peuvent avoir pour but, par exemple, d'attirer une catégorie spécifique de migrants (généralement les individus qualifiés) ou bien chercher à réduire l'immigration totale dans une destination donnée. La dernière section de cette introduction propose un aperçu global des analyses sur les effets associés aux changements de politique migratoire. C'est dans ce cadre que s'insère le *chapitre III* de la présente thèse, pour lequel sera explicité les différentes contributions aux travaux déjà existants dans la littérature.

Les flux migratoires et leurs déterminants[2]

Une myriade de forces économiques et non économiques sont à l'origine de la décision d'émigrer. Les migrants peuvent être "poussés" hors de leur pays d'origine en raison de la détérioration des conditions économiques ou de troubles socio-politiques. À l'inverse, les migrants sont souvent "attirés" vers des destinations qui offrent des salaires élevés, de bonnes infrastructures ou des services de qualité.

Cadre théorique

Les études qui visent à analyser les causes des flux de personnes entre pays utilisent le modèle de gravité, qui a d'abord fait ses preuves dans la littérature sur le commerce international, avant de s'imposer récemment pour décrire les migrations internationales. Cette dynamique est associée à la plus grande disponibilité des données sur la migration, en particulier de nature dyadique (i.e., associées à des paires origine-destination). La littérature se base sur les modèles de maximisation aléatoire de l'utilité (RUM) qui décrivent le problème de décision de localisation auquel les individus font face pour obtenir la valeur attendue de la part des personnes résidant dans un pays d'origine j qui se déplacent dans une destination k à un moment t.

[2] Cette section est largement inspirée du guide publié par Beine *et al.* (2016).

Le modèle canonique RUM de migration décrit l'utilité qu'un individu qui était situé dans le pays j au temps $(t-1)$ dérive du choix du pays k appartenant à l'ensemble des opportunités disponibles pour cet individu au temps t. Cette utilité est divisée en plusieurs parties : une composante déterministique et des coûts migratoires entre j et k qui varient dans le temps, les deux pouvant être modélisés comme une fonction de variables qui sont observables par les chercheurs, et un terme stochastique spécifique à chaque individu et inobservable. Les hypothèses sur la distribution de ce terme (i.e., distribution indépendante et identiquement distribuée des valeurs extrêmes de type 1, McFadden (1974)) déterminent la probabilité attendue que la désignation du pays k représente le choix de l'individu.

Cette définition permet alors de dériver une expression pour les flux migratoires (en niveau) attendus entre deux pays. Ces derniers dépendent (de manière multiplicative) de la capacité du pays d'origine j à envoyer des migrants, de l'attractivité de la destination k, de l'accessibilité de la destination k pour les migrants potentiels en provenance de j, et les flux sont inversement liés à la valeur exponentielle de l'utilité attendue par les migrants potentiels au regard des alternatives disponibles (Small et Rosen, 1981).

Un élément clé de ce modèle a été mis en avant par Bertoli et Fernández-Huertas Moraga (2013), désigné sous le nom de résistance multilatérale à la migration et défini comme l'influence confondante que l'attrait des destinations alternatives (au pays k) exerce sur le taux de migration bilatérale. Celle-ci peut provenir d'hypothèses plus générales (en comparaison avec la distribution adoptée dans le modèle standard de gravité) sur la distribution de la composante stochastique de l'utilité ou de la prise en compte explicite de la nature séquentielle des décisions migratoires. Ignorer cette influence génère des biais dans l'estimation des coefficients des déterminants de la migration. Par exemple, tant Bertoli et Fernández-Huertas Moraga (2013) que Bertoli *et al.* (2013) constatent que l'effet des conditions économiques dans le pays d'origine sur les taux de migration est sur-estimé lorsque l'influence des autres destinations est mise de côté. La possibilité de biais importants est encore plus prononcée lors de l'étude des effets associés aux politiques migratoires. Étant donné que ces politiques ont tendance à être coordonnées entre les pays de destination, plusieurs études trouvent des effets plus importants par rapport aux travaux qui ne prennent pas en compte la résistance multilatérale à la migration (Bertoli et Fernández-Huertas Moraga, 2013, 2015).

Le modèle RUM de migration est silencieux sur la dimension temporelle du problème de décision de localisation auquel sont confrontés les migrants potentiels. L'inclusion d'un indice de temps t suggère que les individus font des choix de localisation répétés au cours de leur vie. Par exemple, un individu qui a décidé de migrer au moment t pourrait décider dans

une période suivante de retourner dans son pays d'origine ou de se rendre dans une autre destination. Dès lors, l'utilité spécifique au lieu de résidence peut être écrite de manière à refléter explicitement la nature séquentielle du problème de décision de localisation, comme proposé par Bertoli *et al.* (2016a).

Résultats empiriques[3]

L'attractivité d'un pays de destination k pour les migrants potentiels d'origine j et les coûts de migration entre j et k sont généralement modélisés comme des fonctions linéaires de deux vecteurs de variables, qui peuvent varier selon toutes les combinaisons des dimensions de l'origine, de la destination et du temps. La revue des travaux empiriques réalisée ci-dessous permet de lever le voile sur certaines conclusions présentes dans la littérature au sujet des facteurs expliquant les flux (et les taux) de migration internationale. Ces dernières sont obtenues à la suite de l'estimation d'équations de gravité avec des données dyadiques et fondées théoriquement sur le modèle RUM de migration.

Le niveau de revenu par habitant est un élément déterminant de l'attractivité de chaque lieu. Un modèle (RUM) de migration n'impose aucune contrainte sur la forme fonctionnelle que doit prendre la relation entre le revenu par habitant et la composante déterministe de l'utilité spécifique à chaque localité. Grogger et Hanson (2011) favorisent une spécification où l'attractivité dépend linéairement du revenu par habitant, tandis que d'autres articles dans la littérature optent pour une spécification de type logarithmique (Mayda, 2010; Bertoli *et al.*, 2013; Bertoli et Fernández-Huertas Moraga, 2013; Ortega et Peri, 2013; McKenzie *et al.*, 2014). La littérature suppose généralement que les perspectives de revenus des migrants potentiels de toutes origines peuvent être mesurées par le PIB par habitant à destination, imposant ainsi principalement l'hypothèse d'une tendance commune des revenus des migrants à destination, Bertoli et Fernández-Huertas Moraga (2013) représentant une exception à cet égard, et minimisant également les préoccupations concernant la causalité inverse. Des améliorations ont été proposées par Grogger et Hanson (2011), qui appliquent des barèmes d'imposition sur les revenus propres à chaque pays pour obtenir des mesures de revenus après impôt, par Grogger et Hanson (2011) et Belot et Hatton (2012), qui récupèrent les revenus spécifiques à l'éducation, et par Beine *et al.* (2019), qui se concentrent sur les salaires plutôt que sur les revenus. Les résultats indiquent une relation positive robuste entre le revenu par habitant et l'attractivité d'une destination.

[3] Cette partie élude intentionnellement les résultats concernant les politiques migratoires, qui seront abordés dans la troisième section de cette introduction.

Le modèle RUM avec des hypothèses de distribution à la McFadden (1974) implique qu'une variation simultanée et identique du (logarithme du) revenu par habitant à l'origine et à la destination n'influence pas le taux de migration bilatérale. Une telle symétrie parfaite disparaît si l'on considère que les migrants potentiels peuvent être confrontés à des contraintes de liquidité qui entravent leurs choix de localisation. Les exigences de crédit peuvent être prises en compte dans le modèle en supposant que les coûts migratoires sont négativement corrélés avec le revenu à l'origine. Si la dépendance des coûts de la migration bilatérale par rapport aux conditions économiques à l'origine n'est pas correctement contrôlée, une augmentation des revenus à l'origine réduirait le taux de migration bilatérale de façon moins importante qu'une diminution identique à la destination, et elle pourrait même accroître l'ampleur des flux migratoires bilatéraux. Le rôle des contraintes de liquidité a donc été pris en compte, par l'inclusion de conditions de revenu d'ordre supérieur dans le pays d'origine (Vogler et Rotte, 2000; Pedersen *et al.*, 2008; Mayda, 2010), en contrôlant pour l'incidence de la pauvreté à l'origine (Belot et Hatton, 2012) ou en divisant l'échantillon en fonction du revenu des pays d'origine (Ortega et Peri, 2013). Les preuves économétriques fournies par Vogler et Rotte (2000), Pedersen *et al.* (2008), et Mayda (2010) suggèrent que les contraintes de crédit entravent les flux migratoires internationaux observés, brouillant l'effet du revenu si elles ne sont pas correctement prises en compte dans l'analyse (Belot et Hatton, 2012).

Un modèle séquentiel de migration implique que le taux de migration bilatérale dépend des attentes concernant l'évolution des conditions économiques dans tous les pays appartenant à l'ensemble des alternatives possibles pour chaque migrant potentiel. Bertoli *et al.* (2016a) montrent empiriquement le rôle très important de ce facteur dans la stimulation des flux migratoires bilatéraux vers l'Allemagne entre 2006 et 2012.

Un autre effet est celui des facteurs environnementaux, et des facteurs climatiques en particulier, sur les migrations internationales. Quatre canaux par lesquels ces déterminants influencent l'émigration sont principalement examinés dans la littérature. Premièrement, les chocs climatiques négatifs diminuent les revenus dans le pays d'origine, agissant sur son attractivité, au travers d'une baisse des salaires ou d'une augmentation du taux d'emploi. Deuxièmement, les chocs peuvent augmenter les coûts de migration s'ils détruisent des actifs, rendant ainsi les exigences de crédit plus contraignantes. Troisièmement, les chocs climatiques néfastes ont tendance à diminuer l'attrait du pays d'origine indépendamment des revenus (par exemple, en raison d'une augmentation de la morbidité), ce qui entraîne à nouveau des incitations à émigrer. Quatrièmement, le canal de la volatilité : des conditions climatiques qui deviennent plus volatiles peuvent conduire les personnes peu enclines au risque à opter pour

la migration. Beine et Parsons (2015) testent ces mécanismes de transmission des effets dans un modèle de gravité des migrations. Les résultats évoquent des éléments robustes en faveur du canal du marché du travail dans le contexte des flux migratoires des pays du Sud vers les pays du Nord.

En outre, il existe des facteurs dyadiques qui influencent les coûts de la migration. Les composantes les plus importantes, qui ne varient pas dans le temps, sont la distance (physique) bilatérale, les liens coloniaux et la proximité linguistique et culturelle entre deux pays. L'effet des liens coloniaux passe principalement par les réseaux de migrants, alors que la proximité linguistique exerce une influence supplémentaire au-delà de celle passant par les groupes de migrants. La plupart des analyses intègrent la langue dans les estimations, soit par l'utilisation de variables traduisant l'existence d'une langue commune (officielle ou parlée) entre deux pays, soit par quelques mesures simples de la proximité linguistique. Des indicateurs plus élaborés ont néanmoins été créés : Belot et Ederveen (2012) et Adserà et Pytliková (2015) emploient diverses mesures de proximité, basées sur des arbres généalogiques établis par des linguistes ou sur des mesures de similarité phonétique entre les langues. Belot et Ederveen (2012) ont également recours à des mesures de proximité culturelle décrivant la distance religieuse entre deux pays et des mesures basées sur des enquêtes qui traduisent l'orientation culturelle des pays, toutes deux favorisant les flux de migrants internationaux.

Une littérature abondante a été consacrée au rôle des réseaux de migration sur l'amplitude et la structure des flux migratoires bilatéraux. Le rôle des réseaux a été analysé à l'aide du modèle de gravité des migrations. Néanmoins, plusieurs défis économétriques rendent cet exercice périlleux pour estimer correctement l'effet de ce facteur. Des travaux se sont focalisés sur des modèles de gravité structurels (Beine *et al.*, 2011; Bertoli et Fernández-Huertas Moraga, 2015), plus adaptés aux enjeux empiriques, et ils aboutissent à des résultats assez consensuels : une augmentation de 10% du stock bilatéral de migrants entraîne une augmentation de 4% du flux de migrants entre deux pays.

Chapitre II et ses contributions

Dans le chapitre II, nous soulignons que l'acquisition d'information sur les pays de destination peut être coûteuse pour les migrants.[4] Nous modélisons les éléments de friction relatifs à l'information dans le modèle d'inattention rationnelle et nous dérivons une solution analytique

[4] Cette hypothèse est à mettre en parallèle avec celle faite dans plusieurs contributions fondamentales sur la modélisation des déterminants des choix migratoires, qui supposent que l'incertitude est entièrement (et sans coût) résolue avant de décider où migrer.

d'une équation de gravité pour les migrations que nous évaluons avec des données pour un grand nombre de pays d'origine et de destination. Le modèle prédit que les flux en provenance de pays où le coût de l'information est plus élevé, i.e. où les a priori sont plus importants, sont moins sensibles aux variations des conditions économiques dans les différentes destinations, car les migrants obtiennent rationnellement moins d'informations avant de décider où aller. L'analyse économétrique révèle une hétérogénéité systématique dans le comportement procyclique des flux migratoires entre les origines, ce qui est cohérent avec l'existence de frictions en matière d'information.

Par rapport à la littérature, l'article offre plusieurs contributions à la fois théoriques et empiriques. Une première série d'innovations est associée aux travaux sur les modèles d'inattention rationnelle. La présente introduction ne portant pas sur ce sujet, je ne détaille pas les avancées réalisées dans ce domaine. La principale conclusion est que nous fournissons des preuves de la pertinence empirique de l'inattention rationnelle dans des situations de choix discrets, complétant ainsi une littérature qui est encore essentiellement théorique. Les migrants semblent donc être rationnellement inattentifs même si les enjeux liés à leurs décisions de localisation sont certainement très élevés (voir, par exemple, McKenzie *et al.* (2010) et Clemens *et al.* (2019)).

D'autre part, notre article est le premier à évaluer, avec des données, une équation de gravité des migrations obtenue à partir d'un modèle contenant des frictions de l'information, Porcher (2019) étant le seul autre article dont nous ayons connaissance, dans son cas, exploitant les flux migratoires internes au Brésil. De plus, nous apportons deux contributions principales à la littérature. Premièrement, nous démontrons qu'une micro-fondation alternative de l'équation de gravité des migrations permet de découvrir et d'interpréter les hétérogénéités systématiques entre les pays d'origine dans la réactivité des flux migratoires par rapport aux conditions économiques changeantes des différents pays de destination. Deuxièmement, notre analyse implique une raison supplémentaire pour laquelle les flux migratoires ont un caractère inertiel, en plus de celle apportée par les externalités positives générées par les réseaux migratoires spécifiques à chaque destination (par exemple, Munshi (2003)), car les frictions au niveau de l'information induisent une distribution plus concentrée des migrants entre les destinations.[5]

[5] Nos résultats révèlent également une dimension supplémentaire d'interdépendance entre les flux migratoires dirigés vers différents pays, au-delà des interactions stratégiques dans les politiques migratoires (Giordani et Ruta, 2013).

Sélection des migrants et des demandeurs d'asile/réfugiés

Les conséquences de l'immigration dans les pays d'origine et de destination dépendent des caractéristiques individuelles des migrants et des différences existantes entre les personnes qui migrent et celles qui restent (les natifs). Les immigrés ne constituent pas un échantillon aléatoire de la population des pays d'origine et la composition des groupes de migrants dépend entièrement de la nature du procédé de sélection qui distingue les individus qui se déplacent de ceux qui ne migrent pas.

Fondations théoriques

La sélection des migrants est un sujet étudié depuis plusieurs décennies dans la littérature. Les analyses principales sont basées sur les théories qui définissent la décision de migrer en fonction des coûts et des bénéfices associés à un changement de pays (Sjaastad, 1962). En partant de l'idée que les caractéristiques observables (Borjas, 1987) et inobservables (Borjas, 1991) des individus influencent les bénéfices obtenus grâce à la migration, Borjas établit un modèle fondateur qui se base sur le cadre proposé par Roy (1951) pour expliquer les choix professionnels. Cette extension permet de déterminer théoriquement quelles sont les personnes qui sont les plus enclines à émigrer.

De manière plus précise, le modèle de Roy-Borjas prévoit que les migrants seront négativement sélectionnés si le pays d'origine offre de meilleurs rendements en rapport aux compétences d'un individu, et donc des niveaux d'inégalités plus élevés, que le pays de destination. Si, au contraire, le pays d'origine a un faible rendement des compétences, donc des inégalités de revenus moindres que le pays d'accueil, alors les migrants seront tirés de manière disproportionnée de la partie supérieure de la distribution des compétences du pays d'origine. Un dernier cas est celui de la sélection des réfugiés, qui peut apparaître lorsque les migrants ont des revenus supérieurs à la moyenne (par rapport à l'ensemble de la population d'origine) dans le pays de destination. Ce schéma peut se produire pour un groupe minoritaire hautement qualifié dont les perspectives ont été déprimées par des préjugés (par exemple, les Juifs européens) ou pour des intellectuels originaires de pays qui ont connu une prise de pouvoir communiste.

Plus récemment, Grogger et Hanson (2011) ont abordé un point théorique important, en suggérant que la migration peut s'expliquer par les différences de salaires absolues plutôt que relatives. Les prédictions de leur modèle sur la sélection des individus indiquent que plus la différence absolue de revenus liée aux compétences entre le pays de destination et le

pays d'origine est grande, plus les migrants associés à cette paire de pays seront éduqués en comparaison avec les non-migrants. Au niveau du choix du pays d'accueil, le modèle prédit que le stock relatif de migrants plus instruits dans une destination augmente avec la différence absolue de revenus entre les travailleurs hautement et peu qualifiés.

Enfin, le modèle Roy-Borjas a fait l'objet de modifications pour s'adapter à la réalité des migrations forcées, c'est-à-dire les déplacements concernant les réfugiés et/ou les demandeurs d'asile qui fuient leur pays d'origine pour des motifs majoritairement non-économiques. Aksoy et Poutvaara (2020) montrent que la sélection des migrants par rapport à leur capital humain dépend non seulement des rendements attribuables à ce capital (qui peuvent varier en fonction du genre de chaque individu), mais aussi des risques associés à un conflit ou à des persécutions dans le pays d'origine et des risques liés à une potentielle migration. Les implications sont les suivantes : si les rendements sont plus élevés dans le pays d'origine et que ce dernier est relativement sûr, alors les migrants seront négativement sélectionnés, en lien avec Borjas (1987). Cependant, si le pays d'origine est confronté à un conflit suffisamment grave, alors la sélection est inversée et les migrants auront tendance à venir de la partie supérieure de la distribution des compétences. Si le rendement du capital humain en fonction du genre est plus faible dans le pays d'origine que le rendement ajusté au risque dans la destination potentielle, les migrants seront positivement sélectionnés même en l'absence de risques dus au fait de rester dans le pays d'origine.

Applications empiriques

La littérature empirique sur la sélection des migrants est vaste et les études existantes offrent des résultats qui varient en fonction de certaines hypothèses au niveau du modèle Roy-Borjas, en fonction des pays analysés ou des données utilisées. Plusieurs travaux ont été publiés sur l'évaluation du processus de sélection dans le cas de flux migratoires spécifiques, notamment les flux entre le Mexique et les États-Unis. Dans un modèle faisant l'hypothèse que les coûts migratoires diminuent avec les compétences, Chiquiar et Hanson (2005) apportent des éléments en faveur d'une sélection intermédiaire (positive) pour les hommes (femmes) qui migrent aux États-Unis en provenance du Mexique. McKenzie et Rapoport (2010) étendent le précédent modèle en intégrant les effets des réseaux de migrants. Ils parviennent à des conclusions similaires, puisque la sélection positive (négative) survient lorsque la taille du réseau est plus faible (grande). En revanche, Fernández-Huertas Moraga (2011) conteste ces résultats et montre une sélection négative des émigrants mexicains, en reproduisant l'analyse précédente avec des données différentes. Cette divergence s'explique à la fois par

l'omission d'éléments non observables et par le sous-dénombrement des migrants non-qualifiés dans les travaux précédents. Enfin, Kaestner et Malamud (2014) montrent que les migrants mexicains viennent du milieu de la distribution en termes d'éducation, mais sont sélectionnés négativement en matière de revenus.

La sélection des migrants vers les pays de l'OCDE a également été étudiée. Avec une hypothèse de coûts de migration aléatoires, Brücker et Defoort (2009) examinent empiriquement la sélection des migrants en termes de capital humain observable pour six pays de l'OCDE. Ils documentent une sélection positive des migrants, qui peut apparaître même lorsque les inégalités de revenus sont plus importantes dans le pays d'origine que dans le pays de destination. Sur un large éventail de pays, Belot et Hatton (2012) soutiennent l'idée que l'effet de la prime d'éducation, prédit par le modèle Roy-Borjas, ne peut se manifester que lorsque des restrictions associées à la pauvreté (contraintes de liquidité) sont ajoutées dans les spécifications estimées. En outre, les coûts de migration et la situation des pays définissent le schéma de sélection, puisque les facteurs standards des modèles de gravité (c'est-à-dire la distance physique, la similarité culturelle et l'héritage colonial) jouent un rôle plus décisif que les incitations salariales ou la politique d'immigration pour expliquer la sélection des individus qui décident de migrer.

D'autre part, des données originales ont été employées pour évaluer la sélection des migrants économiques. McKenzie *et al.* (2010) étudient une expérience conçue à partir d'une loterie de visas pour analyser les gains de revenus issus de la migration. Ils mettent en évidence la sélection positive des migrants Tongiens en Nouvelle-Zélande sur la base de caractéristiques observables et non observables et la nécessité de prendre en compte les deux dimensions du schéma de sélection. Cette conclusion est appuyée par Bertoli *et al.* (2013), qui considèrent un modèle de Roy-Borjas considérant l'hétérogénéité des individus dans leur propension non-observée à migrer. Sur la base de données au niveau individuel en lien avec l'exode des Équatoriens aux États-Unis et en Espagne, ils construisent une équation de choix discrets qui relâche une des hypothèses (indépendance des alternatives non pertinentes) du modèle de gravité. Ils constatent que les différences de revenus contribuent à expliquer la composition des flux migratoires et que les changements de revenus dans une destination particulière ont un effet plus important sur le choix de la destination que sur l'ampleur de la migration. En utilisant des données historiques, Abramitzky *et al.* (2012) analysent la sélection des migrants Norvégiens aux États-Unis pendant la période de migration de masse (1850-1913). Alors que le modèle de Roy prédit que les migrants devraient être sélectionnés de manière négative (car les revenus en Norvège étaient plus dispersés que les revenus aux États-Unis), les résultats

établissent des conclusions neutres sur la sélection des individus nés en milieu rural et une sélection négative parmi les hommes Norvégiens nés en milieu urbain.

Dans l'ensemble, la sélection des migrants économiques a été abordée pour un large panel de pays et dans le cas d'épisodes migratoires variés. En revanche, les travaux empiriques concernant les migrations forcées sont plus rares et, à ce jour, on ne sait que peu de choses sur le mode de sélection des individus qui ont quitté leur pays d'origine pour demander l'asile ou pour se réfugier dans un pays étranger. Birgier *et al.* (2018) donnent des indices sur la sélection des réfugiés politiques ayant fui l'Argentine et le Chili, lorsque ces pays étaient confrontés à un régime militaire (1976-1983 et 1973-1985, respectivement), vers les États-Unis, la Suède et Israël. Ils établissent que le processus de décision de ces réfugiés à propos du choix de leur destination est similaire à celui des migrants économiques. Les travaux descriptifs de Buber-Ennser *et al.* (2016) sur les personnes arrivées en Autriche en 2015, principalement originaires d'Afghanistan, d'Irak et de Syrie, suggèrent que le niveau d'éducation des demandeurs d'asile est élevé par rapport au niveau d'éducation moyen rencontré dans les pays d'origine.

Plus proche du chapitre I de cette thèse, Lange et Pfeiffer (2018) évaluent la sélection en termes de capital humain des demandeurs d'asile (masculins) en Allemagne. Leurs résultats évoquent une sélection positive des demandeurs d'asile originaires de pays du Moyen-Orient et d'Afrique. Enfin, la partie empirique du travail d'Aksoy et Poutvaara (2020) souligne la sélection favorable des réfugiés et complète mes conclusions pour d'autres pays de destination (que l'Allemagne) choisis par les demandeurs d'asile, cette extension géographique représentant la contribution principale de leur analyse.

Chapitre I et ses contributions

Dans le chapitre I, j'examine le schéma de sélection en matière d'éducation des demandeurs d'asile récemment arrivés en Allemagne en provenance de cinq pays d'origine clés : Afghanistan, Albanie, Irak, Serbie et Syrie. L'analyse repose sur des données individuelles uniques collectées en Allemagne, combinées à des informations provenant d'enquêtes menées dans les pays d'origine. Les résultats révèlent une sélection positive en termes d'éducation des demandeurs d'asile qui ont pu fuir l'Irak et la Syrie, et la sélection est neutre pour les individus demandant l'asile en provenance d'Afghanistan. Le schéma de sélection est négatif pour les demandeurs d'asile venant d'Albanie et de Serbie. Je propose une interprétation de ces résultats fondée sur les différences de durée de séjour anticipée dans le pays de destination, sur les coûts de migration auxquels sont confrontés les demandeurs d'asile pour atteindre l'Allemagne et sur la taille des réseaux de migration à destination.

Le chapitre I est un travail empirique et sa contribution principale se situe donc au niveau des données utilisées dans l'analyse. Ces dernières permettent de pallier à plusieurs difficultés rencontrées dans les études précédentes sur le niveau de capital humain des demandeurs d'asile par rapport à la population d'origine (c'est-à-dire, aux individus qui ont décidé de rester dans le pays d'origine).

Les travaux de Buber-Ennser *et al.* (2016) sont entravés par des problèmes de représentativité des données et par le fait que les informations, au niveau individuel, sur les demandeurs d'asile en Autriche ont été comparées avec des données agrégées pour la population d'origine. Concernant l'étude de Lange et Pfeiffer (2018), le principal problème se situe dans la dimension locale de l'enquête sur les demandeurs d'asile, ce qui implique que les informations recueillies ne sont pas représentatives de la population des demandeurs d'asile en Allemagne. Je contribue à la littérature en utilisant des données individuelles et représentatives au niveau national tant pour les demandeurs d'asile en Allemagne que pour la population des cinq pays d'origine considérés.

Par rapport à Aksoy et Poutvaara (2020), mon travail se concentre uniquement sur un pays de destination, à savoir l'Allemagne. Cela me permet de mettre en avant les différences dans la sélection sur l'éducation au sein de la population des demandeurs d'asile dans le pays d'accueil. Comme noté par Borjas et Monras (2017), mes arguments soutiennent que les conditions dans le pays de destination sont susceptibles d'influencer le modèle de sélection des demandeurs d'asile.

Enfin, en plus de contribuer à la littérature sur les schémas de sélection des migrants, l'analyse prolonge les travaux sur les déterminants des demandes d'asile dans les pays développés (Hatton, 2009, 2016; Neumayer, 2004, 2005; Thielemann, 2006). Au lieu d'utiliser des informations agrégées qui permettent de se focaliser uniquement sur le nombre de demandes d'asile entre les pays, mon étude s'appuie sur des données d'enquêtes afin d'identifier les caractéristiques individuelles des personnes qui demandent l'asile en Allemagne. Ainsi, je suis en mesure de déterminer qui migre depuis les principaux pays d'origine plutôt que d'examiner les facteurs macroéconomiques qui déclenchent la migration des demandeurs d'asile.

Sur les effets des politiques migratoires

Les deux sections précédentes permettent de répondre à deux questions fondamentales : quels sont les facteurs qui incitent les individus à migrer? Qui sont les personnes (en particulier, quel est leur niveau d'éducation) qui prennent la décision de quitter le pays d'origine pour

s'installer à l'étranger ? En comprenant mieux les déterminants spécifiques influençant les flux migratoires internationaux et les caractéristiques individuelles des migrants, les gouvernements (des pays de destination) peuvent définir des politiques visant à réguler les flux migratoires et/ou à privilégier la migration d'une catégorie spécifique de migrants.

Approches méthodologiques

L'étude des effets associés aux politiques migratoires peut d'abord s'inscrire dans le cadre théorique du modèle de gravité des migrations. De manière générale, les changements de politique migratoire dans les pays de destination peuvent conduire à des variations dans les coûts de migration. Dès lors, plusieurs analyses ont estimé l'impact des politiques d'immigration *via* l'inclusion de variables apportant des informations relatives aux politiques globales, c'est-à-dire s'adressant à tous les pays d'origine, ou aux politiques bilatérales dans les spécifications du modèle de gravité. En règle générale, la mise en place de mesures de ce type reflète la volonté des pays de destination d'agir sur l'amplitude des flux d'entrée de migrants.

Un autre éventail de politiques d'immigration peut chercher à filtrer uniquement certains individus par rapport à l'ensemble des personnes qui souhaiteraient potentiellement migrer. L'étude des effets de ces politiques requiert de s'écarter du (stylisé) modèle de gravité. Par exemple, plusieurs contributions théoriques s'intéressent au cas des pays de destination qui ont pour objectif global d'améliorer la qualité des migrants, en les sélectionnant sur la base de leurs caractéristiques observables (cette procédure fait référence aux systèmes à base de points). Bertoli et Rapoport (2015) insistent sur le fait que ces politiques négligent deux effets dynamiques importants : le rôle des réseaux de migrants et la réactivité des décisions en matière d'éducation face à la perspective de migration qu'ont les individus. Leur modèle prédit que les réseaux et la qualité des migrants peuvent être associés positivement, lorsque les pays de destination adoptent des politiques d'immigration suffisamment sélectives.

Bertoli *et al.* (2016b) questionnent l'effet des politiques migratoires sélectives sur les caractéristiques non-observables des individus, à l'aide d'un modèle qui utilise les salaires à destination comme indicateur de la qualité des migrants. Ils montrent que le schéma dominant de sélection en fonction de ces caractéristiques influence l'effet d'une hausse de la sélectivité, ce qui peut entraîner une réduction de la qualité des migrants lorsque ceux-ci sont positivement sélectionnés sur la base de leurs attributs non-observables.

Éléments empiriques

En lien avec les politiques d'immigration globales, Ortega et Peri (2013) présentent une extension du travail de Mayda (2010), dans laquelle la mesure de politique migratoire fait référence à un indice de durcissement des conditions d'entrée dans les pays de destination sur la période 1980-2006 pour 15 pays membres de l'OCDE. Cet indicateur, qui n'est pas comparable entre les différents pays, est associé négativement à l'ampleur des flux migratoires entrants dans chaque pays, dans les estimations où la variabilité entre les destinations n'est pas utilisée pour identifier les coefficients.

Beine *et al.* (2020) analysent la manière dont les dispositions des pays en matière de droits des migrants influencent le choix de la destination des migrants potentiels. En combinant des données sur les intentions de migration bilatérale de plus de 140 pays d'origine et sur les politiques de 38 pays de destination sur la période 2007-2014, ils constatent que les migrants potentiels ont tendance à favoriser les destinations qui sont plus ouvertes à l'inclusion des immigrants dans la société. En particulier, un meilleur accès et de meilleures conditions sur le marché du travail, ainsi que l'accès à la nationalité et à la résidence permanente, augmentent considérablement l'attractivité perçue d'un pays de destination.

Concernant les politiques bilatérales, deux grands types de mesures ont été utilisés dans la littérature. Premièrement, il est possible de capturer la prévalence des accords bilatéraux entre pays : par exemple, Grogger et Hanson (2011) et Beine *et al.* (2019) observent des flux bilatéraux de migrants plus élevés lorsque le pays d'origine et le pays de destination sont tous les deux signataires de l'accord de Schengen. De plus, Beine *et al.* (2019) obtiennent des résultats similaires pour les accords bilatéraux entre pays de l'OCDE recueillis par l'Organisation Internationale pour les Migrations (OIM).

La seconde manière de considérer les politiques bilatérales correspond à la prise en compte de l'existence de visa pour entrer dans les pays de destination. Les exemptions de visa, qui n'appartiennent pas *de jure* au cadre juridique qui régit l'admission des immigrants à destination, peuvent faciliter l'entrée légale des migrants, réduisant ainsi les coûts de migration, et reflètent également un traitement préférentiel au niveau dyadique. Bertoli et Fernández-Huertas Moraga (2013) apportent des preuves de l'impact de l'exemption de visa sur les flux migratoires bilatéraux vers l'Espagne. Des conclusions similaires sont mises en avant par Bertoli et Fernández-Huertas Moraga (2015) et Beine et Parsons (2015), ces derniers utilisant des données longitudinales sur les politiques de visa bilatérales collectées par le projet DEMIG de l'Université d'Oxford.

En plus de l'impact direct des changements en termes de politique migratoire, la présence d'externalités a également été reconnue dans la littérature sur les migrations internationales, même si la quantification de ces effets reste rare. Deux exceptions notables sont Bertoli et Fernández-Huertas Moraga (2015), qui établissent des bornes associées aux effets indirects de la politique de visa sur les flux migratoires à travers les destinations, et Beverelli et Orefice (2019), qui documentent l'existence d'une réorientation entre des pays d'origine économiquement similaires et avec un certain degré d'affinité culturelle (par exemple, le fait de partager une langue commune).

Enfin, quelques résultats empiriques sont à signaler au sujet des politiques d'immigration visant à sélectionner les migrants. Aydemir (2011) étudie le système de points mis en place au Canada pour filtrer les individus, indiquant qu'il permet d'obtenir un niveau de compétences des immigrants beaucoup plus élevé que celui qui aurait été obtenu si les individus avaient été distingués en fonction de leurs préférences familiales. Cette sélection positive est le fruit du choix direct des candidats les plus qualifiés, mais aussi la conséquence indirecte du regroupement familial de conjoints avec des compétences élevées. Dans le contexte des États-Unis, Kato et Sparber (2013) évalue la réduction drastique (décision prise en Octobre 2003) du nombre de visas H-1B disponibles pour les travailleurs nés à l'étranger. Leurs estimations suggèrent que cette politique d'immigration restrictive a découragé de manière disproportionnée les étudiants étrangers à haut potentiel de poursuivre leurs études aux États-Unis.

Plus récemment, Bertoli et Stillman (2019) utilisent des données pour les États-Unis afin d'analyser le chevauchement de la distribution des salaires des migrants récents, peu et très instruits, en provenance de divers pays d'origine et après avoir pris en compte les autres caractéristiques observables. Lorsqu'ils font correspondre de manière aléatoire un immigrant hautement qualifié avec un immigrant peu qualifié en provenance du même pays, plus d'un quart du temps le migrant peu qualifié a un salaire horaire plus élevé, malgré une différence statistiquement significative dans le salaire moyen des deux groupes pour la plupart des origines. Pour 98 des 114 pays d'origine de leur échantillon, cette mesure synthétique est supérieure à la valeur correspondante pour les natifs. Cela suggère que le fait de s'appuyer fortement sur l'éducation pour sélectionner les immigrants pourrait ne pas améliorer sensiblement leur qualité.

Naghsh Nejad et Schurer (2019) quantifient les composantes traditionnellement inobservables de la qualité des migrants en Australie, un pays de l'OCDE à forte proportion de migrants et doté d'une politique d'immigration sélective, à l'aide d'indicateurs de la personnalité et des capacités cognitives déjà employés dans la littérature. Ils observent que les immigrants de première et de deuxième génération obtiennent de meilleurs résultats que les natifs en

termes de traits de personnalité socialement bénéfiques. Alors que les migrants de la première génération sont pénalisés sur le plan de la langue, leurs descendants surmontent les difficultés linguistiques et ont de meilleurs résultats que les natifs en matière de capacités cognitives.

Digression sur les déterminants des migrations forcées

Le chapitre III de cette thèse est également associé à la littérature sur les facteurs qui déterminent les stocks de réfugiés et les flux d'asile dans le monde. Les travaux sur ce sujet considèrent généralement une version plus étendue du modèle de gravité des migrations, dans laquelle plusieurs variables spécifiques aux migrations forcées sont ajoutées dans les équations estimées, en plus des facteurs standards comme les conditions économiques dans les pays d'origine et de destination ou les variables dyadiques telles que la distance, les liens coloniaux, la proximité linguistique ou les réseaux de migrants.

En se focalisant sur les pays d'origine, Davenport *et al.* (2003) constatent que le stock de personnes déplacées pouvait s'expliquer principalement par le génocide, la guerre civile, les conflits dissidents et les transitions de régime politique. Moore et Shellman (2007) ont des résultats similaires dans le cadre d'une étude sur les mouvements bilatéraux de réfugiés, en observant aussi l'effet des conflits dans les pays frontaliers et des coûts de migration pour les mouvements plus lointains. Le PIB par habitant dans le pays d'origine a un effet négatif sur les déplacements de réfugiés et sur les flux d'asile vers le monde développé (Hatton, 2009), de sorte que les conditions économiques dans les pays d'origine semblent avoir de l'importance.

Hatton (2016) propose une quantification des effets estimés, qui s'applique dans le cas de son échantillon composé de 19 pays de destination membres de l'OCDE et 48 pays d'origine. Sur la période 2006-2012, les coefficients prédisent que le niveau de terreur politique et l'absence de libertés civiles n'augmenteraient le nombre de demandes d'asile que de 3%, mais avec une grande hétérogénéité entre les pays : une hausse de 50% pour l'Érythrée et le Nigeria, et une augmentation de 108% pour la Syrie. Ces résultats montrent que la terreur politique et les violations des droits de l'homme sont au cœur des départs de réfugiés, alors que, même si l'amélioration des conditions économiques dans les pays d'origine et de transit serait utile, une augmentation de 10% du PIB par habitant des pays d'origine ne réduirait le nombre de demandes d'asile que d'environ 5%.

Plus proche du travail réalisé dans cette thèse, plusieurs études ont introduit des variables afin d'intégrer la politique d'asile des pays de destination comme un déterminant des demandes d'asile, tout en contrôlant pour les facteurs de pression et d'attraction plus standards décrits

précédemment. L'idée principale s'articule autour de l'élaboration d'indices, comparables dans le temps et entre les pays, permettant de suivre la position politique des pays en matière de migration et/ou d'asile. Thielemann (2004, 2006) construit un indice de dissuasion relatif aux politiques d'asile, tandis que Neumayer (2004) fait usage de variables de substitution (par exemple, le pourcentage de portefeuilles ministériels détenus par les partis de gauche) pour rendre compte de la tendance attendue dans les lois promulguées par les pays d'accueil à l'égard des demandeurs d'asile. Plus récemment, Hatton (2016) estime une équation de gravité afin de définir l'effet de trois catégories de politique selon leurs priorités (accès, traitement et bien-être) vis-à-vis des demandes d'asile. Les variables sont basées sur la collecte de données introduite d'abord dans Hatton (2004), puis étendue dans Hatton (2009).

Contributions du chapitre III

Dans le troisième chapitre, je tente de répondre à la question suivante : est-ce que l'adoption d'une liste de pays d'origine sûrs influence le nombre de demandes d'asile déposées dans les États membres de l'OCDE ? Je m'appuie sur un modèle de gravité structurel pour dériver une équation empirique de migration qui est évaluée avec des données dans le but d'estimer l'effet direct de la liste sur le nombre de demandes d'asile entre deux pays. Ce coefficient permet ensuite de résoudre le modèle structurel pour quantifier les externalités provenant d'une expérience contrefactuelle sur la politique des pays sûrs.

La contribution de ce chapitre à la littérature sur les déterminants des migrations (forcées) est double. Premièrement, l'article propose une méthodologie permettant d'estimer les externalités directes et indirectes associées à des changements de politique d'immigration bilatérale imposés par le pays de destination, sur la base d'un modèle de gravité structurel des migrations proposé par Anderson (2011). Deuxièmement, le cadre ci-dessus est appliqué au cas particulier d'une politique migratoire centrée sur les demandeurs d'asile. L'analyse implique la collecte d'informations précises (et uniques) sur l'évolution des listes de pays d'origine sûrs dans les pays de l'OCDE, afin d'établir leurs effets sur le nombre bilatéral de demandes d'asile.

Dans une première étape, j'estime une équation de gravité des migrations pour obtenir l'effet direct de la politique d'asile sur les demandes d'asile logées dans les pays de destination. Le coefficient estimé est ensuite utilisé, dans une seconde étape, pour résoudre le modèle de gravité structurel et quantifier les effets de diversion qui proviennent d'une expérience contrefactuelle sur la liste des pays d'origine sûrs. L'application de cette procédure avec les données disponibles permet de documenter l'effet dissuasif de la politique d'asile sur le nombre bilatéral de demandes d'asile. En particulier, l'analyse empirique révèle que l'introduction d'une

liste de pays d'origine sûrs entraîne une diminution d'environ 30% du nombre de demandes d'asile entre les pays. L'exercice de simulation autour d'un changement hypothétique de la politique d'asile suggère la présence d'effets de diversion sur le volume de demandes d'asile entre *à la fois* les pays d'origine et de destination.

Bibliographie

Abramitzky, R., L. P. Boustan, et K. Eriksson (2012) : "Europe's Tired, Poor, Huddled Masses : Self-Selection and Economic Outcomes in the Age of Mass Migration," *American Economic Review*, 102(5), 1832–56.

Adserà, A., et M. Pytliková (2015) : "The Role of Language in Shaping International Migration," *Economic Journal*, 125(586), F49–F81.

Aksoy, C. G., et P. Poutvaara (2020) : "Refugees' and Irregular Migrants' Self-selection into Europe : Who Migrates Where ?," SOEP papers on Multidisciplinary Panel Data Research 1069, DIW Berlin, The German Socio-Economic Panel (SOEP).

Anderson, J. E. (2011) : "The Gravity Model," *Annual Review of Economics*, 3(1), 133–160.

Aydemir, A. (2011) : "Immigrant Selection and Short-term Labor Market Outcomes by Visa Category," *Journal of Population Economics*, 24(2), 451–475.

Beine, M., S. Bertoli, et J. Fernández-Huertas Moraga (2016) : "A Practitioners' Guide to Gravity Models of International Migration," *World Economy*, 39(4), 496–512.

Beine, M., P. Bourgeon, et J.-C. Bricongne (2019) : "Aggregate Fluctuations and International Migration," *The Scandinavian Journal of Economics*, 121(1), 117–152.

Beine, M., F. Docquier, et Çağlar Özden (2011) : "Diasporas," *Journal of Development Economics*, 95(1), 30–41.

Beine, M., J. Machado, et I. Ruyssen (2020) : "Do Potential Migrants Internalise Migrant Rights in OECD Host Societies ?," *Canadian Journal of Economics*, Forthcoming.

Beine, M., et C. Parsons (2015) : "Climatic Factors as Determinants of International Migration," *The Scandinavian Journal of Economics*, 117(2), 723–767.

Belot, M., et S. Ederveen (2012) : "Cultural Barriers in Migration Between OECD Countries," *Journal of Population Economics*, 25(3), 1077–1105.

Belot, M. V. K., et T. J. Hatton (2012) : "Immigrant Selection in the OECD," *Scandinavian Journal of Economics*, 114(4), 1105–1128.

Bertoli, S., H. Brücker, et J. Fernández-Huertas Moraga (2016a) : "The European Crisis and Migration to Germany," *Regional Science and Urban Economics*, 60, 61–72.

Bertoli, S., V. Dequiedt, et Y. Zenou (2016b) : "Can Selective Immigration Policies Reduce Migrants' Quality ?," *Journal of Development Economics*, 119, 100–109.

Bertoli, S., et J. Fernández-Huertas Moraga (2013) : "Multilateral Resistance to Migration," *Journal of Development Economics*, 102, 79–100.

——— (2015) : "The Size of the Cliff at the Border," *Regional Science and Urban Economics*, 51, 1–6.

Bertoli, S., J. F.-H. Moraga, et F. Ortega (2013) : "Crossing the Border : Self-Selection, Earnings and Individual Migration Decisions," *Journal of Development Economics*, 101, 75–91.

Bertoli, S., et H. Rapoport (2015) : "Heaven's Swing Door : Endogenous Skills, Migration Networks, and the Effectiveness of Quality-Selective Immigration Policies," *The Scandinavian Journal of Economics*, 117(2), 565–591.

Bertoli, S., et S. Stillman (2019) : "All that Glitters is not Gold : Wages and Education for US Immigrants," *Labour Economics*, 61, 101749.

Beverelli, C., et G. Orefice (2019) : "Migration Deflection : The Role of Preferential Trade Agreements," *Regional Science and Urban Economics*, 79, 103469.

Birgier, D. P., C. Lundh, Y. Haberfeld, et E. Elldér (2018) : "Self-Selection and Host Country Context in the Economic Assimilation of Political Refugees in the United States, Sweden, and Israel," *International Migration Review*, 52(2), 524–558.

Borjas, G. J. (1987) : "Self-Selection and the Earnings of Immigrants," *American Economic Review*, 77(4), 531–553.

——— (1991) : "Immigration and Self-Selection," in *Immigration, Trade, and the Labor Market*, pp. 29–76. University of Chicago Press.

Borjas, G. J., et J. Monras (2017) : "The Labor Market Consequences of Refugee Supply Shocks," *Economic Policy*, 32(91), 361–413.

Brücker, H., et C. Defoort (2009) : "Inequality and the Self-Selection of International Migrants : Theory and New Evidence," *International Journal of Manpower*, 30(7), 742–764.

Buber-Ennser, I., J. Kohlenberger, B. Rengs, Z. Al Zalak, A. Goujon, E. Striessnig, M. Potančoková, R. Gisser, M. R. Testa, et W. Lutz (2016) : "Human Capital, Values, and Attitudes of Persons Seeking Refuge in Austria in 2015," *PLOS ONE*, 11(9), 1–29.

Chiquiar, D., et G. Hanson (2005) : "International Migration, Self-Selection, and the Distribution of Wages : Evidence from Mexico and the United States," *Journal of Political Economy*, 113(2), 239–281.

Clemens, M. A., C. Montenegro, et L. Pritchett (2019) : "The Place Premium : Bounding the Price Equivalent of Migration Barriers," *Review of Economics and Statistics*, 101(2), 201–213.

Davenport, C., W. Moore, et S. Poe (2003) : "Sometimes You Just Have to Leave : Domestic Threats and Forced Migration, 1964-1989," *International Interactions*, 29(1), 27–55.

Fernández-Huertas Moraga, J. (2011) : "New Evidence on Emigrant Selection," *The Review of Economics and Statistics*, 93(1), 72–96.

Giordani, P. E., et M. Ruta (2013) : "Coordination Failures in Immigration Policy," *Journal of International Economics*, 89(1), 55–67.

Grogger, J., et G. H. Hanson (2011) : "Income Maximization and the Selection and Sorting of International Migrants," *Journal of Development Economics*, 95(1), 42–57, Symposium on Globalization and Brain Drain.

Hatton, T. J. (2004) : "Seeking Asylum in Europe," *Economic Policy*, 19(38), 5–62.

——— (2009) : "The Rise and Fall of Asylum : What Happened and Why?," *The Economic Journal*, 119(535), F183–F213.

——— (2016) : "Refugees, Asylum Seekers, and Policy in OECD Countries," *American Economic Review*, 106(5), 441–45.

Kaestner, R., et O. Malamud (2014) : "Self-Selection and International Migration : New Evidence from Mexico," *Review of Economics and Statistics*, 96(1), 78–91.

Kato, T., et C. Sparber (2013) : "Quotas and Quality : The Effect of H-1B Visa Restrictions on the Pool of Prospective Undergraduate Students from Abroad," *The Review of Economics and Statistics*, 95(1), 109–126.

Lange, M., et F. Pfeiffer (2018) : "The Human Capital Selection of Young Males Seeking Asylum in Germany," ZEW Discussion Paper 18-040, ZEW Mannheim.

Mayda, A. M. (2010) : "International Migration : A Panel Data Analysis of the Determinants of Bilateral Flows," *Journal of Population Economics*, 23(4), 1249–1274.

McFadden, D. (1974) : "Conditional Logit analysis of Qualitative Choice Behavior," in *Frontiers in Econometrics*, ed. by P. Zarembka, pp. 105–142. New York : Academic Press.

McKenzie, D., et H. Rapoport (2010) : "Self-Selection Patterns in Mexico-US Migration : The Role of Migration Networks," *The Review of Economics and Statistics*, 92(4), 811–821.

McKenzie, D., S. Stillman, et J. Gibson (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.

McKenzie, D., C. Theoharides, et D. Yang (2014) : "Distortions in the International Migrant Labor Market : Evidence from Filipino Migration and Wage Responses to Destination Country Economic Shocks," *American Economic Journal : Applied Economics*, 6(2), 49–75.

Moore, W. H., et S. M. Shellman (2007) : "Whither Will They Go ? A Global Study of Refugees' Destinations, 1965-1995," *International Studies Quarterly*, 51(4), 811–834.

Munshi, K. (2003) : "Networks in the Modern Economy : Mexican Migrants in the US Labor Market," *The Quarterly Journal of Economics*, 118(2), 549–599.

Naghsh Nejad, M., et S. Schurer (2019) : "Cognitive and Non-Cognitive Abilities of Immigrants : New Perspectives on Migrant Quality from a Selective Immigration Country," Discussion Paper 12700, IZA Discussion Papers.

Neumayer, E. (2004) : "Asylum Destination Choice : What Makes Some West European Countries More Attractive Than Others ?," *European Union Politics*, 5(2), 155–180.

——— (2005) : "Bogus Refugees ? The Determinants of Asylum Migration to Western Europe," *International Studies Quarterly*, 49(3), 389–410.

Ortega, F., et G. Peri (2013) : "The Role of Income and Immigration Policies in Attracting International Migrants," *Migration Studies*, 1(1), 47–74.

Pedersen, P. J., M. Pytlikova, et N. Smith (2008) : "Selection and network effects : Migration flows into OECD countries 1990-2000," *European Economic Review*, 52(7), 1160–1186.

Porcher, C. (2019) : "Migration with Costly Information," mimeo, Princeton University.

Roy, A. D. (1951) : "Some Thoughts on the Distribution of Earnings," *Oxford Economic Papers*, 3(2), 135–146.

Sjaastad, L. A. (1962) : "The Costs and Returns of Human Migration," *Journal of Political Economy*, 70(5), 80–93.

Small, K. A., et H. S. Rosen (1981) : "Applied Welfare Economics with Discrete Choice Models," *Econometrica*, 49(1), 105–130.

Thielemann, E. (2004) : "Why Asylum Policy Harmonisation Undermines Refugee Burden-Sharing," *European Journal of Migration and Law*, 6(1), 47–65.

——— (2006) : "The Effectiveness of Governments' Attempts to Control Unwanted Migration," in *Immigration and the Transformation of Europe*, ed. by C. Parsons, et T. Smeeding. Cambridge University Press.

Vogler, M., et R. Rotte (2000) : "The Effects of Development on Migration : Theoretical Issues and New Empirical Evidence," *Journal of Population Economics*, 13(3), 485–508.

Chapter I*

*This chapter has been published in *Demography* (https://doi.org/10.1007/s13524-020-00873-9).

Self-selection of Asylum Seekers: Evidence From Germany[*]

Lucas Guichard[a,b]

[a] *Université Clermont Auvergne, CNRS, IRD, CERDI*
[b] *IAB*[†]

Abstract

I examine the pattern of selection on education of asylum seekers recently arrived in Germany from five key source countries: Afghanistan, Albania, Iraq, Serbia, and Syria. The analysis relies on original individual-level data collected in Germany combined with surveys conducted at origin. The results reveal a positive pattern of selection on education for asylum seekers who were able to flee Iraq and Syria, and the selection is neutral for individuals seeking asylum from Afghanistan and negative for asylum seekers from Albania and Serbia. I provide an interpretation of these patterns based on differences in the expected length of stay at destination, the migration costs faced by asylum seekers to reach Germany, and the size of migration networks at destination.

Keywords: Refugee; Selection, Education, Individual-level data.

[*]I thank Ran Abramitzky, Simone Bertoli, Delphine Boutin, George J. Borjas, Hanna Brenzel, Herbert Brücker, Andreas Hauptmann, Sarah Langlotz, Steffen Sirries, Ignat Stepanok, Jérôme Valette, and the participants in the CEMIR Workshop on Migration Research (Munich, June 2017); the interdisciplinary Workshop on Demographic Change, Migration and Integration (Jena, September 2017); the 4th BAGSS Conference on Challenges for Diverse Societies; the 7th Annual CEPII-OECD Conference on Immigration in OECD Countries (Paris, December 2017); the CERDI PhD Seminar (Clermont-Ferrand, December 2017); and the 13th International German Socio-Economic Panel User Conference (Berlin, July 2018) for their comments and suggestions. Any remaining errors are my own. I acknowledge support received from the *Agence Nationale de la Recherche* of the French government through the program "*Investissements d'avenir*" (ANR-10-LABX-14-01).
[†]Regensburger Strasse, 104, D-90478 Nuremberg; E-mail: Lucas.Guichard2@iab.de.

1.1 Introduction

European countries experienced a short-lived surge in the arrival of asylum seekers from 2014 to 2016. More than 1.2 million first-time asylum applications were registered in the European Union in 2015 (Eurostat, 2016), with Germany receiving approximately three-quarters of the applications lodged that year (BMI, 2017).[1] Because of the ongoing crisis in Syria, most asylum seekers came from there (41.5%), but some originated from other conflict-affected areas (18.1% from Afghanistan and Iraq combined) and from eastern European countries (5.9%). The size, diversity, and potential consequences of the large number of asylum seekers make it important to identify the characteristics of the newcomers, which are likely to affect the socio-economic outcomes of the stayers in the origin country and of the natives at destination.[2]

The push factors behind the decisions of asylum seekers to migrate have been emphasized in the public debate as a pivotal feature differentiating them from economic migrants. The latter are often assumed to be able to choose whether to migrate, whereas asylum seekers are, in principle, forced to flee their country of origin because of threats to their lives. The drivers of economic migration have been widely studied in the literature. However, the determinants that explain who is able to leave the home country to seek asylum abroad have been rarely explored. In this study, I exploit individual-level and representative data related to the recent surge in asylum applications to Germany to improve the current limited knowledge and understanding about the mechanisms fostering the migration decision in the context of forced migration.

This study focuses on the self-selection on education of asylum seekers who arrived in Germany from 2013 or later.[3] It delivers the first insights on this question for individuals drawn from the origin population of five source countries: Afghanistan, Albania, Iraq, Serbia, and Syria. These countries represent 65% of all first-time asylum applications lodged in Germany, and they offer an interesting variety of economic and security conditions at origin, allowing an

[1] The figure represents only a tiny fraction of all refugees. The number of refugees who were able to migrate to Germany is high compared with other European countries but small relative to neighboring countries of the main asylum source countries (United Nations High Commissioner for Refugees, 2016). Thus, the observed pattern of selection is likely not to be representative of the entire population of forced migrants. See the Descriptive Evidence section for further discussion of this issue.

[2] Hanson and McIntosh (2016) argued that the networks created by the surge of asylum seekers are likely to attract future migrants to come to Germany for years. The (short-lived) arrival of asylum seekers could then have a first-order impact on the long-term evolution of immigration to Germany, that the authors were otherwise predicting to decline.

[3] Education is an important characteristic to evaluate the selection of asylum seekers. Specifically, the literature documents a positive correlation between education and wealth, and high-educated individuals are thus likely to be better-off (Card, 1999). Because of data constraints, the analysis can account for only the level of education, which could be seen as a proxy of the socio-economic condition of a given individual.

investigation of variations in the pattern of selection of asylum seekers coming from different countries.

These different conditions are key to describing the origin-specific pattern of selection that prevails for asylum seekers in Germany. Individuals from Afghanistan, Iraq, and Syria are likely to be in danger at home, but asylum seekers from the Balkan region left countries considered to be safe.[4] The level of threats that can be encountered in the origin country largely determines the high (low) rates of acceptance of asylum applications from conflict-affected (Balkan) countries.[5] Accordingly, this could lead to differences in the expected duration of stay in Germany, such that asylum seekers from Afghanistan, Iraq, and Syria have a longer time horizon in the host country, compared with asylum seekers from Albania and Serbia, who are legally entitled to stay at destination only until their applications are rejected, something that almost invariably occurs.[6]

Albanians and Serbians have not needed a visa to enter the European Union since 2010 and 2009, respectively, and this facilitates a legal entry into the Schengen area. Germany was among the countries fearing a surge in asylum applications from Albania and Serbia after the visa requirement was lifted (Bertoli and Fernández-Huertas Moraga, 2015), although this surge did not immediately materialize. Serbians and particularly Albanians started applying for asylum in Germany in large numbers in 2015, when the surge in applications from conflict-affected countries resulted in major delays in the processing of asylum claims.[7] The processing time possibly increased the expected return from lodging an application for Albanians and Serbians, given that they were legally protected from the risk of deportation while their applications were processed, and could get access to welfare benefits.[8]

[4] Throughout this article, Afghanistan, Iraq, and Syria are defined as conflict-affected countries. This denotation is straightforward for Syria because of the ongoing civil conflict, but is less so for Afghanistan and Iraq given that the conflict occurred in the past. The recent status in Europe of asylum seekers from the last two countries is different, mainly because of the conditions at origin. Some areas in Afghanistan are now considered safe, and the EU has signed an agreement (October 4, 2016) with Kabul (European Union and Islamic Republic of Afghanistan, 2016) to implement the deportation of rejected asylum seekers from Afghanistan.

[5] Unfortunately, I do not have information about the ethnicity of asylum seekers because German survey questions (mostly for historical reasons) do not ask about it. Thus, I cannot investigate whether asylum seekers from Albania and Serbia are predominantly Sinti or Roma, a group that is heavily stigmatized and partly persecuted in the Balkan countries. Moreover, this would also have been an additional explanation for the negative selection of asylum seekers from the Balkan region given that Sinti and Roma often do not participate in the general educational system.

[6] In 2015, the acceptance rate was 72.8% for asylum applications from Afghanistan, 98.3% for Iraq, and 97.7% for Syria. Only 0.2% (0.1%) of Albanian (Serbian) asylum seekers were granted a refugee status (Eurostat, 2018).

[7] The decision to seek refuge in Germany from Albania has been related to the influence of smugglers along refugee routes toward European countries, with asylum seekers grossly overestimating the chances of being granted a refugee status. The German Embassy in Tirana implemented ad campaigns to warn people against seeking asylum in Germany (Meisner, 2015).

[8] The origin-specific expected processing time of the asylum applications averaged at 9 and 15.5 months for, respectively, Albania and Serbia in 2015 (see section on empirical results).

Different expected durations of stay in Germany influence the pattern of selection of asylum seekers with respect to education, through the returns to education at destination that increase with the time spent since migration (Dustmann and Glitz, 2011).[9] The longer time horizon of individuals from conflict-affected countries would imply a favorable selection on education.[10] By contrast, Balkan asylum seekers are more likely to be negatively selected because of their greater probability of staying temporarily in Germany. This pattern of selection is consistent with the high (low) migration costs faced by asylum seekers originating from conflict-affected (Balkan) countries. Liquidity constraints on the decision of individuals from Afghanistan, Iraq, and Syria to migrate drive a positive selection with respect to education, whereas Albanians and Serbians encounter low migration costs to move to Germany. Moreover, the migration history of the five selected countries could also play a role in the selection of asylum seekers. Large migration networks from Serbia in Germany before the asylum surge might have facilitated the arrival of asylum seekers from these countries by decreasing the migration costs, resulting in a more negative pattern of selection on education.

I explore these predictions on the selection of asylum seekers in a country-by-country analysis of original data on asylum seekers in Germany. Comprehensive characteristics of asylum seekers are obtained from a survey conducted jointly by the Institute for Employment Research (IAB); the Research Centre on Migration, Integration, and Asylum of the Federal Office of Migration and Refugees (BAMF); and the Socio-Economic Panel (SOEP) at DIW Berlin. The IAB-BAMF-SOEP Refugee Sample allows me to exploit a large set of cases, which includes 4,328 asylum seekers. The data are matched with surveys conducted in the origin countries. Relevant information is combined into country-specific samples, and the empirical analysis uses a logistic model to examine the selection of asylum seekers with respect to education. Individuals claiming asylum in Germany from Iraq and Syria are shown to be positively selected on education, and the results provide mixed evidence on the selection of asylum seekers from Afghanistan. On the other hand, Albanian and Serbian asylum seekers are found to be drawn from the lower tail of the education distribution.

This article is related to various strands of the migration literature, in which the self-selection of immigrants has been widely studied, albeit rarely in the case of asylum seekers. Building on

[9] The computation of the returns to education at origin and at destination would be helpful to predict the selection patterns. Unfortunately, such a computation is not possible with the available data, thus preventing a more structural estimation by first estimating income and then estimating the relationship between location choices and income, as in Bertoli *et al.* (2013).

[10] Recently, Chen *et al.* (2017) studied the self-selection of Pakistani migrants who decided ex ante between temporary and permanent internal migration. The authors demonstrated that the expected duration of stay affects the selection of migrants and that a permanent move drives a positive pattern of selection with respect to education.

the idea that observable and unobservable characteristics influence the (pecuniary) benefits of migration, Borjas (1987, 1991) extended the Roy (1951) model to determine which individuals find migrating optimal. This seminal work was followed by several other contributions (Chiquiar and Hanson, 2005; Chiswick, 1999; Grogger and Hanson, 2011). The implications derived from the Roy-Borjas model have been empirically studied for economic migrants in a variety of migration scenarios. Beginning with Chiquiar and Hanson (2005), analyses on the selection of immigrants from Mexico to the United States (Fernández-Huertas Moraga, 2011; Kaestner and Malamud, 2014; McKenzie and Rapoport, 2010) and from different origin countries to OECD member states (Belot and Hatton, 2012; Brücker and Defoort, 2009; Mayda, 2010)) have flourished. More recently, Aksoy and Poutvaara (2019) extended the Roy-Borjas framework to account for the risks associated with conflicts or persecution. Their model shows that migrants from countries experiencing a major conflict are expected to be positively selected, even when the returns to skill at origin would be higher than in destination countries. Borrowing constraints strengthen the positive pattern of selection: individuals with more education are likely to have more resources and to be willing to leave in times of crises. To date, little is known about the pattern of selection of individuals who left their country of origin to seek asylum abroad. Birgier *et al.* (2016) provided evidence on the selection of political refugees fleeing Argentina and Chile during the military regimes there (1976-1983 and 1973-1985, respectively) to the United States, Sweden, and Israel. They documented that the decision process of these refugees regarding the choice of their destination is similar to those of economic immigrants. The descriptive work of Buber-Ennser *et al.* (2016) on individuals who arrived in Austria in 2015, mainly originating from Afghanistan, Iraq, and Syria, documented that the educational level of these asylum seekers was high relative to the average level of education in the origin countries. However, the analysis was hindered by representativeness issues regarding some of the data used and by the fact that individual-level information about asylum seekers in Austria was compared only with aggregate data of the origin population. More closely related to the current study, Lange and Pfeiffer (2018) evaluated the human capital selection of male asylum seekers in Germany. Their results suggested a positive selection of asylum seekers from Middle Eastern and African countries, who had 22% more years of schooling than the same-aged individuals in the origin country. The main difficulty of this study is the local dimension of the survey of asylum seekers, which implies that collected information is not representative of the asylum population in Germany. I contribute to the literature through the use of individual-level and nationally representative data for both the asylum seekers in Germany and the home-country population. Empirical support provided by Aksoy and Poutvaara (2019) highlights the favorable selection of refugees and complements my findings for other asylum destination countries. Their main contribution is to extend

the analysis to other destination (or transit) countries of the recent refugee arrival in Europe. Focusing on Germany, I am able to depict differences in the selection on education *within* the refugee population in the host country, and I attempt to provide an interpretation of the observed patterns. Last, a recent work by Borjas and Monras (2017, p. 376) noted that conditions at destination may influence the pattern of selection of asylum seekers. In line with this argument, I argue that the origin-specific expected length of stay in Germany is likely to explain differences in the selection of asylum seekers with respect to education.

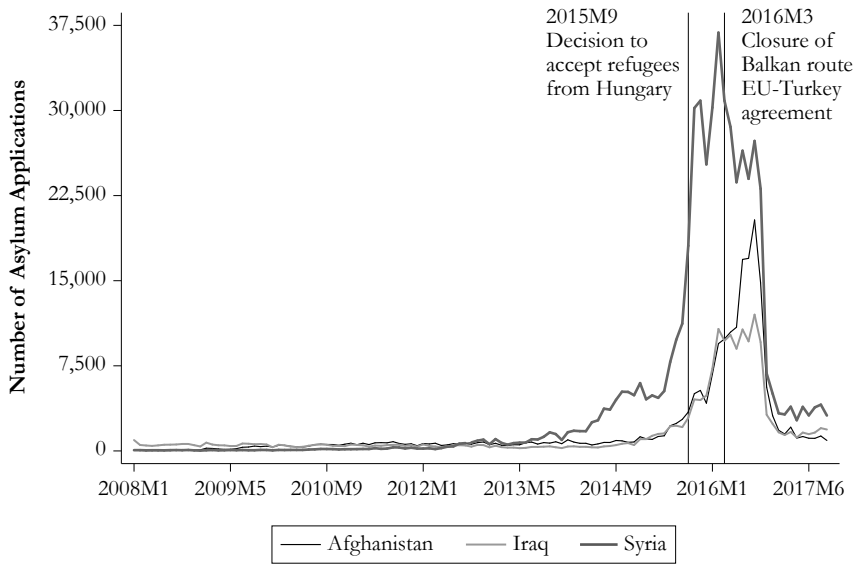
In addition to contributing to the self-selection literature, this analysis extends work on the determinants of asylum applications to developed countries (Hatton, 2009, 2016; Neumayer, 2004, 2005; Thielemann, 2006). Instead of using aggregate information that allows a focus on only the sheer scale of asylum applications, my study relies on survey data to identify the characteristics of individuals seeking asylum in Germany. As a result, I am able to evaluate who migrates from the main asylum source countries rather than analyzing the macroeconomic forces that trigger the migration of asylum seekers.

1.2 Selected Countries of Origin and Data Sources

1.2.1 Selected Countries of Origin

The recent evolution of the number of asylum applications lodged in Germany has implied several changes in the German asylum policy and raised interest in studying the characteristics of the newcomers. The civil conflict in Syria forced the migration of 5.5 million Syrians, and most asylum seekers who were able to move to Europe went to Germany (United Nations High Commissioner for Refugees, 2016). The related surge in the number of Syrian asylum applications is reflected in Fig. 1.1. Germany also experienced a large influx of asylum seekers fleeing turmoil in Afghanistan and Iraq (also shown in Fig. 1.1) as well as asylum seekers from the Balkan region (Fig. 1.2). However, the pattern is remarkably different between Albania and Serbia. The evolution of asylum applications from Albania is similar to the one of conflict-affected countries, whereas asylum claims from Serbia were more evenly spread (between 0 and 2,000 applications) over a longer period.

Figure 1.1: Asylum applications in Germany from conflict-affected countries



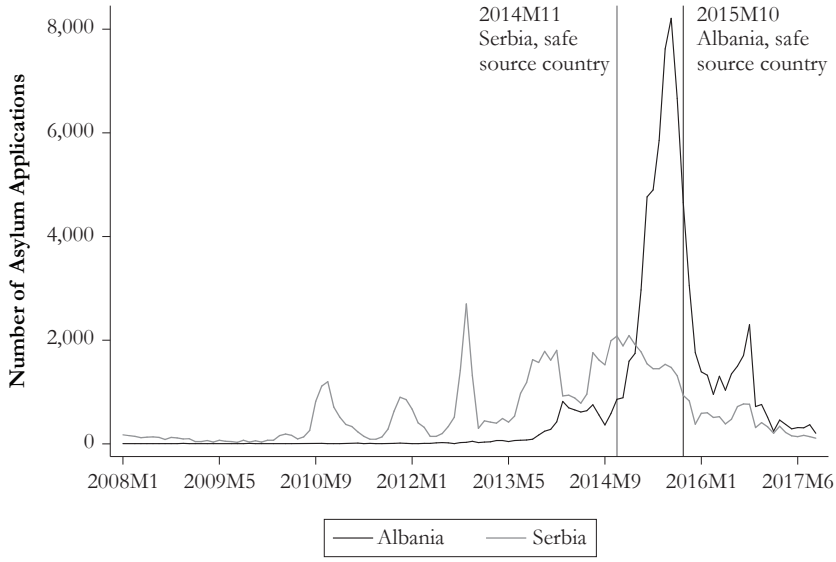
Source: Author's elaboration based on Eurostat (2017b).

This large influx of asylum seekers has prompted several changes in the asylum policy of destination countries. European countries closed the Western Balkan route (March 9, 2016) and implemented an agreement with Turkey shortly thereafter (March 18, 2016). The latter aimed to address the overwhelming arrival of smuggled asylum seekers going across the Aegean Sea from Turkey to the Greek islands, by allowing Greece to deport to Turkey "all new irregular migrants" (European Council, 2016) arriving since March 20, 2016. In return, EU member states agreed to increase the resettlement of Syrian refugees residing in Turkey, enhance visa liberalization for Turkish nationals, and expand existing financial support for the refugee population in Turkey. These decisions can certainly explain the downward slope in the number of applications beginning in mid-2016. At the national level, German authorities reacted to the inflow of asylum seekers from the Balkan region by repeatedly modifying its list of safe countries of origin.[11] Serbia was included in November 2014 (along with the Republic of

[11] The *safe country of origin* concept is a presumption that certain countries can be designated as safe for their nationals to the extent that "it can be shown that there is generally and consistently no persecution as defined in Article 9 of Directive 2004/83/EC, no torture or inhuman or degrading treatment or punishment and no threat by reason of indiscriminate violence in situations of international or internal armed conflict" (Council of the European Union, 2005). This implies that a claim for international protection by an applicant from a safe country of origin is likely to be considered unfounded.

Macedonia and Bosnia-Herzegovina),[12] and Albania was added in October 2015 (with Kosovo and Montenegro).[13] This policy change triggered a decrease in the number of asylum claims from these two countries.

Figure 1.2: Asylum applications in Germany from Balkan countries



Source: Author's elaboration based on Eurostat (2017b).

Asylum seekers in Germany are mainly from conflict-affected countries (i.e., Afghanistan, Iraq, Syria) and the Balkan region (Albania, Serbia, Kosovo), but also come from a few other countries (e.g., Eritrea, Somalia, Iran, and Pakistan). The sheer scale of the asylum surge is shown in Table 1.1, which reports the number of asylum seekers across origins recorded by the Federal Office for Migration and Refugees (BAMF) between the beginning of 2013 and the end of January 2016. As a consequence of the ongoing crisis in Syria, 41.5% of the asylum seekers originate from this country; individuals from Afghanistan and Iraq correspond, respectively, to 9.8% and 8.3% of asylum seekers.

The self-selection of asylum seekers from the origin population is examined for a limited number of source countries: Afghanistan, Albania, Iraq, Serbia, and Syria. These five countries represent roughly 65% of all recent asylum seekers in Germany (Table 1.1). Moreover, they

[12] Law on classification of further states as safe countries of origin and on the facilitation of access to the labor market for asylum seekers and tolerated foreigners (Federal Ministry of Justice and Consumer Protection, Germany, 2014).

[13] Law for an acceleration of asylum procedures (Federal Ministry of Justice and Consumer Protection, Germany, 2015).

offer an interesting variety with respect to economic and security conditions at origin, which lead to differences in the migration costs and the origin-specific duration of stay in Germany. These differences, in turn, are likely to affect the observed pattern of selection of asylum seekers.

1.2.2 Individual-Level Data

A Survey of Asylum Seekers in Germany

The IAB-BAMF-SOEP Refugee Sample is used to extract comprehensive information for individuals who fled their home country to seek asylum in Germany. The study surveyed recently arrived asylum seekers on a broad range of topics and included questions on their socio-economic attributes, migration experience, past and current living conditions, and labor market experience as well as attitudes about some socio-political issues (democracy, religion, and gender equality). I rely on the first wave of the survey, which was conducted in 2016 and covers 4,328 adult asylum seekers who arrived in Germany since 2013.

The sample was drawn from the Central Register of Foreigners (AZR) of the BAMF, making the survey representative of asylum seekers who arrived in Germany between January 1, 2013, and January 31, 2016, and were registered as asylum seekers by the end of June 2016 (for details on the design, methodology, and response rate of the survey, see Kroh *et al.* (2017).). Individuals with a higher likelihood of being granted refugee status in Germany at the time of the sampling (i.e., those from Afghanistan, Iraq, and Syria), women, and persons over age 30 were oversampled. Given this oversampling, I use appropriate weighting methods so that the results can be interpreted as representative of the asylum population.

Table 1.1: Composition of the recent arrival of asylum seekers in Germany

	Asylum seekers (AZR)	Asylum seekers (IAB-BAMF-SOEP)
Total	529,078	4,328
	(100.0)	(100.0)
Syria	219,673	2,181
	(41.5)	(42.6)
Afghanistan	51,709	527
	(9.8)	(13.6)
Iraq	44,138	538
	(8.3)	(8.7)
Albania, Serbia	31,104	164
	(5.9)	(3.8)
Others	182,454	918
	(34.5)	(31.3)

Notes: The first column represents cases in the register of foreigners (AZR) at BAMF, for whom the entry in Germany occurred between January 1, 2013, and January 31, 2016. The second column corresponds to asylum seekers surveyed in the IAB-BAMF-SOEP Refugee Sample. Numbers for Albania and Serbia also include Kosovo. Shares by column are reported in parentheses, and for the IAB-BAMF-SOEP Refugee Sample are weighted to be representative.
Source: Author's calculations based on Brücker *et al.* (2016) and IAB-BAMF-SOEP Refugee Sample.

Country-Specific Surveys of the Origin Population

This section proposes a brief overview of the data combined with information about asylum seekers in Germany to build the origin-specific samples required to carry out the empirical analysis (for descriptive statistics for each survey of the origin population, see online appendix section A.1). Two important comments should be made regarding the surveys conducted in the asylum source countries under focus. On the one hand, because the situation in Syria makes it difficult (if not impossible in some areas) to conduct surveys, I must rely on data collected in 2006, before the surge of asylum seekers in Germany.[14] On the other hand, the five samples are representative of the national origin population, and this holds regardless of the main purpose of each survey.[15] The representativeness is key and allows me to assess the selection of asylum seekers by avoiding potential biases that could arise if one were to compare the recent asylum seekers in Germany with a selected group at origin.

[14]The time lapse between the data collection and the asylum arrival implies that the analysis is not directly comparing asylum seekers with individuals who stayed in the source country but instead with the entire population at origin, thus diluting any pattern of self-selection.

[15]Considering Syria, the Central Microdata Catalog of the World Bank clearly indicates that the sample is "nationally representative and cover the whole of Syria" (UNICEF, 2006).

Afghanistan, Iraq, and Syria

Information about the origin population for Afghanistan comes from the Asia Foundation, which conducted the Survey of the Afghan People (SAP) yearly from 2004 to 2016. The SAP is a public opinion survey that explores social, economic, and political issues in Afghanistan. The study has gathered the opinions of more than 87,000 persons, providing an interesting portrait of individual perceptions and their evolution over time. I pool six recent waves (2011-2016) to build the sample of individuals who have stayed in Afghanistan.

Data for Iraq are drawn from the Living Standards Measurement Study (LSMS) of the World Bank. More specifically, I exploit the Household Socio-Economic Survey (HSES), which was implemented for the second time in Iraq in 2012-2013 (Organization for Statistics and Information Technology, COSIT and Kurdistan Regional Statistics Office, KRSO, 2012-2013). The main objective of the study is to provide information to measure and analyze poverty throughout the country, but it also evaluates the socio-economic situation of individuals in Iraq. The total sample size is 24,944 households, which corresponds to 176,042 individuals.

Individual-level data on the Syrian population are rarely available, particularly for recent years. I am nonetheless able to derive representative information from UNICEF's 2006 Multiple Indicator Cluster Survey (Central Bureau of Statistics, Syria, 2006). The primary goal of the survey is to deliver insights on the situation of children and women in Syria, but I can extract some relevant socio-economic characteristics for this study. UNICEF successfully interviewed 19,870 households, among which 107,365 individuals were listed. Of the full sample, I keep only 55,277 observations because of restrictions on the age of individuals (18-64); the survey involved a large number of individuals younger than age 18. The data cover 28,297 men and 26,980 women, among whom 49% and 46.1%, respectively, are aged 18-30.

Albania and Serbia

The LSMS of the World Bank is also the data source for Albania (Institute of Statistics of Albania, 2012). This multi-purpose study, which aimed to measure and evaluate the living conditions and the poverty situation in the country, was conducted several times (2002, 2003, 2004, 2005, 2008, and 2012). I use data from the last round of the survey (i.e., 2012), in which 6,671 households and a total of 25,335 individuals were interviewed. The sample contains 16,108 cases, with 8,084 men and 8,024 women, and respective shares of individuals aged 18–30 of 35.1% and 30.2%.

Finally, information for the origin population of Serbia is obtained from the European Union Statistics on Income and Living Conditions (EU-SILC). Surveys in Serbia have been administered since 2013, and the 2013-2015 waves are pooled to form the sample under focus (European Union Statistics, 2013-2015). The EU-SILC provides data on income, poverty, social exclusion, and living conditions, and it is specifically designed to be suitable for comparative statistics across European countries. At the individual level, data on the socio-economic attributes and the labor market characteristics of the interviewees are available.

1.2.3 Harmonization of Data Sources

The aforementioned data are combined to build five origin-specific samples. These samples are the result of the matching of information of the population in the source countries and the asylum seekers surveyed in Germany. The related harmonization is straightforward for several socio-demographic factors (e.g., age, gender, and marital status) given that they are commonly defined and measured across the different surveys. However, this procedure is more demanding and time-consuming for the level of education and the perceived level of insecurity in the home country. Section A.2 of the online appendix provides details on the methodology followed to link available information between the various data sources. The final number of observations in each sample is given in Table 1.2.[16] For instance, the final sample for Syria is composed of 54,014 individuals, among whom 3.8% are asylum seekers who recently arrived in Germany.

Table 1.2: Size of the respective origin-specific samples by migration status

	Afghanistan	Iraq	Syria	Albania	Serbia
Origin Population	50,406	80,722	51,968	14,829	33,395
	(99.1)	(99.4)	(96.2)	(99.7)	(99.8)
Asylum Seekers	442	485	2,046	46	43
	(0.9)	(0.6)	(3.8)	(0.3)	(0.1)
Total	50,848	81,207	54,014	14,875	33,438
	(100.0)	(100.0)	(100.0)	(100.0)	(100.0)

Note: Respective shares are reported in parentheses.
Source: Author's calculations based on SAP (2011-2016), COSIT and KRSO (2012-2013), Central Bureau of Statistics (2006), Institute of Statistics of Albania (2012), EU-SILC (2013-2015), and IAB-BAMF-SOEP Refugee Sample.

[16] The number of cases presented in Table 1.2 differs from the one reported in Table 1.1 because of sample restrictions on the age of individuals and from missing values with respect to the level of education (Table A2.1.1, online appendix).

1.3 Empirical Analysis

The combination of individual-level data for the five countries under focus paves the way for an empirical analysis of the characteristics that shape the selection of asylum seekers from the origin population. The set of variables considered in each country-specific sample is described in section A.3 of the online appendix, and weighted summary statistics are presented in section A.4. The study aims to shed light on the self-selection of asylum seekers with respect to education. Therefore, I mainly present and discuss findings related to differences in the observed level of education between asylum seekers and the home country population.

1.3.1 Descriptive Evidence

The country-specific distributions of education of the population at origin and the asylum seekers in Germany are shown in Table 1.3. The pattern of selection of asylum seekers from conflict-affected countries seems to be positive: the share who attended tertiary education is higher among asylum seekers than among their counterparts in the origin population. In the case of Syria, the figures reveal that 16.5% of asylum seekers are highly educated, compared with only 5.7% of the home country population. On the other hand, asylum seekers from Albania and Serbia appear to be negatively selected with respect to education. In the case of Serbia, only 4.9% of individuals in the origin population did not attend more than primary education, but the share peaks at 81.4% for asylum seekers.[17]

The last two rows of Table 1.3 present the statistics related to the test of independence (chi-square test) and the likelihood-ratio test for proportions. These tests are used to compare the country-specific distribution of education of the origin population with that of asylum seekers. The results indicate that the two distributions are significantly different with respect to education for all origins considered in the analysis.

[17] Using the IAB-SOEP Migration Sample (2015), I compare the education profile of Balkan asylum seekers with the one of economic migrants, who migrated to Germany mainly through family reunification. The selection pattern with respect to education of Serbian migrants is close to the one of the origin population, whereas economic migrants from Albania seem to be selected from the middle of the education distribution at origin. The statistics are available from the author upon request.

Table 1.3: Origin-specific distribution of education by migration status

	Afghanistan		Iraq		Syria		Albania		Serbia	
	Origin	Asylum	Origin	Asylum	Origin	Asylum	Origin	Asylum	Origin	Asylum
Primary or less	77.6	74.7	77.1	71.1	52.5	52.5	48.6	58.7	4.9	81.4
Secondary	20.0	19.0	18.0	17.9	41.8	31.0	36.7	39.1	77.7	18.6
Tertiary	2.4	6.3	4.9	10.9	5.7	16.5	14.8	2.2	17.4	
Total	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
Chi test	27.5***		37.7***		434.4***		6.0**		533.5***	
LR test	19.4***		28.6***		322.4***		8.8**		173.6***	

Notes: Reported figures correspond to the share of individuals in each cell. Weighted statistics can be found in the summary statistics presented in section A.4 of the online appendix. Chi-Square Test is the test of independence, and Likelihood Ratio Test is the likelihood-ratio test for proportions.
Source: Author's calculations based on SAP (2011-2016), COSIT and KRSO (2012-2013), Central Bureau of Statistics (2006), Institute of Statistics of Albania (2012), EU-SILC (2013-2015), and IAB-BAMF-SOEP Refugee Sample.

The IAB-BAMF-SOEP survey includes questions about the self-assessed relative income and economic position of asylum seekers relative to the home country population. The related statistics are introduced in Table 1.4 for each country of origin and show that 19% to 29% of asylum seekers originating from the three conflict-affected countries self-report being better-off (i.e., above average in both dimensions) compared with the origin population. This provides evidence of a positive pattern of selection with respect to their economic situation before their migration to Germany. By contrast, asylum seekers from Balkan countries come from the lower end of the income distribution, as suggested by the fact that 76% to 87% of Albanians and Serbians seeking asylum in Germany self-assess their economic position as being below the average of the home country population.

Table 1.4: Self-assessed income and economic position relative to the home country population

	Afghanistan		Iraq		Syria		Albania		Serbia	
	Income	Econ.	Income	Econ.	Income	Econ.	Income	Econ.	Income	Econ.
Below average	29.5	21.6	29.9	25.3	34.2	17.9	76.0	76.9	87.3	85.3
Average	48.8	50.1	50.9	56.0	43.3	53.0	21.0	20.7	12.7	14.7
Above average	21.7	28.4	19.2	18.7	22.5	29.1	3.0	2.4	0.0	0.0
Total	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0

Notes: Reported figures correspond to the weighted share of asylum seekers in each cell. Income refers to the following question: "If you compare your net income at that time with the income of other people in your country, how would you describe your level of net income there?" Economic Position pertains to the following question: "How would you estimate your financial situation at that time with the income of other people in your country?" For each question, five answers were available: (1) well above average, (2) above average, (3) average, (4) below average, and (5) well below average. I group (1) and (2) in the "above average' " category, while (4) and (5) are grouped in the "below average" category.
Source: Author's calculations based on IAB-BAMF-SOEP Refugee Sample.

The pattern of selection observed for the three conflict-affected countries refers to asylum seekers who were able to flee their home country and successfully reached Germany. However, only a tiny fraction of all asylum seekers managed to arrive in Europe. More than 300,000 asylum seekers in Germany come from Afghanistan, Iraq, or Syria (Table 1.1). At the European scale, this figure is high, but it is not high compared with asylum seekers hosted by neighbors of the main asylum source countries (Fig. 4; United Nations High Commissioner for Refugees (2016, p. 15)). The actual difference in the number of asylum seekers suggests that the recent asylum population in Germany is likely to represent a selected subsample of all asylum seekers who were able to leave their origin country. More specifically, it raises questions about whether the pattern of selection depends on (1) the selection of asylum seekers who left their home country or (2) the selection of asylum seekers who managed to go to Germany among those who fled Afghanistan, Iraq, or Syria. In other words, can the pattern of selection be extended for conflict-affected countries to other asylum seekers who ran away from their origin country without migrating to Germany? Based on information collected in the fourth wave of the Arab Barometer (2018), we can evaluate the distribution of education of Syrian refugees who have migrated to Jordan and Lebanon. In Jordan, the share of refugees with primary education or less is 46%, but the share of tertiary-educated refugees is 8.3%. In Lebanon, the shares of low-educated and high-educated refugees are 57.3% and 6.3%, respectively. These figures suggest that refugees in Jordan are slightly positively selected (to a lower extent than Syrian asylum seekers in Germany), whereas refugees in Lebanon are relatively similar to the education profile of the origin population (Table tab3). Notice, however, that other data sources (e.g.,

Verme *et al.* (2016)) have revealed a different pattern of selection on education for Syrian refugees who fled to Lebanon and Jordan. All things considered, the results outlined in this article are likely not to apply to the entire population of forced migrants, indicating that asylum seekers in Germany may represent a selected subsample of this population.

1.3.2 Empirical Strategy and Results

Collected information can be used to study the self-selection of asylum seekers with respect to education while other characteristics that can affect the pattern of selection are controlled for. The empirical strategy relies on the estimation of origin-specific logistic regressions with the following specification:

$$P(Y_{ij} = 1|\boldsymbol{X_{ij}}) = \frac{\exp(\boldsymbol{\beta}_j'\boldsymbol{X_{ij}})}{1 + \exp(\boldsymbol{\beta}_j'\boldsymbol{X_{ij}})} \tag{1.1}$$

where Y_{ij} is a binary indicator taking the value 1 if an individual i left her home country j to seek asylum in Germany, and 0 otherwise. $\boldsymbol{X_{ij}}$ represents individual attributes of asylum seekers: (pre-migration) level of education; age; age squared; gender; marital status; and, sporadically, perceptions about security conditions at origin, (premigration) ability to speak German, information about religion, and occupational status before migration. Notice that the set of covariates changes for each country-specific estimation because of differences in the availability of data across the surveys of the origin population.[18]

The results are presented through both the predicted probabilities of seeking asylum in Germany for each level of education and the average marginal effects, which are calculated for each individual with their observed values of covariates and then averaged across all individuals. The estimates are displayed for the three conflict-affected countries (Table 1.5) and for the two Balkan countries (Table 1.6); section A.6 in the online appendix reports the standard coefficients. The level of education of asylum seekers in Germany is evaluated with respect to the distribution of education of the origin population. In each sample, the variable is divided into three levels of education: primary or less, secondary, and tertiary education. High-educated individuals (i.e., individuals who attended tertiary education in the home country) represents the benchmark category for all countries except Serbia, which has individuals with secondary or more education as reference group, and the average marginal effects are

[18]Information used to build some control variables relates to the premigration period. Section A.5 of the online appendix demonstrates that the retrospective language proficiency in German of asylum seekers is not correlated with the time they have spent in Germany.

interpreted accordingly.

The first three columns of Table 1.5 provide evidence of a positive selection on education for asylum seekers from Afghanistan. The average marginal effects are negative and significant, indicating that asylum seekers in Germany are more likely to be highly educated than those who stayed in Afghanistan. This pattern of selection is consistent with the assumption that asylum seekers originate from a better-off subsample of the Afghan population. This, in turn, could reflect the fact that only certain individuals can afford the relatively high migration costs required to migrate to Germany. By contrast, poorer Afghans might have ended up in neighboring countries or remained at home. Moreover, the favorable selection of asylum seekers persists when the subjective perceptions about the level of insecurity (column 2) and the retrospective language proficiency in German (column 3) are added into the specification. The positive coefficient of the former suggests that asylum seekers left the country because they felt more unsafe at home than their nonasylum counterparts. This result is in line with the literature on the (macroeconomic) determinants of asylum migration, which shows that higher values of the Political Terror Scale and less individual freedom (Freedom House) push individuals out of their origin country (Hatton, 2009, 2016).

Table 1.5: Self-selection of asylum seekers from conflict-affected countries

Dependent variable: 1 if an individual has migrated, 0 otherwise Benchmark group: Tertiary education								
	Afghanistan			Iraq		Syria		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Probability of Migrating								
Level of Education								
Primary or less	0.009***	0.009***	0.009***	0.006***	0.006***	0.041***	0.041***	0.041***
	(0.001)	(0.001)	(0.001)	(0.000)	(0.000)	(0.001)	(0.001)	(0.001)
Secondary	0.006***	0.007***	0.007***	0.005***	0.006***	0.026***	0.026***	0.026***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Tertiary	0.017***	0.018***	0.014***	0.011***	0.014***	0.090***	0.087***	0.087***
	(0.003)	(0.004)	(0.003)	(0.002)	(0.002)	(0.005)	(0.005)	(0.005)
Average Marginal Effects								
Level of Education								
Primary or less	-0.008**	-0.009**	-0.005	-0.006***	-0.008***	-0.049***	-0.046***	-0.045***
	(0.003)	(0.004)	(0.003)	(0.002)	(0.002)	(0.005)	(0.005)	(0.005)
Secondary	-0.011***	-0.012***	-0.007**	-0.006***	-0.008***	-0.064***	-0.061***	-0.060***
	(0.003)	(0.004)	(0.003)	(0.002)	(0.002)	(0.005)	(0.005)	(0.005)
Age	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)	(0.000)
Male	0.004***	0.005***	0.004***	0.003***	0.003***	0.020***	0.019***	0.019***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.002)	(0.002)	(0.002)
Married	-0.010***	-0.009***	-0.009***	-0.003***	-0.003***			
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)			
Insecurity		0.018***	0.018***		0.022***			
		(0.002)	(0.002)		(0.002)			
Speaks German			0.038***					
			(0.004)					
PTS							0.094***	
							(0.005)	
FH CL								0.155***
								(0.008)
Observations	50,848	50,848	50,848	81,207	81,207	54,014	54,014	54,014
McFadden's R^2	0.028	0.074	0.097	0.018	0.103	0.038	0.146	0.164

Notes: All models are estimated using logistic regressions. McFadden's $R^2 = 1 - ln(L_M)/ln(L_0)$, with L_M, the likelihood of the estimated model and L_0, the likelihood of the model without predictors. Robust standard errors are shown in parentheses. ***, **, and * denote significance at the 1, 5, and 10 percent levels, respectively. PTS corresponds to Political Terror Scale, while FH CL is the Civil Liberties index from the Freedom House.
Source: Author's calculations based on SAP (2011–2016), COSIT and KRSO (2012–2013), Central Bureau of Statistics (2006), Political Terror Scale from Gibney et al. (2017), Freedom House (2017), and IAB-BAMP-SOEP Refugee Sample.

The selection on education of asylum seekers from Iraq is detailed in columns 4 and 5 of Table 1.5. The results reveal a positive pattern of selection of asylum seekers with respect to the origin population. More specifically, the estimates in column 5 imply that the probability of seeking asylum in Germany is 1.4% for individuals with tertiary education but only 0.6% for

individuals with a primary education or less. Consequently, the average marginal effects are negative and significant, showing that Iraqi asylum seekers have a lower likelihood of being low- and secondary-educated relative to those who remained in Iraq. Similar to Afghanistan, the observed selection could be explained by the high migration costs needed to reach Germany, and asylum seekers have been forced to flee Iraq because they feared for their own security.

The analysis focuses then on the pattern of selection of asylum seekers from Syria, and the last three columns of Table 1.5 document the relevant probabilities and average marginal effects. They all depict a positive selection of asylum seekers with respect to premigration education. The probability of migrating is 8.7% for high-educated individuals more than twice the likelihood of seeking asylum in Germany for low-educated individuals (4.1%). Differences across education groups can be shown with the average marginal effects. They highlight that the probability of claiming asylum in Germany decreases by 4.5 to 6.4 percentage points for individuals with low or secondary education, compared with high-educated individuals. Consistent with the literature, asylum seekers have been pushed out of Syria by a greater level of political terror and a worsening of civil liberties in their home country.[19]

The results show a favorable pattern of selection on education for asylum seekers originating from conflict-affected countries. By contrast, the findings are strikingly different for asylum seekers from the Balkan region. Both Albanians and Serbians who recently arrived in Germany via the asylum channel are negatively selected on education, as suggested by the estimates reported in Table 1.6. On the one hand, the probability of migrating to Germany for low- and secondary-educated individuals is positive and significant, but that for high-educated individuals from Albania is insignificant. On the other hand, the average marginal effects indicate that the differences in the probability of migrating are positive and significant for asylum seekers from Albania with low and secondary education and for low-educated asylum seekers from Serbia. Besides information on education, I am also able to take into account other characteristics in the specifications. The negative pattern of selection of Albanian asylum seekers still prevails when the retrospective ability to speak German (column 2) or religious affiliation (column 3) are included in the list of covariates. Serbian individuals seeking asylum in Germany tend to have held (in the origin country) positions as a worker rather an employee, compared with individuals who have no work experience (column 5). Controlling

[19] I do not have individual-level information on the perceived level of insecurity in Syria, which I circumvent by matching the sample with aggregate data from the Political Terror Scale (Gibney *et al.*, 2017) and the Freedom House (Freedom House, 2017) based on the date of departure for asylum seekers and on the survey date for the origin population. The latter implies values of PTS (3, from the U.S. State Department) and FH (6, for Civil Liberties) indices that clearly do not represent the situation in Syria after 2011. I solve the issue by randomly assigning a year (between 2006 and 2016) to each individual in the MICS sample and use it as the base year to merge with the aggregate variables. An overview of the outcome of this procedure is given in Table A1.3 in the online appendix.

for this variable, however, mitigates the observed negative selection of asylum seekers from Serbia. Last, the decision to migrate taken by Serbian asylum seekers is not influenced by the perceived level of insecurity in the home country (column 6). This finding supports the idea that they did not leave Serbia because they were threatened there and could potentially reinforce the fact that Serbia can be considered as a safe source country. This outcome clearly contrasts with the conclusions for asylum seekers from Afghanistan, Iraq, and Syria.

Table 1.6: Self-selection of asylum seekers from Balkan countries

Dependent variable: 1 if an individual has migrated, 0 otherwise Benchmark group: Tertiary education (Albania), Secondary or more (Serbia)						
	Albania			Serbia		
	(1)	(2)	(3)	(4)	(5)	(6)
Probability of Migrating						
Level of education						
Primary or less	0.004***	0.004***	0.004***	0.045***	0.020***	0.126***
	(0.001)	(0.001)	(0.001)	(0.008)	(0.003)	(0.018)
Secondary	0.003***	0.003***	0.003***	0.000***	0.000***	0.001***
	(0.001)	(0.001)	(0.001)	(0.000)	(0.000)	(0.000)
Tertiary	0.000	0.000	0.000			
	(0.000)	(0.000)	(0.000)			
Average Marginal Effects						
Level of education						
Primary or less	0.003***	0.004***	0.004***	0.045***	0.020***	0.125***
	(0.001)	(0.001)	(0.001)	(0.008)	(0.003)	(0.019)
Secondary	0.003***	0.003***	0.003***			
	(0.001)	(0.001)	(0.001)			
Age	-0.000**	-0.000***	-0.000**	-0.000***	-0.000***	-0.000***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Male	-0.001	-0.001	-0.001	0.001*	0.001*	0.004**
	(0.001)	(0.001)	(0.001)	(0.000)	(0.000)	(0.001)
Married	0.002	0.002	0.002	0.002***	0.001***	0.007***
	(0.002)	(0.002)	(0.002)	(0.000)	(0.000)	(0.002)
Speaks German		0.010***				
		(0.002)				
Religion						
Atheist			0.026*			
			(0.014)			
Orthodox/other			0.001			
			(0.002)			
Catholic			0.001			
			(0.002)			
Occupation						
No work					0.001***	
					(0.000)	
Worker					0.008***	
					(0.001)	
Self-employed					0.001	
					(0.001)	
Insecurity						-0.001
						(0.001)
Observations	14,875	14,875	14,875	33,438	33,438	8,400
McFadden's R^2	0.054	0.074	0.073	0.435	0.515	0.518

Notes: All models are estimated using logistic regressions. McFadden's $R^2 = 1 - ln(L_M)/ln(L_0)$, with L_M, the likelihood of the estimated model and L_0, the likelihood of the model without predictors. Robust standard errors are shown in parentheses. Muslim is the benchmark category to analyze the religious affiliation of Albanians. Employee (both with and without supervision tasks) is the reference group to interpret the occupational status in Serbia. Information about insecurity in Serbia is available only in the 2013 wave, which explains the number of observations reported in column 6.

Source: Author's calculations based on Institute of Statistics of Albania (2012), EU-SILC (2013–2015), and IAB-BAMF-SOEP Refugee Sample.

1.3.3 Robustness Checks

This section presents the results obtained from the estimation of the baseline specifications with an alternative estimator and different country-specific subsamples. These results confirm the conclusions derived for all countries except Afghanistan, for which estimates are found to be sensitive to sample selection.

Selection of Asylum Seekers Through the Analysis of Rare Events

The study is based on country-specific information, and the relative number of asylum seekers in some of the origin-specific samples (i.e., Albania and Serbia) is small (Table 1.2). Thus, the maximum likelihood estimation of the logistic model might suffer from small-sample biases. To ensure that the results are not affected by this issue, I estimate the fit models with penalized maximum likelihood estimation following the methodology proposed by Firth (1993). This procedure leads to the average marginal effects compiled in Table A7.1 in the online appendix. These effects are highly similar to the estimates obtained in the Empirical Strategy and Results section, ruling out potential biases affiliated with the low number of individuals seeking asylum in Germany contained in the dependent variable.

Selection With Respect to Urban/Rural Origin Population

The benchmark analysis does not control for potential information about the place of departure of asylum seekers in the origin country. However, the positive pattern of selection of asylum seekers from conflict-affected countries might be driven by the fact that they fled urban areas, which could on average host more high-educated individuals. The reverse occurs for Balkan countries, and the negative selection might be the consequence of asylum seekers originating from rural areas, where the average level of education is likely to be lower than in cities.

This question cannot be directly evaluated because the IAB-BAMF-SOEP Refugee Sample does not include data on the starting point of migration from the home country to Germany. On the other hand, the origin-specific surveys allow me to determine whether individuals are located in an urban or a rural area. One way to address the lack of information relative to asylum seekers is to assume that all those who are from conflict-affected (Balkan) countries come from urban (rural) locations in their source country. The empirical study is then replicated to check whether the observed pattern of selection is the result of the selection of asylum seekers with respect to the urban/rural composition of the origin population. The related average

marginal effects are presented in Table A7.2 in the online appendix.

The positive (negative) pattern of selection documented for asylum seekers from Iraq and Syria (Albania and Serbia) is not altered when their level of education is compared with that of individuals who live in urban areas. By contrast, Afghan asylum seekers are now negatively selected with respect to the urban origin population, and this raises questions on the robustness of the results reported in Table 1.5. However, the conservative assumption that all asylum seekers from Afghanistan who fled toward Germany are from urban locations might not be consistent with the urban/rural composition that prevails in the country. With only 25% (United Nations, 2018) of the Afghan population living in urban areas in 2015, it is likely that some asylum seekers originate from rural areas and would better be compared with the rural population to correctly analyze the robustness of the pattern of selection on education.

Self-selection of Male Asylum Seekers

In the migration literature, the self-selection of migrants has been mainly studied among men or by dividing the sample into men and women. The origin-specific samples used in the baseline analysis did not make this distinction. The arrival of female asylum seekers in Germany might follow the initial departure of men, and this mechanism could distort the results obtained previously. Therefore, I reestimate the various specifications presented earlier by considering only male asylum seekers. The resulting average marginal effects, displayed in Table A7.3 in the online appendix, support the findings depicted with the full-sample estimates for all countries except Afghanistan, for which the education profile of asylum seekers with respect to the origin population appears to be sensitive to choices made about sample selection.

Self-selection of Family-Tied Asylum Seekers

All asylum seekers surveyed in the IAB-BAMF-SOEP Refugee Sample have been considered in the benchmark estimations. However, the probability of seeking asylum in Germany for some individuals might depend on their family ties, so the chosen destination could also be the residence of at least one of their relatives. Although no explicit question on family reunification has been asked of asylum seekers, those who are susceptible to migrating to Germany through the family channel are identified as having at least one of the following two characteristics: (1) they left the origin country because some of the family members moved abroad, and/or (2) they chose Germany because they have relatives already living there. Then, the analysis

is reproduced with samples that involve only those individuals with potential family links in Germany to check for potential differences in the pattern of selection on education of family-tied asylum seekers.

The estimates for education are provided in Table A7.4 in the online appendix. All specifications have been replicated, but the baseline group has been switched from tertiary education to secondary education or more for Balkan countries because of constraints on the number of asylum seekers in the dependent variable. The results obtained with the origin-specific subsamples reveal that a positive selection on education still arises for asylum seekers from Iraq and Syria, whereas Serbian asylum seekers are negatively selected on education with respect to the origin population. Moreover, family-tied asylum seekers from Albania and Afghanistan are comparable in terms of education to their nonasylum counterparts.

1.4 Interpretation of the Empirical Results

This section provides arguments and supportive evidence to understand the findings obtained in the econometric analysis. The main goal is to interpret the observed difference in the pattern of selection of asylum seekers from the two groups of origins.

Economic and security conditions differ across origin countries. Individuals from Afghanistan, Iraq, or Syria are likely to be threatened or persecuted at home. The individuals in Albania and Serbia, however, are not considered to be endangered, which explains why these two countries were included in the list of safe source countries. This difference has consequences on the probability of being granted the refugee status in Germany. Indeed, the recognition rate is relatively high for asylum seekers from conflict-affected countries but is extremely low for asylum seekers from the Balkan region. In 2015, 72.8% of asylum applications from Afghanistan were accepted by Germany, and the acceptance rates for Iraq and Syria were 98.3% and 97.7%, respectively. However, Germany approved only 0.2% and 0.1% of asylum claims from, respectively, Albania and Serbia (Eurostat, 2018).

Origin countries also differ in terms of migration costs. More specifically, the median total cost of migration (i.e., the sum of the costs associated with transport, accommodation, and smuggling) is 2,015 euros for conflict-affected countries but is only 280 euros for Balkan countries. Moreover, the median time to reach Germany from conflict-affected countries is 23

days, compared with only 2 days for Balkan countries.[20] The lower figures for asylum seekers from Albania and Serbia highlight that the door was rather open between the Balkan region and Germany. Higher metrics for asylum seekers from Afghanistan, Iraq, and Syria imply that it was more difficult for them to reach the host country.

Both the migration costs and the origin-specific asylum recognition rate are consistent with differences in the pattern of selection of asylum seekers who recently arrived in Germany. The higher costs and acceptance rates faced by asylum seekers from conflict-affected countries would lead to a more positive selection on education. On the one hand, savings are likely to be positively correlated with skills, such that the liquidity constraints on the decision to migrate would determine a positive pattern of selection. On the other hand, asylum seekers from unsafe areas have a higher recognition rate, which allows them to expect to stay longer (or even permanently) in Germany. Because transferring human capital across borders takes time, the returns to education at destination is an increasing function of the time spent there (Dustmann and Glitz, 2011). Even though asylum seekers might enjoy limited returns to education on the German labor market in the early stages of their stay, the time horizon could be sufficiently long for the income gains from migration to become an increasing function of education.

Asylum seekers from Balkan countries encounter lower costs of migration and are able to enter the destination without a visa, enhancing the attractiveness of migrating to Germany. However, Germany considers Albania and Serbia to be safe. This implies that the probability of acceptance of asylum claims is close to zero and that, after an asylum claim is denied, asylum seekers can either leave the host country (voluntarily or by force) or stay in Germany as undocumented migrants.[21] In principle, this should compel them to remain only temporarily in Germany; in fact, the limited legal time refers to the period required to process the asylum applications. Misusing the asylum channel as a legal temporary migration scheme might exclusively be attractive for low-educated individuals, such that the income gains from migration are a decreasing function of education. If the asylum seekers decide to remain as undocumented migrants, they could stay longer in Germany but would be able to work only in the informal sector, where the returns to education are lower than in the formal labor

[20] The median cost of migration and the median time to reach Germany come directly from the IAB-BAMF-SOEP Refugee Sample. Information about the latter is obtained from answers to the following question: "How many days did it take to travel from your country of birth to Germany?"

[21] Voluntary returns from Germany are implemented via the REAG/GARP programme run by the International Organisation for Migration. In 2016, Albania (31.3%) and Serbia (11.4%) were the two most important countries of origin involved in voluntary returns (European Migration Network/Federal Office for Migration and Refugees, 2016, Fig.5, p 60).

market.[22] This would also coincide with a negative pattern of selection of asylum seekers with respect to education.

The perspective from being able to stay only temporarily in the receiving country raises questions about whether the time taken to process the applications could be beneficial for asylum seekers from Albania or Serbia. First, they are protected from deportation to their home country during the claim processing time. Because of the insufficient capacity of German authorities to process the surge of asylum applications in 2015, the number of pending cases increased sharply, mechanically increasing the time needed to process these claims.[23] However, priorities given to process the claims from some origin countries might have resulted in differences in the expected processing time across countries.

To explore this idea, I compute the origin-specific expected processing time of asylum applications in Germany.[24] The average time to determine whether the request would be accepted in 2015 was high for Afghanistan (25 months) and Iraq (15.5 months), whereas Syrian asylum claims were processed more quickly (4.5 months). The figures for Albania (9 months) and Serbia (15.5 months) indicate that the expected time to process the asylum claims was high when Balkan asylum seekers arrived in Germany. Recall that this metric corresponds to the temporary legal period whereby asylum seekers from the Balkan region can stay in the host country. This, in turn, implies that the expected duration of stay was substantial upon arrival in Germany, which might have fostered low-educated individuals to claim asylum there.[25] Second, individuals could seek asylum with the aim of working in Germany, regardless of whether the job is in the formal or informal labor market. This motive could have been strengthened by origin-specific network ties that result from past (legal or illegal) migration to Germany. However, this potential channel is likely to be at play when the size of the network is relatively large. On the one hand, legal migration is proxied with the stock of valid residence permits at the end of the year in Germany (Eurostat, 2017a). Among the five origin countries, 60%

[22] Recently, Borjas (2017) documented that the rate of returns to schooling (i.e., the coefficient of years of schooling in the log wage regression) is lower for undocumented immigrants in the United States than for legal immigrants or native workers.

[23] In 2015, the Federal Office for Migration and Refugees (BAMF) took on average 5.2 months to determine whether the request would be accepted; the waiting period was 7.1 months in 2016, increasing to 8.1 months in the fourth quarter of 2016 (Federal Government, Germany, 2017).

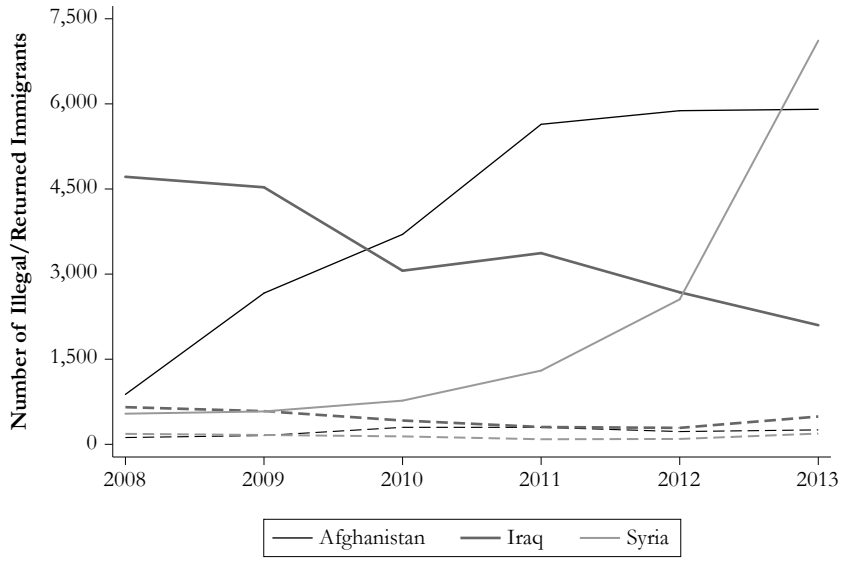
[24] I calculate the expected processing time by comparing the number of origin-specific pending applications at the end of month $(t-1)$ with the number of months (from 1 to 36), over which it is necessary to cumulate applications (from $t-1$) to reach the number of pending applications. I then take the average of the generated variable for each year and each origin country (Eurostat, 2017b,c).

[25] In principle, I could use the date of arrival in Germany to check whether asylum seekers from Balkan countries, who arrived when the time to process the applications and the recognition rate were high, are less negatively selected than those who arrived when the decision time and acceptance rates were lower. However, the small number of observations prevents an econometric analysis of the mechanism. Descriptive evidence supporting the aforementioned assumption is available from the author upon request.

of the residence permits that were valid in 2013 (6% of all residence permits) were held by immigrants from Serbia. On the other hand, illegal migration is proxied through the evolution of the number of found illegal immigrants in Germany and the number of individuals who returned to their origin country after they received an order to leave. The number of undocumented immigrants from Afghanistan increased over time, but the reverse occurred for illegal immigrants from Iraq. Following the onset of the civil conflict, the number of illegal immigrants from Syria rose sharply.

Compared with the aforementioned figures, the number of deported immigrants is relatively constant and small, mainly because of the security conditions that prevail in the home country (Fig. 1.3). The two Balkan countries are strikingly different with respect to illegal migration (Fig. 1.4). The figures are, on average, small for Albania: 750 found illegal immigrants and 250 returned individuals. They are higher for Serbia and stand at, respectively, 4,000 and 2,700. The opportunity to come and stay illegally in Germany could have been enhanced by the lower migration costs faced to reach the host country from the Balkan region and by the fact that Albanians and Serbians can legally enter the Schengen area without a visa.

Figure 1.3: Illegal immigrants and returned individuals from conflict-affected countries.

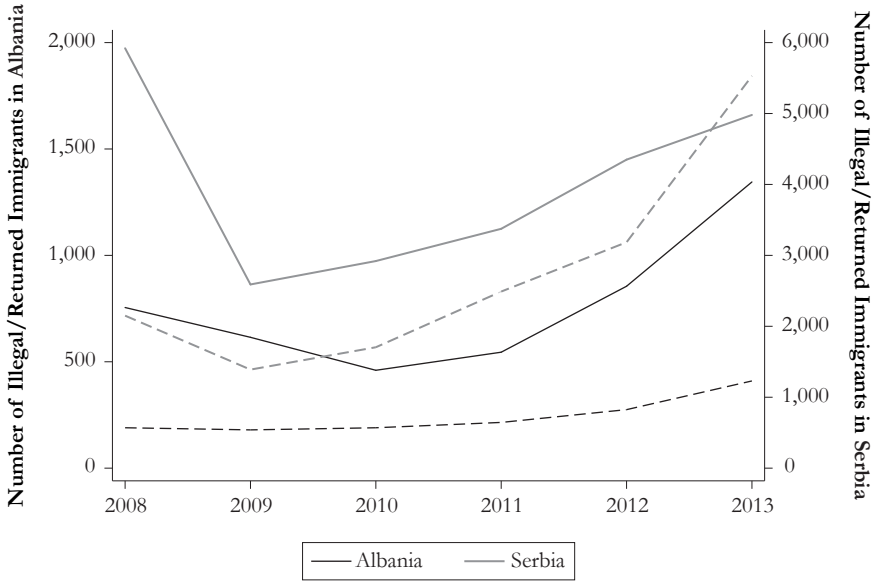


Notes: The solid lines represent illegal immigrants; the dashed lines correspond to returned individuals.
Source: Author's elaboration based on Eurostat (2017d,e).

These descriptive statistics outline that past migration from Serbia to Germany and the related size of the network may have facilitated the arrival of Serbian asylum seekers and their entry

into the German labor market. In addition, several stepwise changes in the German asylum policy improved conditions for accessing the labor market. The adjustments led to a reduction in waiting time to request a permit to work from 12 months to 9 months (from September 2013 until October 2014; Federal Ministry of Justice and Consumer Protection, Germany (2013)), and further to 3 months (from November 2014),[26] provided that asylum seekers from the Balkan region were registered before September 2015; from this date onward, they were no longer allowed to work during the application processing time.[27] In the IAB-BAMF-SOEP Refugee Sample, the last condition is fulfilled for most asylum seekers from Albania and Serbia (86%). At the end of the waiting period, compliance with various labor market regulations is assessed, so that asylum seekers can effectively be allowed to work in Germany. Altogether, Albanian and Serbian asylum seekers might have been attracted by the German labor market, but they would have encountered different hurdles when trying to find a job (at least, in the formal economy).

Figure 1.4: Illegal immigrants and returned individuals from Balkan countries



Notes: The solid lines represent illegal immigrants; the dashed lines correspond to returned individuals. The left (right) axis refers to figures for Albania (Serbia)

Source: Author's elaboration based onEurostat (2017d,e).

[26] Law for the classification of additional states as safe countries of origin and to facilitate the access to the labor market for asylum seekers and tolerated foreigners (Federal Ministry of Justice and Consumer Protection, Germany, 2014).

[27] Law for the acceleration of asylum procedures (Federal Ministry of Justice and Consumer Protection, Germany, 2015).

Finally, claiming refugee status directly benefits asylum seekers through the allocation of welfare provisions during the time required to review their application. The amounts depend on a range of characteristics, such as whether asylum seekers are living in public or private housing and the composition of one's own family. For instance, if they are hosted in a government facility, two adult persons living in the same household as their partner each receive 129 euros per month, but the amount is 194 euros if they reside in a private dwelling (Federal Ministry of Justice and Consumer Protection, Germany, 2019). Given the low migration costs involved in migrating to Germany from the Balkan region and the longer time needed to evaluate the asylum claims, it might then have been economically worthwhile for Albanians and Serbians to seek asylum in Germany and receive welfare benefits until they were notified about their application.

1.5 Conclusion

The distinction between asylum seekers and economic migrants is often made in the public debate based on the factors fostering the decision to migrate for each group. Unlike the determinants of economic migration, the drivers behind who is able to make her way to another country from the main asylum source countries have been rarely explored. The few studies that have focused on the self-selection of individuals in the context of forced migration are either related to past episodes of migration (Birgier *et al.*, 2016) or based on data that are imperfectly representative of the origin population (Buber-Ennser *et al.*, 2016) or of the asylum population at destination (Lange and Pfeiffer, 2018). This study contributes to the literature through the use of individual-level and representative information for both asylum seekers in Germany and the population at origin. Specifically, the analysis is built on original data about asylum seekers in Germany complemented with surveys conducted in five key source countries, which offers an interesting variety of economic and security conditions at origin. The pattern of selection of asylum seekers from the origin population is examined with respect to education. The country-specific investigations provide evidence of positive selection on education for asylum seekers who fled Iraq and Syria, and shows mixed evidence for asylum seekers from Afghanistan. By contrast, individuals seeking asylum in Germany from Albania and Serbia are negatively selected relative to the home country population.

These patterns of selection on education are interpreted using differences in the expected duration of stay in Germany and in migration costs faced by asylum seekers when migrating to Germany. Specifically, I describe the decision of Albanians and Serbians to seek asylum in Germany (where their claims are almost certainly rejected) through the high expected

processing time of their applications, which corresponds to the temporary legal period of stay in the host country. Lower expected duration of stay and migration costs may have triggered the observed negative selection on education of asylum seekers from the Balkan region, whereas a higher time horizon in Germany and migration costs may have driven the positive selection of asylum seekers from conflict-affected countries. Moreover, some network ties in Germany might have facilitated the arrival of asylum seekers from Serbia. This work suggests that the set of factors -especially the premigration socio-economic status- influencing the decision of asylum seekers to migrate do not involve a sharp discontinuity with the determinants associated with migration decisions of economic migrants.

Bibliography

Aksoy, C. G., and P. Poutvaara (2019): "Refugees' Self-Selection into Europe: Who Migrates Where?," ifo Working Paper Series 289, ifo Institute - Leibniz Institute for Economic Research at the University of Munich.

Arab Barometer (2018): "Arab Barometer Wave IV (2016-2017)," Available from https://www.arabbarometer.org/survey-data/data-downloads/.

Belot, M. V. K., and T. J. Hatton (2012): "Immigrant Selection in the OECD," *Scandinavian Journal of Economics*, 114(4), 1105–1128.

Bertoli, S., and J. Fernández-Huertas Moraga (2015): "The Size of the Cliff at the Border," *Regional Science and Urban Economics*, 51, 1–6.

Bertoli, S., J. Fernández-Huertas Moraga, and F. Ortega (2013): "Crossing the Border: Self-Selection, Earnings and Individual Migration Decisions," *Journal of Development Economics*, 101, 75–91.

Birgier, D. P., C. Lundh, Y. Haberfeld, and E. Elldér (2016): "Self-Selection and Host Country Context in the Economic Assimilation of Political Refugees in the United States, Sweden, and Israel," *International Migration Review*, (52), 524–558.

BMI (2017): "280,000 asylum seekers in 2016," Press Release (January 11, 2017), Retrieved from https://www.bmi.bund.de/SharedDocs/pressemitteilungen/DE/2017/01/asylantraege-2016.html.

Borjas, G. J. (1987): "Self-Selection and the Earnings of Immigrants," *American Economic Review*, 77(4), 531–553.

——— (1991): "Immigration and Self-Selection," in *Immigration, Trade, and the Labor Market*, pp. 29–76. University of Chicago Press.

——— (2017): "The Earnings of Undocumented Immigrants," Working Paper 23236, National Bureau of Economic Research.

Borjas, G. J., and J. Monras (2017): "The Labour Market Consequences of Refugee Supply Shocks," *Economic Policy*, 32(91), 361–413.

Brücker, H., and C. Defoort (2009): "Inequality and the Self-Selection of International Migrants: Theory and New Evidence," *International Journal of Manpower*, 30(7), 742–764.

Brücker, H., N. Rother, and J. Schupp (2016): "IAB-BAMF-SOEP-Befragung von Geflüchteten: Überblick und erste Ergebnisse," Politikberatung Kompakt 116, DIW Berlin, [IAB-BAMF-SOEP Survey of Refugees: Overview and first Results].

Buber-Ennser, I., J. Kohlenberger, B. Rengs, Z. Al Zalak, A. Goujon, E. Striessnig, M. Potančoková, R. Gisser, M. R. Testa, and W. Lutz (2016): "Human Capital, Values, and Attitudes of Persons Seeking Refuge in Austria in 2015," *PLOS ONE*, 11(9), 1–29.

Card, D. (1999): "The Causal Effect of Education on Earnings," in *Handbook of Labor Economics*, ed. by O. C. Ashenfelter, and D. Card, vol. 3, pp. 1801–1863. Elsevier.

Central Bureau of Statistics, Syria (2006): Demascus, Syria; and New York, NY: Central Bureau of Statistics, Ministry of Health; and Strategic Information Section, Division of Policy and Planning, UNICEF. Retrieved from http://mics.unicef.org/surveys.

Chen, J. J., K. Kosec, and V. Mueller (2017): "Temporary and Permanent Migrant Selection: Theory and Evidence of Ability-Search Cost Dynamics," IZA Discussion Paper No. 9639 (Revised version).

Chiquiar, D., and G. Hanson (2005): "International Migration, Self-Selection, and the Distribution of Wages: Evidence from Mexico and the United States," *Journal of Political Economy*, 113(2), 239–281.

Chiswick, B. R. (1999): "Are Immigrants Favorably Self-Selected?," *American Economic Review*, 89(2), 181–185.

Council of the European Union (2005): "Asylum Procedures Directive, Directive 2005/85/EC - Annex II on the *Designation of safe countries of origin for the purposes of Articles 29 and 30(1)*," Retrieved from https://www.asylumlawdatabase.eu/en/content/en-asylum-procedures-directive-directive-200585ec-1-december-2005.

Dustmann, C., and A. Glitz (2011): "Migration and Education," in *Handbook of The Economics of Education*, ed. by S. M. Eric A. Hanushek, and L. Woessmann, vol. 4, pp. 327–439. North-Holland.

European Council (2016): "EU-Turkey Statement," [Press release]; Retrieved from https://www.consilium.europa.eu/en/press/press-releases/2016/03/18/eu-turkey-statement/.

European Migration Network/Federal Office for Migration and Refugees (2016): "Migration, Integration, Asylum: Political Developments in Germany 2015," (Annual Policy Report).

European Union and Islamic Republic of Afghanistan (2016): "Joint way forward on migration issues between Afghanistan and the EU," Retrieved from https://eeas.europa.eu/sites/eeas/files/eu_afghanistan_joint_way_forward_on_migration_issues.pdf.

European Union Statistics (2013-2015): "Income and Living Conditions-Serbia," Information available from https://ec.europa.eu/eurostat/web/income-and-livingconditions/overview.

Eurostat (2016): "Asylum in the EU Member States - Record number of over 1.2 million first time asylum seekers registered in 2015," News Release 44/2016 (March 4, 2016), Retrieved from http://ec.europa.eu/eurostat/documents/2995521/7203832/3-04032016-AP-EN.pdf.

——— (2017a): "All valid permits by reason, length of validity and citizenship on 31 December of each year," [migr_resvalid]; Retrieved from http://appsso.eurostat.ec.europa.eu/nui/show.do?dataset=migr_resvalid&lang=en.

——— (2017b): "Asylum and first time asylum applicants by citizenship, age and sex-Monthly data," [migr_asyappctzm]; Retrieved from http://appsso.eurostat.ec.europa.eu/nui/show.do?dataset=migr_asyappctzm&lang=en.

——— (2017c): "Persons subject of asylum applications pending at the end of the month by citizenship, age and sex-Monthly data," [migr_asypenctzm]; Retrieved from https://appsso.eurostat.ec.europa.eu/nui/show.do?dataset=migr_asypenctzm&lang=en.

——— (2017d): "Third country nationals found to be illegally present-Annual data," [migr_eipre]; Retrieved from https://appsso.eurostat.ec.europa.eu/nui/show.do?dataset=migr_eipre&lang=en.

——— (2017e): "Third country nationals returned following an order to leave-Annual data," [migr_eirtn]; Retrieved from http://appsso.eurostat.ec.europa.eu/nui/show.do?dataset=migr_eirtn&lang=en.

——— (2018): "First instance decisions on applications by citizenship, age and sex-Annual aggregated data," [migr_asydcfsta]; Retrieved from https://appsso.eurostat.ec.europa.eu/nui/show.do?dataset=migr_asydcfsta&lang=en.

Federal Government, Germany (2017): "Response to parliamentary question by the left: Antwortder Bundesregierung auf die Kleine Anfrage der Abgeordneten Ulla Jelpke, Frank Tempel, Sevim Dağdelen, weiterer Abgeordneter und der Fraktion DIE LINKE," (Drucksache 18/11262, pp. 12-13). Retrieved from http://dipbt.bundestag.de/doc/btd/18/112/1811262.pdf.

Federal Ministry of Justice and Consumer Protection, Germany (2013): *Bundesgesetzblatt (BGBl)* [Federal Law Gazette] (Vol. I, No. 54, p. 3480). Cologne, Germany: Bundesanzeiger Verlagsgesellschaft. Retrieved from https://www.bgbl.de/xaver/bgbl/start.xav?start=//*%5B@attr_id=%27bgbl113s3474.pdf%27%5D#__bgbl.

——— (2014): "Law on classification of further states as safe countries of origin and on the facilitation of access to the labour market for asylum seekers and tolerated foreigners," In *Bundesgesetzblatt (BGBl)* [Federal Law Gazette] (Vol. I, No. 49, p. 1649). Cologne, Germany: Bundesanzeiger Verlagsgesellschaft. Retrieved from https://www.bgbl.de/xaver/bgbl/

start.xav?start=%2F%2F*%5B%40attr_id%3D%27bgbl114049.pdf%27%5D#__bgbl__%2F%2F*%5B%40attr_id%3D%27bgbl114049.pdf%27%5D__1583492944831.

——— (2015): "Law for an acceleration of asylum procedures," In *Bundesgesetzblatt (BGBl)* [Federal Law Gazette] (Vol. I, No. 23, pp. 1722-1723). Cologne, Germany: Bundesanzeiger Verlagsgesellschaft. Retrieved from https://www.bgbl.de/xaver/bgbl/start.xav?startbk=Bundesanzeiger_BGBl&start=//*%5B@attr_id=%27bgbl115s1722.pdf%27%5D#__bgbl__%2F%2F*%5B%40attr_id%3D%27bgbl115s1722.pdf%27%5D__1583495046572.

——— (2019): *Asylum Seeker Benefits Act (AsylbLG), §3a Requirement rates for basic services*; Retrieved from https://www.gesetze-im-internet.de/asylblg/__3a.html.

Fernández-Huertas Moraga, J. (2011): "New Evidence on Emigrant Selection," *Review of Economics and Statistics*, 93(1), 72–96.

Firth, D. (1993): "Bias Reduction of Maximum Likelihood Estimates," *Biometrika*, 80(1), 27–38.

Freedom House (2017): "Freedom in the world 2017," Retrieved from https://freedomhouse.org/report/freedom-world.

Gibney, M., L. Cornett, R. Wood, P. Haschke, D. Arnon, A. Pisanò, and G. Barrett (2017): "Political Terror Scale 1976-2017," Retrieved from http://www.politicalterrorscale.org.

Grogger, J., and G. H. Hanson (2011): "Income Maximization and the Selection and Sorting of International Migrants," *Journal of Development Economics*, 95(1), 42–57.

Hanson, G., and C. McIntosh (2016): "Is the Mediterranean the New Rio Grande? US and EU Immigration Pressures in the Long Run," *Journal of Economic Perspectives*, 30(4), 57–82.

Hatton, T. J. (2009): "The Rise and Fall of Asylum: What Happened and Why?," *The Economic Journal*, 119(535), F183–F213.

——— (2016): "Refugees, Asylum Seekers, and Policy in OECD Countries," *American Economic Review*, 106(5), 441–45.

Institute of Statistics of Albania (2012): "Albania Living Standards Measurement Survey 2012," (Ref. ALB_2012_LSMS_v01_M) [Data set]; Retrieved from https://microdata.worldbank.org/index.php/catalog/1970.

Kaestner, R., and O. Malamud (2014): "Self-Selection and International Migration: New Evidence from Mexico," *Review of Economics and Statistics*, 96(1), 78–91.

Kroh, M., S. Kühne, J. Jacobsen, M. Siegert, and R. Siegers (2017): *Sampling, Nonresponse, and Integrated Weighting of the 2016 IAB-BAMF-SOEP Survey of Refugees (M3/M4) - Revised Version*, vol. 477 : Series C, Data Documentations. Berlin, Germany: The German Socio-Economic Panel Study at DIW Berlin.

Lange, M., and F. Pfeiffer (2018): "The Human Capital Selection of Young Males Seeking Asylum in Germany," ZEW Discussion Paper 18-040, ZEW Mannheim.

Mayda, A. M. (2010): "International Migration: A Panel Data Analysis of The Determinants of Bilateral Flows," *Journal of Population Economics*, 23(4), 1249–1274.

McKenzie, D., and H. Rapoport (2010): "Self-Selection Patterns in Mexico-US Migration: The Role of Migration Networks," *Review of Economics and Statistics*, 92(4), 811–821.

Meisner, M. (2015): "Ad campaign in Albania: German embassy warns against *economic asylum*," [Press release]; Retrieved from https://www.euractiv.com/section/global-europe/news/ad-campaign-inalbania-german-embassy-warns-against-economic-asylum/.

Neumayer, E. (2004): "Asylum Destination Choice: What Makes Some West European Countries More Attractive Than Others?," *European Union Politics*, 5(2), 155–180.

——— (2005): "Bogus Refugees? The Determinants of Asylum Migration to Western Europe," *International Studies Quarterly*, 49(3), 389–410.

Organization for Statistics and Information Technology, COSIT and Kurdistan Regional Statistics Office, KRSO (2012-2013): "Iraq Household Socio-Economic Survey 2012-2013," (Ref. IRQ_2012_IHSES_v02_M) [Data set]; Retrieved from https://microdata.worldbank.org/index.php/catalog/2334.

Roy, A. D. (1951): "Some Thoughts on the Distribution of Earnings," *Oxford Economic Papers*, 3(2), 135–146.

Thielemann, E. (2006): "The Effectiveness of Governments' Attempts to Control Unwanted Migration," in *Immigration and the Transformation of Europe*, ed. by C. Parsons, and T. Smeeding. Cambridge University Press.

UNICEF (2006): "Multiple Indicator Cluster Survey 2006: Study notes," Retrieved from http://microdata.worldbank.org/index.php/catalog/53/study-description.

United Nations (2018): "World urbanization prospects: 2018 revision," New York, NY: United Nations Population Division. Retrieved from https://data.worldbank.org/indicator/SP.URB.TOTL.IN.ZS?locations=AF.

United Nations High Commissioner for Refugees (2016): "Global trends: Forced displacement in 2016," (Report). Geneva, Switzerland: UNHCR.

Verme, P., C. Gigliarano, C. Wieser, K. Hedlund, M. Petzoldt, and M. Santacroce (2016): *The Welfare of Syrian Refugees*, no. 23228 in World Bank Publications. The World Bank.

Appendix

A.1 Statistics of origin-specific surveys

Table A1.1: Number of cases for Afghanistan by wave, gender and age

	Men		Women		
Wave	Up to 30	Over 30	Up to 30	Over 30	Total
2011	1,475	2,028	1,309	1,344	6,156
	(12.7)	(13.6)	(10.8)	(11.3)	(12.2)
2012	2,046	2,643	1,570	1,484	7,743
	(17.6)	(17.7)	(13.0)	(12.5)	(15.3)
2013	2,230	2,896	1,509	1,592	8,227
	(19.2)	(19.4)	(12.5)	(13.4)	(16.3)
2014	1,701	2,219	2,376	2,203	8,499
	(14.6)	(14.8)	(19.7)	(18.5)	(16.8)
2015	1,811	2,205	2,345	2,342	8,703
	(15.6)	(14.7)	(19.3)	(19.7)	(17.2)
2016	2,364	2,970	2,970	2,953	11,257
	(20.3)	(19.9)	(24.6)	(24.8)	(22.3)
Total	11,627	14,961	12,079	11,918	50,585
	[23.0]	[29.6]	[23.9]	[23.6]	(100.0)

Notes: Percentages by column are displayed in parentheses, whereas shares by line are given in brackets. Individuals aged below 18 and above 64 are not included.
Source: Author's elaboration based on SAP (2011-16).

Table A1.2: Number of cases for Iraq by year, gender and age

	Men		Women		
Year	Up to 30	Over 30	Up to 30	Over 30	Total
2012	19,800	23,005	18,772	23,020	84,597
	(96.1)	(95.9)	(96.3)	(96.1)	(96.1)
2013	798	975	724	946	3,443
	(3.9)	(4.1)	(3.7)	(4.0)	(3.9)
Total	20,598	23,980	19,496	23,966	88,040
	[23.4]	[27.2]	[22.1]	[27.2]	(100.0)

Notes: Percentages by column are displayed in parentheses, whereas shares by line are given in brackets. Individuals aged below 18 and above 64 are not included.
Source: Author's elaboration based on COSIT and KRSO (2012–2013).

Table A1.3: Adjustments of PTS and FH indices

	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	Mean | Total
PTS (US State Department)	3	3	4	4	4	5	5	5	5	5	5	4.36
FH (Civil Liberties)	6	6	6	6	6	6	7	7	7	7	7	6.55
Number of cases	5,142	4,931	4,918	5,023	5,070	5,005	4,990	5,079	5,090	5,023	5,006	55,277

Note: Number of cases corresponds to individuals in Central Bureau of Statistics (2006).
Source: Author's elaboration based on Gibney et al. (2017) and Freedom House (2017).

Table A1.4: Number of cases for Serbia by wave, gender and age

	Men		Women		
Wave	Up to 30	Over 30	Up to 30	Over 30	Total
2013	1,614	4,778	1,505	4,982	12,879
	(34.1)	(34.9)	(34.6)	(35.4)	(35.0)
2014	1,598	4,546	1,448	4,662	12,254
	(33.7)	(33.2)	(33.3)	(33.2)	(33.3)
2015	1,524	4,361	1,394	4,411	11,690
	(32.2)	(31.9)	(32.1)	(31.4)	(31.7)
Total	4,736	13,685	4,347	14,055	36,823
	[12.9]	[37.2]	[11.8]	[38.2]	(100.0)

Notes: Percentages by column are displayed in parentheses, whereas shares by line are given in brackets. Individuals aged below 18 and above 64 are not included.
Source: Author's elaboration based on EU-SILC (2013-15).

A.2 Data harmonisation

A.2.1 Level of education

The empirical analysis focuses on the selection of asylum seekers from the origin population with respect to education[28]. Two steps are implemented to combine available information on the educational attainment of individuals[29]. First, the answers about the level of education in each questionnaire are divided into six different categories, i.e. no formal education, primary, lower secondary, upper secondary education, vocational training and university. Tables A2.1.1 and A2.1.2 detail the procedure that has been followed to assign answers about education to each group for the asylum seekers and the origin population, respectively.

Nevertheless, the six categories are not included in every survey, e.g. studies carried out in Afghanistan and Syria do not contain information on vocational training. Therefore, the initial binary indicators are grouped, so that the final variable of interest is composed of three levels: (*i*) Primary education or less, which refers to cases without education and with primary education (*ii*) Secondary education, which contains individuals who attended lower, upper secondary education and vocational training and (*iii*) Tertiary education, which encloses

[28] When discussing about the level of education of asylum seekers hosted by Germany, we only consider the education obtained in the origin country. The paper does not account for further education which might have been attended in the host country.

[29] The analysis does not include individuals, who were attending education in the origin country at the time the surveys were carried out.

those who went to university.

Table A2.1.1: Level of education of asylum seekers

	Question	Answers	Frequency	Code
School certificate	In which year did you last attended a general school?	I did not attend school	13.11	A
	Have you attended school in another country than Germany?	No		A
	With what kind of graduation you finished school there?	Left school with no qualifications	23.57	B
		Middle school leaving certificate	20.87	C
		Practical-based further education certificate	8.08	D
		General-based further education certificate	23.44	E
		Certificate from a different school	2.47	F
		No information	8.46	
Vocational training	Were you in a country other than Germany in a vocational training or have you studied in another country than Germany?	No	74.90	
	What kind of vocational training was that?	(Long) training in a company		
	How did you finish this training?	Aborted prematurely	0.33	1
		Completed without certificate	0.93	2
		Completed with certificate	2.73	3
		Attended a vocational school		
		Aborted prematurely	0.19	4
		Completed without certificate	0.21	5
		Completed with certificate	1.83	6
		Other training		
		Aborted prematurely	0.06	7
		Completed without certificate	0.31	8
		Completed with certificate	0.48	9
		University with practical or theoretical orientation visited / Graduate degree		
		Aborted prematurely	4.49	10
		Completed without certificate	0.99	11
		Completed with certificate	10.69	12
		No information	1.86	
Level of education		No education	36.45	A,B
		Primary education	17.24	C
		Lower secondary	17.66	(D,E,F) + (1,2,4,5,7,8)
		Upper secondary	5.04	(C,D,E,F) + (10,11)
		Vocational	5.04	(3,6,9)
		University	10.69	12
		No information	7.87	

Notes: Statistics reported in the fourth column are associated to the full sample of asylum seekers and are therefore not representative of the five selected origin countries. Figures are weighted to be representative of the (recent) asylum population in Germany.
Source: Author's elaboration based on IAB-BAMF-SOEP Refugee Sample.

Table A2.1.2: Level of education in the origin country

Origin	Question	Answers	Frequency	Categories
Afghanistan	What is the highest level of school you completed?	Never went to school	55.87	No education
		Islamic education at Madrassa	0.52	No education
		Informal schooling at home or at a literacy class	1.88	No education
		Primary school, incomplete (classes 1 to 5)	8.22	Primary
		Primary school, complete (finished class 6)	4.71	Primary
		Secondary education, incomplete (classes 7 to 8)	4.73	Primary
		Secondary education, complete (finished class 9)	2.81	Lower secondary
		High school (classes 10 to 12)	5.05	Lower secondary
		High school incomplete (classes 10-11)	2.58	Lower secondary
		High school complete (finished class 12)	7.79	Upper secondary
		14^{th} grade incomplete (class 13)	0.66	Upper secondary
		14^{th} grade complete (finished class 14)	2.18	Upper secondary
		University education incomplete	0.54	University
		University education or above	2.42	University
		No information	0.04	
Iraq	What was the highest certificate you attained?	No certificate / Never went to school	40.04	No education
		Elementary	25.31	Primary
		Intermediate	8.39	Lower secondary
		Basic	0.37	Lower secondary
		Secondary	3.95	Upper secondary
		Diploma from an institution	6.00	Upper secondary
		Vocational	1.23	Vocational
		Bachelor degree	5.88	University
		Higher diploma	0.09	University
		Master degree	0.23	University
		PhD (Doctorate)	0.10	University
		No information	8.41	
Syria	What is the highest level of school you attended?	Pre-school / Never attended school	19.66	No education
		Primary	32.14	Primary
		Preparatory	21.04	Lower secondary
		Secondary	10.51	Upper secondary
		Post secondary institute	5.83	Upper secondary
		University and higher	5.18	University
		No information	5.64	
Albania	What is the highest grade you have completed in school?	None	2.73	No education
		8 or 9 years school	44.46	Primary
		Technicum < 2 years	0.71	Lower secondary
		Gymnasium (secondary general)	27.41	Upper secondary
		Vocational 2-3 years	1.94	Vocational
		Vocational 4/5 years	2.80	Vocational
		University - Albania	12.25	University
		University - Abroad	0.15	University
		Master - Albania	0.68	University
		Master - Abroad	0.02	University
		Doctorate/PhD - Albania	0.08	University
		Doctorate/PhD - Abroad	0.01	University
		No information	6.77	
Serbia	Highest ISCED level attained	No education	1.11	No education
		Pre-primary education	2.04	No education
		Primary education	7.60	Primary
		Lower secondary education	16.50	Lower secondary
		Upper secondary education	45.49	Upper secondary
		Post-secondary non tertiary education	1.42	Vocational
		First and second stage of tertiary education	16.72	University
		No information	9.12	

Notes: For Syria, an individual is assigned to one category if at least one grade has been completed at a given level of education, otherwise the level immediately below is attributed. For Serbia, the original questionnaire (2013) asked the acquired education level, with the following answers: (*i*) No school, (*ii*) 4^{th} grade of primary school, (*iii*) 5^{th}-8^{th} grade of primary school, (*iv*) Primary school, (*v*) 3-year secondary school, (*vi*) 4-year secondary school, (*vii*) Specialisation after secondary school, (*viii*) College - 1^{st} faculty degree, (*ix*) Higher education and (*x*) Doctoral studies. However, figures related to this classification are not accessible. Source: Author's elaboration based on the full sample (i.e., without restriction on individuals' age) of the following datasets: SAP (2011-16), COSIT and KRSO (2012–2013), Central Bureau of Statistics (2006), Institute of Statistics of Albania (2012) and EU-SILC (2013-15).

A.2.2 Insecurity in the home country

The subjective perceptions on insecurity (labelled *Insecurity*) in the origin site are reported in three samples (Afghanistan, Iraq and Serbia). The starting point to match information among the different data sources is the answers collected in the IAB-BAMF-SOEP Refugee Sample. Asylum seekers were asked the following semi-open query: "What were the reasons for leaving your country of origin?" and, among all propositions, four of them are retained to build a binary variable indicating whether people felt threatened (at least, by one of the selected propositions) before the migration to Germany. The possible answers are: (*i*) "Fear or violent conflicts or war", (*ii*) "Fear or forced recruitment by military or armed groups", (*iii*) "Persecution" and (*iv*) "Discrimination (ethnic, religious, etc.)".

Table A2.2: Subjective perceptions of insecurity in the origin country

Country	Question	Available answers
Afghanistan	How often do you fear for your own personal safety or security or for that of your family these days?	Never, rarely → Insecurity = 0 Sometimes, often, always → Insecurity = 1
Iraq	In general, how satisfied or unsatisfied are you with you local security level?	Very satisfied → Insecurity = 0 Fairly, not very, not at all satisfied → Insecurity = 1
Serbia (only 2013)	How safe do you feel in the area where you live?	Very safe → Insecurity = 0 Mostly, not feel quite safe, not feel safe at all → Insecurity = 1

Source: Author's elaboration based on SAP (2011-16), COSIT and KRSO (2012-13) and EU-SILC (2013).

Then, the variable is combined with relevant individual characteristics encompassed in the three origin-specific data sets. For each case, a binary indicator has been derived from the ordinal answers that were available for the respondents. Table A2.2 presents the question about insecurity and the procedure followed to assign the replies to to the variable.

A.3 Definition of the variables

Table A3: Definition of the variables

Variable	Definition	Additional comments
Age	Age at the time of the survey (18 to 64)	Include age squared in the regressions
Education*	Level of education. Divided into 3 binary categories: primary or less, secondary and tertiary education	
Speaks German*	1 if an individual is able to speak German, 0 otherwise	Available only for Afghanistan and Albania
Male	1 if an individual is a male, 0 otherwise	
Married	1 if an individual is married, 0 otherwise	Not available for Syria
Insecurity*	1 if an individual feels unsafe in the origin country, 0 otherwise	Not available for Syria and Albania. Replaced by PTS and FH indices for Syria
Religion	Religious affiliation. Divided into 4 binary categories: Atheist, Muslim, Orthodox/Other and Catholic	Available only for Albania
Occupational status*	Divided into 4 dummies: never worked, worker, self-employed and employee	Available only for Serbia

Notes: Variables with a star superscript denote characteristics of asylum seekers that refer to the pre-migration period. For instance, the occupational status corresponds to the position held before they left the origin country.

A.4 List of variables and summary statistics

Table A4.1: List of variables and summary statistics - Afghanistan

Mig	Variable	Mean	SD	N	Mig	Variable	Mean	SD	N
	Level of education					**Level of education**			
	Primary or less	0.76	0.43	50,406		Primary or less	0.70	0.46	442
	Secondary	0.21	0.41	50,406		Secondary	0.26	0.44	442
	Tertiary	0.03	0.17	50,406		Tertiary	0.04	0.20	442
0	Age	33.70	11.59	50,406	1	Age	27.93	8.79	442
	Speaks German	0.00	0.03	50,406		Speaks German	0.05	0.22	442
	Male	0.52	0.50	50,406		Male	0.77	0.42	442
	Married	0.79	0.40	50,406		Married	0.40	0.49	442
	Insecurity	0.63	0.48	50,406		Insecurity	0.94	0.24	442

Notes: Mig refers to the migration status. It takes the value 0 for the origin population and 1 for asylum seekers.
Source: Author's elaboration based on SAP (2011-16) and IAB-BAMP-SOEP Refugee Sample.

Table A4.2: List of variables and summary statistics - Iraq

Mig	Variable	Mean	SD	N	Mig	Variable	Mean	SD	N
	Level of education					**Level of education**			
	Primary or less	0.70	0.46	80,722		Primary or less	0.68	0.47	485
	Secondary	0.23	0.42	80,722		Secondary	0.20	0.40	485
	Tertiary	0.07	0.26	80,722		Tertiary	0.12	0.33	485
0	Age	35.88	12.21	80,722	1	Age	30.74	9.54	485
	Male	0.49	0.50	80,722		Male	0.73	0.44	485
	Married	0.74	0.44	80,722		Married	0.50	0.50	485
	Insecurity	0.60	0.49	80,722		Insecurity	0.98	0.14	485

Notes: Mig refers to the migration status. It takes the value 0 for the origin population and 1 for asylum seekers.
Source: Author's elaboration based on COSIT and KRSO (2012-13) and IAB-BAMP-SOEP Refugee Sample.

Table A4.3: List of variables and summary statistics - Syria

Mig	Variable	Mean	SD	N	Mig	Variable	Mean	SD	N
	Level of education					**Level of education**			
	Primary or less	0.52	0.50	51,968		Primary or less	0.47	0.50	2,046
	Secondary	0.42	0.49	51,968		Secondary	0.36	0.48	2,046
	Tertiary	0.06	0.23	51,968		Tertiary	0.16	0.37	2,046
0	Age	35.28	12.50	51,968	1	Age	31.11	10.31	2,046
	Male	0.51	0.50	51,968		Male	0.74	0.44	2,046
	PTS	4.36	0.77	51,968		PTS	4.97	0.23	2,046
	FH CL	6.55	0.50	51,968		FH CL	6.99	0.08	2,046

Notes: Mig refers to the migration status. It takes the value 0 for the origin population and 1 for asylum seekers. PTS corresponds to Political Terror Scale, while FH CL is the Civil Liberties index from Freedom House.
Source: Author's elaboration based on Central Bureau of Statistics (2006), Gibney et al. (2017), Freedom House (2017) and IAB-BAMP-SOEP Refugee Sample.

Table A4.4: List of variables and summary statistics - Albania

Mig	Variable	Mean	SD	N	Mig	Variable	Mean	SD	N
	Level of education					**Level of education**			
	Primary or less	0.46	0.50	14,829		Primary or less	0.48	0.50	46
	Secondary	0.38	0.49	14,829		Secondary	0.52	0.51	46
	Tertiary	0.16	0.36	14,829		Tertiary	0.00	0.05	46
	Age	41.23	13.14	14,829		Age	30.23	10.09	46
	Speaks German	0.01	0.08	14,829		Speaks German	0.20	0.40	46
0	Male	0.50	0.50	14,829	1	Male	0.72	0.45	46
	Married	0.76	0.43	14,829		Married	0.42	0.50	46
	Religion					**Religion**			
	Atheist	0.01	0.08	14,829		Atheist	0.02	0.13	46
	Muslim	0.82	0.38	14,829		Muslim	0.86	0.35	46
	Orthodox/Other	0.08	0.27	14,829		Orthodox/Other	0.03	0.18	46
	Catholic	0.10	0.29	14,829		Catholic	0.09	0.29	46

Notes: Mig refers to the migration status. It takes the value 0 for the origin population and 1 for asylum seekers.
Source: Author's elaboration based on Institute of Statistics of Albania (2012) and IAB-BAMP-SOEP Refugee Sample.

Table A4.5: List of variables and summary statistics - Serbia

Mig	Variable	Mean	SD	N	Mig	Variable	Mean	SD	N
	Level of education					**Level of education**			
	Primary or less	0.04	0.20	33,395		Primary or less	0.81	0.39	43
	Secondary	0.76	0.43	33,395		Secondary	0.19	0.39	43
	Tertiary	0.20	0.40	33,395		Tertiary			43
	Age	44.07	12.57	33,395		Age	28.24	7.40	43
	Male	0.50	0.50	33,395		Male	0.72	0.46	43
0	Married	0.61	0.49	33,395	1	Married	0.58	0.50	43
	Occupation					**Occupation**			
	Never worked	0.15	0.35	33,395		Never worked	0.53	0.50	43
	Worker	0.03	0.17	33,395		Worker	0.40	0.50	43
	Self-employed	0.13	0.34	33,395		Self-employed	0.05	0.21	43
	Employee	0.69	0.46	33,395		Employee	0.02	0.13	43
	Insecurity	0.60	0.49	8,358		Insecurity	0.62	0.49	42

Notes: Mig refers to the migration status. It takes the value 0 for the origin population and 1 for asylum seekers.
Source: Author's elaboration based on EU-SILC (2013-15) and IAB-BAMP-SOEP Refugee Sample.

A.5 Consistency of language proficiency

This section addresses a potential concern associated to retrospective (i.e., linked to the pre-migration period) questions that were asked in the Refugee Sample. For instance, the ability to speak German is defined from the following question: "How well could you speak the German language before you move to Germany?". An issue will arise if the answers are a function of the time spent in Germany, so that they will be contaminated by the current language aptitude of asylum seekers. To evaluate whether this problem might exists in the empirical investigation, the relevant variable is regressed on the years and months since arrival in Germany. The results obtained using the linear probability model are given in Table A5.

Table A5: Language proficiency and time spent in Germany

Dependent variable: 1 if an individual is able to speak German before migration, 0 otherwise						
	All		Afghanistan		Albania	
Years since arrival	0.003		−0.007		−0.025	
	(0.004)		(0.011)		(0.035)	
Months since arrival		0.000		−0.001		−0.006
		(0.000)		(0.001)		(0.004)
Observations	3,326	3,229	521	499	52	50

Notes: All models are estimated using OLS. Robust standard errors in parentheses. ***, **, and * denote significance at the 1, 5, and 10 percent levels, respectively. All represents the five selected origin countries.
Source: Author's elaboration based on SAP (2011-16), Institute of Statistics of Albania (2012) and IAB-BAMF-SOEP Refugee Sample.

All coefficients are not significant, regardless of the time variable taken into account. This outcome supports the idea that the time spent in the host country is likely not to influence the estimates of the language variable included in the specifications associated to Afghanistan and Albania.

A.6 Results of the logistic regressions

Table A6.1: Self-selection of asylum seekers from conflict-affected countries

Dependent variable: 1 if an individual has migrated, 0 otherwise Benchmark group: Tertiary education								
	Afghanistan			Iraq		Syria		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Level of Education								
Primary or less	-0.626***	-0.719***	-0.434*	-0.688***	-0.914***	-0.838***	-0.841***	-0.842***
	(0.216)	(0.215)	(0.255)	(0.152)	(0.154)	(0.068)	(0.071)	(0.071)
Secondary	-0.984***	-1.037***	-0.794***	-0.730***	-0.844***	-1.316***	-1.326***	-1.329***
	(0.226)	(0.228)	(0.267)	(0.177)	(0.178)	(0.071)	(0.074)	(0.074)
Age	0.213***	0.194***	0.199***	0.125***	0.127***	0.165***	0.166***	0.166***
	(0.040)	(0.039)	(0.039)	(0.027)	(0.028)	(0.015)	(0.015)	(0.015)
Age^2	-0.003***	-0.003***	-0.003***	-0.002***	-0.002***	-0.002***	-0.002***	-0.002***
	(0.001)	(0.001)	(0.001)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Male	0.477***	0.554***	0.531***	0.494***	0.553***	0.553***	0.560***	0.560***
	(0.107)	(0.107)	(0.108)	(0.096)	(0.096)	(0.048)	(0.049)	(0.049)
Married	-1.109***	-1.091***	-1.086***	-0.423***	-0.437***			
	(0.133)	(0.129)	(0.130)	(0.103)	(0.103)			
Insecurity		2.162***	2.161***		3.811***			
		(0.191)	(0.191)		(0.337)			
Speaks German			4.506***					
			(0.406)					
PTS							2.700***	
							(0.146)	
FH CL								4.493***
								(0.231)
Observations	50,848	50,848	50,848	81,207	81,207	54,014	54,014	54,014
McFadden's R^2	0.028	0.074	0.097	0.018	0.103	0.038	0.146	0.164

Notes: All models are estimated using logistic regressions and the reported coefficients are in log-odds units. McFadden's $R^2 = 1 - ln(L_M)/ln(L_0)$, with L_M, the likelihood of the estimated model and L_0, the likelihood of the model without predictors. Robust standard errors in parentheses. ***, **, and * denote significance at the 1, 5, and 10 percent levels, respectively. PTS corresponds to Political Terror Scale, while FH CL is the Civil Liberties index from Freedom House.

Source: Author's elaboration based on SAP (2011-16), COSIT and KRSO (2012-13), Central Bureau of Statistics (2006), Gibney et al. (2017), Freedom House (2017) and IAB-BAMP-SOEP Refugee Sample.

Table A6.2: Self-selection of asylum seekers from Balkan countries

Dependent variable: 1 if an individual has migrated, 0 otherwise Benchmark group: Tertiary education (Albania), Secondary or more (Serbia)						
	Albania			Serbia		
	(1)	(2)	(3)	(4)	(5)	(6)
Level of education						
Primary or less	2.180**	2.553**	2.277**	5.357***	4.699***	5.382***
	(1.004)	(1.071)	(0.985)	(0.402)	(0.445)	(0.464)
Secondary	2.057**	2.355**	2.066**			
	(1.016)	(1.056)	(1.012)			
Age	0.160	0.160	0.158	0.236**	0.226*	0.178
	(0.122)	(0.121)	(0.120)	(0.112)	(0.118)	(0.127)
Age^2	-0.003*	-0.003*	-0.003*	-0.005***	-0.005***	-0.004**
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
Male	-0.359	-0.400	-0.383	0.609*	0.662**	1.009**
	(0.306)	(0.305)	(0.307)	(0.344)	(0.334)	(0.406)
Married	0.727	0.778	0.762	1.346***	1.415***	1.814***
	(0.533)	(0.531)	(0.528)	(0.369)	(0.356)	(0.487)
Speaks German		3.122***				
		(0.664)				
Religion						
Atheist			2.441***			
			(0.554)			
Orthodox / Other			0.366			
			(0.619)			
Catholic			0.399			
			(0.422)			
Occupation						
No work					2.740***	
					(0.998)	
Worker					4.802***	
					(1.016)	
Self-employed					2.044	
					(1.254)	
Insecurity						-0.335
						(0.374)
Observations	14,875	14,875	14,875	33,438	33,438	8,400
McFadden's R^2	0.054	0.074	0.073	0.435	0.515	0.518

Notes: All models are estimated using logistic regressions. McFadden's $R^2 = 1 - ln(L_M)/ln(L_0)$, with L_M, the likelihood of the estimated model and L_0, the likelihood of the model without predictors. Robust standard errors in parentheses. ***, **, and * denote significance at the 1, 5, and 10 percent levels, respectively. Muslim is the benchmark category to analyse the religious affiliation of Albanians. Employee (both with and without supervision tasks) is the reference group to interpret the occupational status in Serbia. Information about insecurity in Serbia is only available in the 2013 wave, which explains the number of observations reported in column (6).

Source: Author's elaboration based on Institute of Statistics of Albania (2012), EU-SILC (2013-15) and IAB-BAMF-SOEP Refugee Sample.

A.7 Robustness checks

Table A7.1: Selection of asylum seekers through the analysis of rare events

Dependent variable: 1 if an individual has migrated, 0 otherwise
Benchmark group: Tertiary education, Secondary or more (Serbia)

	Afghanistan			Iraq		Syria		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Primary or less	-0.008^{**}	-0.010^{***}	-0.005^{*}	-0.006^{***}	-0.008^{***}	-0.049^{***}	-0.046^{***}	-0.045^{***}
	(0.003)	(0.004)	(0.003)	(0.002)	(0.002)	(0.005)	(0.005)	(0.005)
Secondary	-0.011^{***}	-0.012^{***}	-0.008^{**}	-0.006^{***}	-0.008^{***}	-0.064^{***}	-0.061^{***}	-0.061^{***}
	(0.003)	(0.004)	(0.003)	(0.002)	(0.002)	(0.005)	(0.005)	(0.005)
McFadden's R^2	0.028	0.075	0.098	0.018	0.103	0.038	0.146	0.164
Observations	50,848	50,848	50,848	81,207	81,207	54,014	54,014	54,014

	Albania			Serbia		
	(10)	(11)	(12)	(13)	(14)	(15)
Primary or less	0.003^{***}	0.004^{***}	0.004^{***}	0.047^{***}	0.022^{***}	0.129^{***}
	(0.001)	(0.001)	(0.001)	(0.008)	(0.004)	(0.019)
Secondary	0.003^{***}	0.003^{***}	0.003^{***}			
	(0.001)	(0.001)	(0.001)			
McFadden's R^2	0.053	0.077	0.078	0.455	0.540	0.549
Observations	14,875	14,875	14,875	33,438	33,438	8,400

Notes: All models are estimated using penalized logistic regressions and the reported coefficients are the average marginal effects. McFadden's $R^2 = 1 - ln(L_M)/ln(L_0)$ with L_M, the likelihood of the estimated model and L_0, the likelihood of the model without predictors. Robust standard errors in parentheses. ***, **, and * denote significance at the 1, 5, and 10 percent levels, respectively.
Source: Author's elaboration based on SAP (2011-16), COSIT and KRSO (2012-13), Central Bureau of Statistics (2006), Institute of Statistics of Albania (2012), EU-SILC (2013-15) and IAB-BAMF-SOEP Refugee Sample.

Table A7.2: Selection with respect to urban/rural origin population

Dependent variable: 1 if an individual has migrated, 0 otherwise
Benchmark group: Tertiary education, Secondary or more (Serbia)

	Afghanistan			Iraq		Syria		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Primary or less	0.030**	0.024	0.037***	−0.004*	−0.008***	−0.041***	−0.037***	−0.037***
	(0.015)	(0.015)	(0.014)	(0.002)	(0.003)	(0.007)	(0.006)	(0.006)
Secondary	−0.037***	−0.038***	−0.026**	−0.007***	−0.010***	−0.079***	−0.074***	−0.074***
	(0.013)	(0.014)	(0.013)	(0.002)	(0.003)	(0.007)	(0.006)	(0.006)
McFadden's R^2	0.061	0.115	0.134	0.019	0.113	0.042	0.164	0.184
Observations	6,515	6,515	6,515	45,972	45,972	30,736	30,736	30,736

	Albania			Serbia	
	(10)	(11)	(12)	(13)	(14)
Primary or less	0.004**	0.005***	0.005***	0.124***	0.062***
	(0.002)	(0.002)	(0.002)	(0.018)	(0.010)
Secondary	0.006**	0.007***	0.006***		
	(0.003)	(0.002)	(0.002)		
McFadden's R^2	0.048	0.075	0.093	0.500	0.566
Observations	7,277	7,277	7,277	9,240	9,240

Notes: All models are estimated using logistic regressions and the reported coefficients are the average marginal effects. McFadden's $R^2 = 1 - ln(L_M)/ln(L_0)$ with L_M, the likelihood of the estimated model and L_0, the likelihood of the model without predictors. Robust standard errors in parentheses. ***, **, and * denote significance at the 1, 5, and 10 percent levels, respectively. Due to data limitations, column (6) of Table 8 can not be replicated.
Source: Author's elaboration based on SAP (2011-16), COSIT and KRSO (2012-13), Central Bureau of Statistics (2006), Institute of Statistics of Albania (2012), EU-SILC (2013-15) and IAB-BAMF-SOEP Refugee Sample.

Table A7.3: Selection on education of male asylum seekers

Dependent variable: 1 if an individual has migrated, 0 otherwise

Benchmark group: Tertiary education

	Afghanistan			Iraq		Syria		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Primary or less	−0.003	0.001	−0.005	−0.006***	−0.008***	−0.043***	−0.039***	−0.039***
	(0.004)	(0.003)	(0.004)	(0.002)	(0.003)	(0.007)	(0.006)	(0.006)
Secondary	−0.009**	−0.005*	−0.011***	−0.005**	−0.007***	−0.068***	−0.063***	−0.063***
	(0.004)	(0.003)	(0.004)	(0.002)	(0.003)	(0.007)	(0.006)	(0.006)
McFadden's R^2	0.038	0.066	0.095	0.021	0.118	0.031	0.139	0.163
Observations	26,781	26,781	26,781	40,337	40,337	27,746	27,746	27,746

Benchmark group: Secondary education or more

	Albania			Serbia		
	(10)	(11)	(12)	(13)	(14)	(15)
Primary or less	0.002*	0.003*	0.002*	0.054***	0.024***	0.156***
	(0.001)	(0.002)	(0.001)	(0.012)	(0.005)	(0.034)
McFadden's R^2	0.027	0.062	0.034	0.487	0.626	0.595
Observations	7,524	7,524	7,524	16,815	16,815	3,953

Notes: All models are estimated using logistic regressions and the reported coefficients are the average marginal effects. McFadden's $R^2 = 1 - ln(L_M)/ln(L_0)$ with L_M, the likelihood of the estimated model and L_0, the likelihood of the model without predictors. Robust standard errors in parentheses. ***, **, and * denote significance at the 1, 5, and 10 percent levels, respectively.

Source: Author's elaboration based on SAP (2011-16), COSIT and KRSO (2012-13), Central Bureau of Statistics (2006), Institute of Statistics of Albania (2012), EU-SILC (2013-15) and IAB-BAMF-SOEP Refugee Sample.

Table A7.4: Selection on education of family-tied asylum seekers

Dependent variable: 1 if an individual has migrated, 0 otherwise

Benchmark group: Tertiary education

	Afghanistan			Iraq		Syria		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Primary or less	−0.002	−0.002	−0.001	−0.002*	−0.003**	−0.022***	−0.022***	−0.022***
	(0.002)	(0.002)	(0.002)	(0.001)	(0.001)	(0.003)	(0.003)	(0.003)
Secondary	−0.002	−0.002	−0.001	−0.002**	−0.003**	−0.028***	−0.028***	−0.028***
	(0.002)	(0.002)	(0.002)	(0.001)	(0.001)	(0.003)	(0.003)	(0.003)
McFadden's R^2	0.017	0.049	0.068	0.008	0.078	0.025	0.122	0.130
Observations	50,499	50,499	50,499	80,882	80,882	52,680	52,680	52,680

Benchmark group: Secondary education or more

	Albania			Serbia		
	(10)	(11)	(12)	(13)	(14)	(15)
Primary or less	0.000	0.000	0.000	0.012***	0.012*	0.047***
	(0.000)	(0.000)	(0.000)	(0.004)	(0.007)	(0.017)
McFadden's R^2	0.026	0.069	0.066	0.368	0.369	0.445
Observations	14,836	14,836	14,836	33,406	33,406	8,369

Notes: All models are estimated using logistic regressions and the reported coefficients are the average marginal effects. McFadden's $R^2 = 1 - ln(L_M)/ln(L_0)$ with L_M, the likelihood of the estimated model and L_0, the likelihood of the model without predictors. Robust standard errors in parentheses. ***, **, and * denote significance at the 1, 5, and 10 percent levels, respectively.

Source: Author's elaboration based on SAP (2011-16), COSIT and KRSO (2012-13), Central Bureau of Statistics (2006), Institute of Statistics of Albania (2012), EU-SILC (2013-15) and IAB-BAMF-SOEP Refugee Sample.

Chapter II*

*This chapter has been accepted for publication by the *Journal of International Economics*.

Rational inattention and migration decisions*

Simone Bertoli[a], Jesús Fernández-Huertas Moraga[b], and Lucas Guichard[c]

[a] *Université Clermont Auvergne, CNRS, CERDI, IUF, IZA*†

[b] *Universidad Carlos III de Madrid, IZA*‡

[c] *IAB, Université Clermont Auvergne, CNRS, CERDI*§

Abstract

Acquiring information about destinations can be costly for migrants. We model information frictions in the rational inattention framework and obtain a closed-form expression for a migration gravity equation that we bring to the data. The model predicts that flows from countries with a higher cost of information or stronger priors are less responsive to variations in economic conditions in the various destinations, as migrants rationally get less information before deciding where to move. The econometric analysis reveals systematic heterogeneity in the pro-cyclical behavior of migration flows across origins that is consistent with the existence of information frictions.

Keywords: international migration; information; rational inattention; gravity equation.

JEL codes: F22; D81; D83.

*The Authors are grateful to the Editor Treb Allen and to two anonymous referees for their comments, and to Rabah Arezki, Erhan Artuç, Atanas Christev, Mark Dean, Vianney Dequiedt, Lionel Fontagné, Nicholas Hanley, Joël Machado, Thierry Mayer, David McKenzie, Çağlar Özden, Panu Poutvaara, Ariell Reshef, Mark Schaffer, Victor Stephane, Jérôme Valette, and to the participants at various conferences and seminars for their suggestions; the Authors are also grateful to Sergio Correia, Paulo Guimarães and Thomas Zylkin for sharing the Stata command `ppmlhdfe` with us before it became publicly available, and to Olivier Santoni for providing valuable research assistance; Simone Bertoli and Lucas Guichard acknowledge the support received from the *Agence Nationale de la Recherche* of the French government through the program "*Investissements d'avenir*" (ANR-10-LABX-14-01); Jesús Fernández-Huertas Moraga acknowledges the financial support from the *Ministerio de Ciencia e Innovación* (Spain), grants PID2019-111095RB-I00 and MDM 2014-0431; the usual disclaimers apply.

†CERDI, Avenue Léon-Blum, 26, F-63000, Clermont-Ferrand; email: simone.bertoli@uca.fr.

‡Madrid, 126, E-28903, Getafe (Madrid); email: jesferna@eco.uc3m.es.

§Regensburger Strasse, 104, D-90478 Nuremberg; email: lucas.guichard2@iab.de.

"Before making a choice, one may have an opportunity to study the actions and their payoffs; however, in most cases it is too costly to investigate to the point where the payoffs are known with certainty. As a result, some uncertainty about the payoffs remains when one chooses among the actions even if complete information was available in principle."

(Matějka and McKay, 2015, p. 272)

2.1 Introduction

Human migration is portrayed as an investment decision that should be based on a comparison of the private returns for the migrant in each of the potential destinations (Sjaastad, 1962), but the key elements that lead to the choice of the preferred destination are unlikely to be readily available. The migrant needs first to gather information about the attractiveness of the various countries she could opt for. However, some of the seminal contributions to the modeling of the determinants of migration choice assume that uncertainty is fully (and costlessly) resolved before deciding where to migrate.[1] In particular, this is the case for the canonical micro-foundations of migration gravity equations that rely on discrete choice models *à la* McFadden (McFadden, 1974). In contrast, there is empirical evidence revealing that potential migrants can have inaccurate expectations on their earnings abroad (McKenzie et al., 2013) or about the costs and risks associated to migrating (Shrestha, 2020).

This suggests that the uncertainty surrounding the utility at destination might not be entirely resolved when a migrant has to come up with a decision, and the size of the remaining uncertainty could be endogenously determined. The literature on rational inattention (Sims, 1998, 2003), which has been recently applied to discrete choice situations (Matějka and McKay, 2015; Caplin et al., 2019), provides us with a framework to think about how costs associated to information acquisition and processing would influence the specification of the migration gravity equation that is brought to the data.

How can we enhance our understanding of the determinants of international migration flows if we take into account the uncertainty that migrants face, and the costly actions that they can take to narrow it down? We estimate a gravity equation whose specification is derived from the analysis of a location-decision problem with information frictions. We obtain a closed-form expression for optimal choice probabilities under suitable assumptions on the priors held

[1] Borjas (1987) assumes that migration decisions are based on a comparison of "potential incomes" at origin and at destination (p. 532), with the latter being known before migrating, in line with the analysis by Roy (1951) on the occupational choice between hunting and fishing that explicitly assumes that "[e]very man, too, has a fairly good idea of what his annual output is likely to be in both occupations" (p. 137).

by the migrants about the distribution of destination-specific utility, following Dasgupta and Mondria (2018).[2] The main testable implication of this model is that the responsiveness of bilateral migration flows with respect to variations in the attractiveness of alternative destinations is larger when migrants have a stronger incentive to acquire information before deciding where to move. We refer to this incentive as the value of information, which is related to the ratio between the variance of the prior distribution of destination-specific utility and the marginal cost of receiving signals about the actual attractiveness of the various alternatives in the choice set. The distribution of past migration flows across destinations can be used to infer the (unobserved) value of information, and we exploit this property to estimate the model.

We draw on data on bilateral migration flows between 1960 and 2015 from Abel (2018) to build an origin-specific and time-varying measure of the value of information for international migrants, which is inversely related to the share of cumulated past flows directed to the main destination.[3] We estimate a gravity equation where the destination-specific utility depends on an interaction between income per capita at destination and our empirical counterpart of the value of information. The results are in line with the theoretical model: a one standard deviation increase in our proxy for the value of information determines an increase in the estimated elasticity between 0.063 and 0.083.[4] Our estimates imply that the elasticity of the bilateral migration rate with respect to income per capita for China is 0.182-0.241 higher than the corresponding elasticity for Mexico, which represents a paradigmatic case of migration flows concentrated in just one single destination, namely the United States. Our results are robust when we exclude the main origin-specific destination from the sample, so that they are are not driven by a lower procyclicality of the migration flows directed to just one destination but rather, as the theory predicts, to all foreign countries. Our results are inconsistent with the predictions stemming from a canonical random utility maximization model with unobserved heterogeneity, where the variance of the stochastic component of utility is origin-specific. This alternative full-information model would imply that the coefficient of our interaction term should have the opposite sign to the one that we obtain when estimating our gravity equation.

The econometric evidence that we provide is fully robust when we allow for additional hetero-

[2] Dasgupta and Mondria (2018) have drawn on Matějka and McKay (2015) to extend the N-country Ricardian model of trade by Eaton and Kortum (2002), introducing costly acquisition of information on the prices of goods in different exporting countries.

[3] Our reliance on the distribution of past migration flows across destinations to measure the value of information acquisition is closely related to the use of past market shares in Caplin et al. (2016).

[4] Consistently with a theoretical result derived by Dasgupta and Mondria (2018), we obtain a non-significant coefficient for this interaction term when we measure the value of information using the past share of migrants in destinations other than the main one.

geneity in the coefficient of income at destination either across origins or at the dyadic level. Specifically, we let this coefficient vary also with the level of income of the migrant-sending country, with its past total emigration rate, and with dyadic correlates of migration costs, such as the size of migrant networks at destination, geographic, cultural or linguistic distance. This, in turn, implies that our results cannot be explained by a full-information model with a richer and more flexible specification of the deterministic component of utility, where the effect of income at destination depends in a multiplicative way on other variables, which might also be correlated with the past distribution of flows across destinations. Thus, the results of the estimation of our theory-based gravity equation suggest that variations in economic conditions in a given destination country influence more incoming migration flows from origins where migrants (rationally) invest more in information acquisition.

This paper is mainly related to two strands of literature, namely (*i*) the theoretical analyses of discrete choice models with costly information acquisition (Matějka and McKay, 2015; Caplin et al., 2019; Fosgerau et al., 2020; Steiner et al., 2017), and (*ii*) the analysis of the determinants of international migration flows through micro-founded specifications of the gravity equation (see, for instance, Mayda, 2010; Grogger and Hanson, 2011; Bertoli and Fernández-Huertas Moraga, 2013; Ortega and Peri, 2013).[5] With respect to (*i*), we make three distinct contributions to the literature on rational inattention. First, we prove that all alternatives are chosen with positive probability,[6] once we assume that utility is identically and independently distributed according to a conjugate of a Gumbel distribution (Cardell, 1997) around a destination-specific expected value.[7] Second, we show that the optimal total investment in information acquisition is negatively related to the expected utility associated to the alternative that is, *a priori*, most attractive, but that the migrant chooses to receive more informative (and hence costly) signals about the alternatives that are less likely to be selected. This latter theoretical result is reminiscent of the evidence about the redirection of attention towards less attractive options in the so-called lemon-dropping markets in Bartoš et al. (2016). Third, we provide evidence of the empirical relevance of rational inattention in discrete choice situations, complementing a strand of literature that is still mostly theoretical.[8]

[5] Batista and McKenzie (2018) have recently tested in the lab these micro-foundations, notably allowing players to pay a cost to reduce the uncertainty about the payoffs associated to the various destinations.

[6] This is a natural property in models of industrial organization, e.g., Brown and Jeon (2020), where profit-maximizing rules out prices that would bring the demand to zero, but needs to be demonstrated in settings in which the attractiveness of the various alternatives is not endogenously determined.

[7] "Determining the empirical content of the rational inattention model with nonexchangeable priors [...] is an active area of research" (Natenzon, 2019, p. 445), and our paper thus also contributes to develop the analysis of models where the priors about the distribution of utility are alternative-specific.

[8] "The model of [rational inattention] is well suited for a boom in empirical work, which has not yet occurred" (Maćkowiak et al., 2018, p. 27).

Migrants appear to be rationally inattentive even though the stakes related to their location decisions are certainly very high (see, for instance, McKenzie et al., 2010 and Clemens et al., 2019).

As far as (*ii*) is concerned, ours is the first paper bringing to the data a migration gravity equation derived from a model with information frictions, with Porcher (2019) being the only other paper we are aware of, in his case exploiting internal migration flows in Brazil. Furthermore, we make two main contributions. First, we demonstrate that an alternative micro-foundation of the migration gravity equation allows for uncovering and interpreting systematic heterogeneities across origins in the responsiveness of migration flows with respect to varying economic conditions in the various destination countries. Second, our analysis implies an additional reason why migration flows have an inertial character, over and above the positive externalities generated due to destination-specific migration networks (e.g., Munshi, 2003), as information frictions induce a more concentrated distribution of migrants across destinations.[9]

The rest of the paper is structured as follows: Section 2.2 introduces information frictions in a standard location-choice problem, solving it under suitable distributional assumptions, and deriving its testable implications; Section 2.3 briefly presents the main data sources, it describes how we bring the model to the data, and it presents basic descriptive statistics. Section 2.4 presents the results of the econometric analysis, and Section 2.5 concludes.

2.2 Theoretical model

Consider a migrant from the origin country j who has to select her preferred destination from a choice set A including N alternatives, i.e., foreign countries, so that we analyze the choice of the destination conditional upon migrating. Let $v_{jk} = w_k - c_{jk}$ denote the utility, or payoff, associated to alternative $k \in A$, and let $\boldsymbol{v}_j \in \mathbb{R}^N$ represent the vector of payoffs, which we will be referring to as the state of the world. We omit the origin subscript j to avoid cluttering the notation, but the distribution of payoffs (because of the dyadic migration costs c_{jk}) and all the other parameters of the model can be origin-specific.

We denote by $F(\boldsymbol{v})$ the belief held by the migrant on the distribution of the state of the world; we assume that $F(\boldsymbol{v})$ is differentiable, and we denote by $f(\boldsymbol{v})$ the probability density function. We define $\overline{v}_k \equiv \int_{\boldsymbol{v}} v_k f(\boldsymbol{v})\,\mathrm{d}\boldsymbol{v}$, and we assume that the expected value of the payoff is finite.

[9] Our results also reveal an additional dimension of interdependence between migration flows directed to different countries, beyond the strategic interactions in migration policies (Giordani and Ruta, 2013).

Without loss of generality, we also assume that $\overline{v}_k \geq \overline{v}_h$, when $k < h$, $\forall k,h \in A$.

The migrant can obtain a signal $\boldsymbol{s} \in \mathbb{R}^N$ about the payoffs in the various alternatives in the choice set, choosing both where to focus her attention (e.g., some destinations might be completely disregarded), and how much information to acquire before deciding where to migrate. More precise signals, i.e., signals that convey more information about the state of the world, are more costly, and the cost of information acquisition is proportional to the mutual information between the signal $\boldsymbol{s}$ and the state $\boldsymbol{v}$. The parameter $\lambda > 0$ translates the reduction in the entropy of $\boldsymbol{v}$ (Shannon, 1948) induced by the chosen information acquisition strategy into the same metrics as the payoffs.[10] The migrant behaves as a Bayesian expected utility maximizer, selecting the alternative in A with the highest expected payoff given the posterior distribution of $\boldsymbol{v}$ that has been induced by the signal $\boldsymbol{s}$, i.e., $F(\boldsymbol{v}|\boldsymbol{s})$.

Letting $S_k \subseteq \mathbb{R}^N$ be the set of signals that induces the migrant to select $k \in A$, the probability of opting for alternative k under the state of the world $\boldsymbol{v}$ is given by:

$$\mathcal{P}_k(\boldsymbol{v}) \equiv \int_{\boldsymbol{s} \in S_k} F(\mathrm{d}\boldsymbol{s}|\boldsymbol{v})$$

A key property of this model is that the migrant is never going to acquire distinct signals that lead to the choice of the same alternative, as in this case costly information would be acquired but not acted upon. This implies that the mutual information between the state and the signal is the same as the mutual information between the state and the alternative. This fundamental result (see Lemma 1 in Matějka and McKay, 2015), coupled with the symmetry of mutual information, implies that we can cast the location-decision problem facing the migrant in terms of the selection of the conditional choice probabilities $\mathcal{P}_k(\boldsymbol{v})$, $\forall k \in A$. The location-decision problem that the migrant faces can thus be described as follows:[11]

$$\max_{\mathcal{P}=\{\mathcal{P}_a(\boldsymbol{v})\}_{a=1}^N} \sum_{a=1}^{N} \int_{\boldsymbol{v}} v_a \mathcal{P}_a(\boldsymbol{v}) f(\boldsymbol{v})\, \mathrm{d}\boldsymbol{v} - C(\mathcal{P}), \tag{2.1}$$

where:

$$C(\mathcal{P}) \equiv \sum_{a=1}^{N} C_a(\mathcal{P}),\ C_a(\mathcal{P}) = \lambda \left(-\mathcal{P}_a \ln \mathcal{P}_a + \int \mathcal{P}_a(\boldsymbol{v}) \ln \mathcal{P}_a(\boldsymbol{v}) f(\boldsymbol{v})\, \mathrm{d}\boldsymbol{v} \right), \tag{2.2}$$

[10] This parameter is invariant across alternatives in the choice set; if λ was alternative-specific, conditional choice probabilities would no longer have the functional form derived by Matějka and McKay (2015).

[11] In the expression for entropy, we adopt the convention that $0 \ln(0) = 0$.

with $\mathcal{P}_a \equiv \int_{\boldsymbol{v}} \mathcal{P}_a(\boldsymbol{v}) f(\boldsymbol{v})\, \mathrm{d}\boldsymbol{v}$, and subject to the constraints:

$$\mathcal{P}_a(\boldsymbol{v}) \geq 0, \forall a \in A, \forall \boldsymbol{v} \in \mathbb{R}^N, \sum_{a=1}^{N} \mathcal{P}_a(\boldsymbol{v}) = 1, \forall \boldsymbol{v} \in \mathbb{R}^N. \tag{2.3}$$

The location-decision problem described in (2.1)-(2.3) is characterized by the parameter $\lambda > 0$, and by the function $f(\boldsymbol{v})$ that denotes the distribution of the vector of payoffs.

2.2.1 Solution of the model

Matějka and McKay (2015) prove in Theorem 1 that the optimal conditional probability for $k \in B$ is given by:[12]

$$\mathcal{P}_k(\boldsymbol{v}) = \frac{\mathcal{P}_k e^{v_k/\lambda}}{\sum_{a \in B} \mathcal{P}_a e^{v_a/\lambda}} \tag{2.4}$$

where $\mathcal{P}_k \equiv \int_{\boldsymbol{v}} \mathcal{P}_k(\boldsymbol{v}) f(\boldsymbol{v})\, \mathrm{d}\boldsymbol{v}$. We denote by $B \subseteq A$ the consideration set (Caplin et al., 2019), i.e., the set of alternatives that are chosen with positive probability.

If we plug the expression for $\mathcal{P}_k(\boldsymbol{v})$ in (2.4) in the original maximization problem in (2.1), this can be expressed only in terms of the unconditional probabilities:[13]

$$\max_{\mathcal{P}_1,\ldots,\mathcal{P}_N} \int_{\boldsymbol{v}} \lambda \ln \left[\sum_{a \in B} \mathcal{P}_a e^{v_a/\lambda} \right] f(\boldsymbol{v})\, \mathrm{d}\boldsymbol{v} \tag{2.5}$$

The analytical challenges that are related to the solution of the model are that (*i*) we do not know what is the composition of the set B,[14] and (*ii*) a closed-form expression for the integral in (2.5) does not, in general, exist.

2.2.1.1 Consideration set

With respect to point (*i*), the number of potential sets of alternatives that correspond to the solution of the maximization problem in (2.1) stands in general at $2^N - 1$. If $v_k = \overline{v}_k + \epsilon_k$ and ϵ_k is identically and independently distributed for all alternatives $k \in A$, then there are just N different subsets of A that can be the consideration set, and these are nested. This is implied by Theorem 2 in Caplin et al. (2019); when payoffs are independently distributed

[12] See Fosgerau et al. (2020) on the relationship between the use of Shannon entropy to define $C(\mathcal{P})$ and the functional form of optimal conditional choice probabilities.

[13] This is Lemma 2 in Matějka and McKay (2015).

[14] Caplin et al. (2019) derive necessary and sufficient conditions to have that $P_k > 0$.

across alternatives, if $k \in B$, i.e., $\mathcal{P}_k > 0$, then $l \in B$ if:

$$\int_{-\infty}^{+\infty} e^{(\overline{v}_l+\epsilon_l)/\lambda} f(\epsilon_l)\,\mathrm{d}\epsilon_l \geq \int_{-\infty}^{+\infty} e^{(\overline{v}_l+\epsilon_k)/\lambda} f(\epsilon_k)\,\mathrm{d}\epsilon_k \tag{2.6}$$

When ϵ_k and ϵ_l are identically distributed, then the distribution of the payoff for alternative k is first-order stochastically dominated by the distribution of the payoff for alternative l, $\forall l < k$. Thus, if an alternative $k \in B$, then $l \in B$ for all alternatives $l = 1,...,k-1$, and the consideration set can only be of the type $B_k = \{1,...,k\}$, with $k = 1,...,N$.

2.2.1.2 Solving for unconditional probabilities

As far as point (*ii*) is concerned, a closed-form solution for the unconditional probabilities can be obtained by assuming that the distribution of payoffs is the same across all alternatives, so that $\overline{v}_k = \overline{v}$, $\forall k \in A$, or by allowing for alternative-specific values of the expected payoff under suitable distributional assumptions. If payoffs are identically distributed for all alternatives, then the consideration set is $B_N = A$, and all alternatives are chosen with probability $1/N$.[15]

The second option is to introduce the same distributional assumptions as in Dasgupta and Mondria (2018), Brown and Jeon (2020) and Porcher (2019). We can thus assume that $v_k = \overline{v}_k + \epsilon_k$, where ϵ_k is identically and independently drawn according to a Cardell distribution $C(\lambda)$, with $\lambda \in (0,1)$. The key property of this distribution, whose density is fully supported on the real line, is that it is the (unique) conjugate of the EVT-1 distribution: when η_k is EVT-1 and ϵ_k is an independent $C(\lambda)$ random variable, then $\epsilon_k + \lambda\eta_k$ follows an EVT-1 distribution (Cardell, 1997).[16,17]

With these distributional assumptions, once we fix λ we are also pinning down σ^2, but the ratio between the variance of the payoffs and the marginal cost of acquiring information can take any positive value when $\lambda \in (0,1)$, as $\sigma^2/\lambda = \frac{(1-\lambda^2)}{\lambda}(\pi^2/6)$. Thus, we can represent a location-decision problem with an arbitrary quantity associated to the ratio between the value of acquiring information, which depends on the extent to which payoffs vary with the state of the world, and the marginal cost of acquiring information.

[15] See Proposition 1 in Matějka and McKay (2015).

[16] The variance σ^2 of $C(\lambda)$ is equal to $(1-\lambda^2)\pi^2/6$, so that the variance of $v_k + \lambda\epsilon_k$ is equal to $\pi^2/6$, i.e., the variance of a Gumbel distribution with a scale parameter equal to 1.

[17] As with a Gumbel distribution, the difference of two independent $C(\lambda)$ random variables follows a logistic distribution (Cardell, 1997), with scale parameter equal to $\sqrt{1-\lambda^2}$.

2.2.2 Optimal unconditional probabilities

The integral in (2.5) can be solved given the distributional assumptions that we have just introduced, and the constrained maximization problem simplifies to:[18]

$$\max_{\mathcal{P}_1,...,\mathcal{P}_k} \ln \left[\sum_{a \in B_k} e^{\overline{v}_a + \lambda \ln(\mathcal{P}_a)} \right] \tag{2.7}$$

The maximization problem in (2.7) can be solved for an arbitrary set B_k, with $k = 1, ..., N$; the solution is given by:[19]

$$\mathcal{P}_h^{B_k} = \frac{e^{\overline{v}_h/(1-\lambda)}}{\sum_{a \in B_k} e^{\overline{v}_a/(1-\lambda)}} \tag{2.8}$$

We can show that the expected utility from choosing with positive probability the alternatives in the set B_k monotonically increases with k, so that the consideration set is given by $B_N = A$, i.e., all alternatives are always selected with positive probability.[20] The optimal unconditional probabilities are given by:

$$\mathcal{P}_h = \frac{e^{\overline{v}_h/(1-\lambda)}}{\sum_{a \in A} e^{\overline{v}_a/(1-\lambda)}} \tag{2.9}$$

Notice, as proved by Dasgupta and Mondria (2018), that $\mathcal{P}_h$ is a non-monotonic function of λ for $h = 2, ..., N-1$, while $\mathcal{P}_1$ ($\mathcal{P}_N$) is monotonically increasing (decreasing) in λ, as:

$$\frac{\partial \ln \mathcal{P}_h}{\partial \lambda} = \frac{e^{\overline{v}_h/(\lambda-1)}}{(1-\lambda)^2} \left(\overline{v}_h - \sum_{a=1}^{N} \mathcal{P}_a \overline{v}_a \right) \tag{2.10}$$

The sign of the partial derivative in (2.10) depends on the difference between $\overline{v}_h$ and a probability-weighted average of the payoffs of all alternatives, which is unambiguously lower (higher) than the payoff of the most (least) attractive alternative. We will exploit the fact that:

$$\frac{\partial \ln \mathcal{P}_1}{\partial \lambda} > 0 \tag{2.11}$$

in the empirical analysis to obtain information on the unobserved value of this key parameter from observed past migration flow data.

[18] See the proof in the Appendix A.1.1.

[19] See the proof in the Appendix A.1.2.

[20] See the proof in the Appendix A.1.3; this property also implies that, as in Brown and Jeon (2020), the model can still admit a closed-form solution in the presence of unobserved individual heterogeneity, as all alternatives are always included in the (individual) consideration set.

2.2.3 Closed-form conditional choice probabilities

If we plug in (2.4) the expression for the unconditional choice probabilities in (2.9), when $B_k = A$:

$$\mathcal{P}_k(\boldsymbol{\epsilon}) = \frac{\mathcal{P}_k e^{v_k/\lambda}}{\sum_{a\in A}\mathcal{P}_a e^{v_a/\lambda}} = \frac{e^{\frac{\epsilon_k}{\lambda}+\frac{\overline{v}_k}{\lambda(1-\lambda)}}}{\sum_{a\in A} e^{\frac{\epsilon_a}{\lambda}+\frac{\overline{v}_a}{\lambda(1-\lambda)}}} \tag{2.12}$$

The conditional choice probability $\mathcal{P}_k(\boldsymbol{\epsilon})$ can be written as a function of the *un*conditional choice probabilities $\mathcal{P}_a$, which only depend on the vector of expected payoffs $\overline{v}$, and on the vector $\boldsymbol{\epsilon}$ of the realizations of the deviation of the actual payoffs from their expected values.

2.2.4 Optimal cost of information acquisition

We can gain further insights on features of the solution of the location-choice problem with costly information acquisition by analyzing a simplified version of the model where $A = \{1,2\}$, which gives us the opportunity to present the results graphically.[21] Without loss of generality, we can set $\overline{v}_2 = 0$; the optimal conditional probability $\mathcal{P}_1(x)$ of selecting alternative 1 is thus given by:

$$\mathcal{P}_1(x) = \left[1 + \left(\frac{1-\mathcal{P}_1}{\mathcal{P}_1}\right)^{1/\lambda} e^{-x/\lambda}\right]^{-1} \tag{2.13}$$

where $x = \epsilon_1 - \epsilon_2$ follows a logistic distribution, with the cumulative distribution:

$$G(x) = \left(1 + e^{-x/\sqrt{1-\lambda^2}}\right)^{-1}.$$

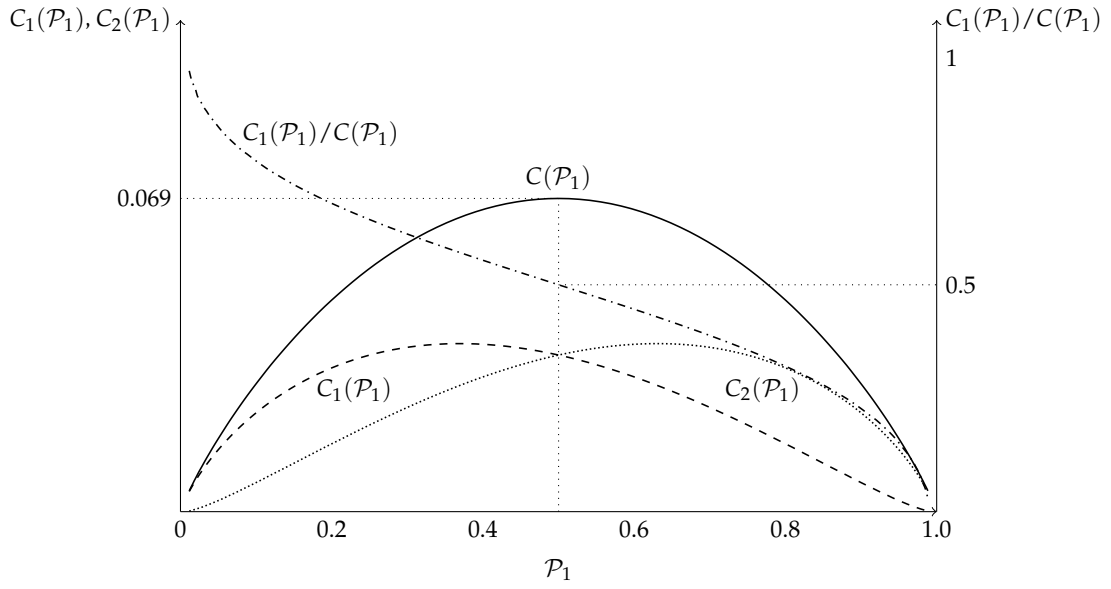
We can thus rewrite the two alternative-specific costs of information acquisition as follows:

$$C_1(\mathcal{P}_1) = \lambda\left(-\mathcal{P}_1 \ln \mathcal{P}_1 + \int_{-\infty}^{+\infty} \mathcal{P}_1(x)\ln \mathcal{P}_1(x) g(x)\,\mathrm{d}x\right) \tag{2.14}$$

where $g(x) = \partial G(x)/\partial x$, and $C_2(\mathcal{P}_1) = C_1(1-\mathcal{P}_1)$.

[21] We describe below how these results generalize to the case in which $N > 2$.

Figure 2.1: Total and (absolute and relative) alternative-specific optimal cost of information



Notes: the integral that enters into the expression for $C_1(\mathcal{P}_1)$ is solved numerically for $\lambda = 0.1$.

The integrand function that appears in (2.14) does not admit a closed-form primitive, but we can gain insights on the total and alternative-specific investment in information acquisition by numerically solving for $C_1(\mathcal{P}_1)$.[22] Figure 2.1 plots the values of $C(\mathcal{P}_1)$, $C_1(\mathcal{P}_1)$ and $C_2(\mathcal{P}_1)$ against $\mathcal{P}_1$ when $\lambda = 0.1$ (left-hand side vertical axis), as well as the value of the ratio $C_1(\mathcal{P}_1)/C(\mathcal{P}_1)$ (right-hand side axis).[23]

Several features of the evolution of the cost of information acquisition, and of its distribution between the two alternatives, with respect to $\mathcal{P}_1$, are worth emphasizing:[24] First, the total cost of information acquisition $C(\mathcal{P}_1)$ is maximized when $\mathcal{P}_1 = 1/2$,[25] and it is monotonically increasing (decreasing) in $\mathcal{P}_1$ when $\mathcal{P}_1 < 1/2$ ($\mathcal{P}_1 > 1/2$). Second, the migrant invests more in information acquisition about the alternative that is *a priori* less attractive, as $C_1(\mathcal{P}_1) <$

[22] This is done by computing the value of the integral in (2.14) with 2,000 draws for x; the computation is repeated 2,000 times, and we then average $C_1(\mathcal{P}_1)$ over these replications; we then define $C_2(\mathcal{P}_1) = C_1(1 - \mathcal{P}_1)$.

[23] We thank an anonymous referee for pushing us to explore the uneven allocation of attention across alternatives in the choice set.

[24] These properties are independent from the value of λ, and are demonstrated analytically when the two alternatives are *ex ante* identical in the Appendix A.1.4; an increase in λ exerts an ambiguous effect on $C(\mathcal{P}_1)$, while it unambiguously reduces the optimal reduction in the entropy of the payoffs, i.e., $C(\mathcal{P}_1)/\lambda$.

[25] We have that $C(1/2) = 0.069$; as $\lambda = 0.1$, the reduction in entropy stands at 0.69; as the entropy of the distribution of the priors is approximately equal to $2(1+\gamma) \approx 3.14$, where $1+\gamma$ is the entropy of a univariate Gumbel distribution, so the entropy is reduced by approximately 22 percent with the optimal signal acquisition strategy.

$C_2(\mathcal{P}_1)$ when $\mathcal{P}_1 > \mathcal{P}_2$. Third, the alternative-specific investment in information acquisition is maximized when the probability of choosing an alternative is below 1/2. Fourth, the share of the total cost of information acquisition that is directed towards alternative 1, i.e., $C_1(\mathcal{P})/C(\mathcal{P})$, monotonically declines with the probability of selecting alternative 1. In terms of the signals, the migrant rationally decides to receive a more precise signal with respect to the payoff of the less attractive alternative, so that for this alternative the conditional choice probabilities vary more with respect to $x = \epsilon_1 - \epsilon_2$.

When the choice set A includes N alternatives, we can follow Bunch and Rocke (2016) to obtain independent draws of the payoffs from a $C(\lambda)$ distribution, and numerically compute the value of $C(\mathcal{P})$. This reveals that the properties that we have just described extend to an arbitrary number of alternatives. Notably, $C(\mathcal{P})$ is maximized when $\mathcal{P}_k = 1/N$, $\forall k \in A$,[26] and the cost of information acquisition for the alternative that is *a priori* most attractive is always below the cost of information acquisition for at least another alternative in the choice set A.[27]

2.2.5 Elasticities

The semi-elasticity of the choice probability $\mathcal{P}_k(\boldsymbol{\epsilon})$ with respect ϵ_k and the expected value of this semi-elasticity are given by:

$$\frac{\partial \ln[\mathcal{P}_k(\boldsymbol{\epsilon})]}{\partial \epsilon_k} = \frac{1}{\lambda}\left[1 - \mathcal{P}_k(\boldsymbol{\epsilon})\right], \mathbb{E}_{\boldsymbol{\epsilon}}\left(\frac{\partial \ln[\mathcal{P}_k(\boldsymbol{\epsilon})]}{\partial \epsilon_k}\right) = \frac{1}{\lambda}\left(1 - \mathcal{P}_k\right) \tag{2.15}$$

Thus, this elasticity is higher for alternatives whose unconditional probability of being chosen is low; this can be related to how the alternative-specific investment in information acquisition $C_k(\mathcal{P})$ is related to the unconditional choice probability $\mathcal{P}_k$. We can write down the corresponding expressions for the elasticities with respect to $\overline{v}_k$:

$$\frac{\partial \ln[\mathcal{P}_k(\boldsymbol{\epsilon})]}{\partial \overline{v}_k} = \frac{1}{\lambda(1-\lambda)}\left[1 - \mathcal{P}_k(\boldsymbol{\epsilon})\right], \mathbb{E}_{\boldsymbol{\epsilon}}\left(\frac{\partial \ln[\mathcal{P}_k(\boldsymbol{\epsilon})]}{\partial \overline{v}_k}\right) = \frac{1}{\lambda(1-\lambda)}\left(1 - \mathcal{P}_k\right) \tag{2.16}$$

The ratio between (2.15) and (2.16) stands at $1-\lambda$: when λ increases, the relative size of the average elasticity of $\mathcal{P}_k(\boldsymbol{\epsilon})$ with respect to deviations of the payoff from its expected value declines, as the migrant is (rationally) receiving less precise signals about the payoff.

[26] We can also demonstrate that $C(\mathcal{P})$ monotonically increases with N, while $C_k(\mathcal{P}) = C(\mathcal{P})/N$, $\forall k \in A$, monotonically declines with the size of the choice set when alternatives are *ex ante* identical.

[27] A corollary of this property is that $C_k(\mathcal{P})$ is maximized when $\mathcal{P}_k < 1/2$, and we can demonstrate that $C_k(\mathcal{P})$ is an hump-shaped function of $\mathcal{P}_k$.

2.2.6 Testable implication

Our location-choice model with costly information acquisition implies that (*i*) the responsiveness of optimal conditional choice probabilities to variations in the expected value of the payoff in one alternative in the choice set is negatively related to λ, the parameter that determines the marginal cost of information acquisition, as shown in equation (2.15), and that (*ii*) there is a monotonic positive relationship between λ and $\mathcal{P}_1$, i.e., the unconditional probability of opting for the most attractive destination, as shown in (2.11). The econometric analysis will exploit point (*ii*) to build the empirical counterpart of λ from the distribution of past (origin-specific) international migration flows, and bring to the data the testable implication described at point (*i*).

2.3 From the theory to the data

We describe here the source of our panel data on bilateral international migration flows, and how we build the empirical proxy for the cost of information acquisition λ (or, more precisely, for $1/\lambda$), which we will term the value of information. We also present basic descriptive statistics, focusing in particular on our variable of interest.

2.3.1 Data on bilateral migration flows

Our main data source is represented by Abel (2018), which provides data on the bilateral migration flows $m_{jkt} \geq 0$ between the origin j and the destination k across 203 countries for five-year periods, starting in t, between 1960 and 2015. Abel (2018) extends the methodology presented by Abel and Sander (2014) for inferring gender-specific bilateral migration flows from census-based data on the stock of individuals (by country of birth) residing in each country. More precisely, Abel (2018) recovers the minimal amount of bilateral flows that are required to match the observed evolution of stock data, once these have been adjusted for demographic events. The stock data are taken from Özden et al. (2011) between 1960 and 2000, and from United Nations Population Division (2015a) for later years, and are combined with demographic information from United Nations Population Division (2015b) to obtain the estimates on flows. To our knowledge, the dataset generated by Abel (2018) is the most comprehensive in terms of both time and geographical coverage produced to date on interna-

tional migration flows.[28] As discussed below, these two aspects are critical to generate from the data what we define as the value of information, the key variable that allows us to recover the effect of income at destination in a way that reflects the presence of information frictions in the location-decision problem that migrants face. The sample over which we conduct our econometric analysis includes the entire set of countries covered by Abel (2018): for the period between 1980 and 2015, we have 263,008 observations on bilateral migration flows over seven consecutive five-year periods.[29,30] The average value of m_{jkt} stands at 957.4, with a standard deviation of 15,472.4, and 61.2 percent of zero flows.

2.3.2 Measurement of the value of information

Eq. (2.11) suggests that we can build from the data a suitable empirical counterpart of the (unknown) origin-specific value of information. The location-decision problem presented in Section 2.2 is static, while the availability of longitudinal data on bilateral migration flows allows us to build an empirical counterpart of the value of information that is possibly time-varying. More precisely, we proxy $\mathcal{P}_1$ with the share of migration flows directed from j to the main foreign destination in a period up to t. We rely on $p(r)_{jt}$, defined as follows:[31]

$$p(r)_{jt} \equiv \max_k \left[\frac{\sum_{t-r}^{t} m_{jkt}}{\sum_{t-r}^{t} \sum_{l \in A} m_{jlt}} \right], r = \{5,10,15,20\} \tag{2.17}$$

It is interesting to note that 105 different countries represent the main destination, and hence determine the value of information, for at least one of the 1,347 origin-year pairs in our estimation sample; unsurprisingly, the United States are the most typical main destination accumulating most of the flows for a particular origin, but this happens only in 20.7 percent of the cases; the second most typical main destination is Russia, for 7.6 percent of all origin-year pairs, and five Sub-Saharan African countries (namely, South Africa, Ethiopia, Nigeria, the Democratic Republic of the Congo, and Ivory Coast) appear among the 20 countries that most frequently play the role of main destination, thus revealing the importance of having a

[28] Our empirical evidence is robust to using only the bilateral flow data in Abel (2018) that are based solely on migrant stocks from Özden et al. (2011), thus avoiding possible inconsistencies at the junction between the two underlying data sources, and to defining bilateral migration flows as the variations in the stock of j-born individuals residing in destination k derived from Özden et al. (2011).

[29] Migration flows before 1980 are used to measure the value of information (see Section 2.3.2 below).

[30] This is below $203 \times 202 \times 7 = 287,042$ as we have missing information of GDP per capita at destination for some destination-year pairs; more precisely, we lose completely 14 minor destination countries, which represent less than 0.9 percent of total migration flows in Abel (2018).

[31] Notice that $p(r)_{jt}$ in (2.17) is defined provided that the total flow originating from j between year $t-r$ and t is positive; this is always the case except for 31 origin-year pairs when $r = 5$, 14 origin-year pairs when $r = 10$, 7 when $r = 15$, and 6 when $r = 20$.

comprehensive set of destinations covered in the data. As $\mathcal{P}_1$ is a monotonically increasing function of λ, as demonstrated in (2.10), while the value of information is negatively related to λ, we measure it through the following transformation of $p(r)_{jt}$:[32]

$$w(r)_{jt} = -\ln[p(r)_{jt}] \tag{2.18}$$

To give concrete examples, we have that 97.0 percent of flows from Mexico between 1990 and 1995 were directed to the United States, so that $w(5)_{\text{MEX1995}} = -\ln(0.970) = 0.031$. Over the same period, 25.4 percent of migration flows from China were directed to the main destination (United States), and this implies that $w(5)_{\text{CHN1995}} = -\ln(0.254) = 1.371$. Thus, the empirical counterpart of the value of information in (2.18) suggests that Chinese migrants valued information more than Mexican migrants in the five-year period starting in 1995.

Table 2.1: Descriptive statistics for the empirical counterparts of the value of information

	mean	s.d.	min	max	obs.
$w(5)_{jt}$	0.86	0.53	0.00	2.49	257,086
$w(10)_{jt}$	0.92	0.52	0.00	2.40	260,332
$w(15)_{jt}$	0.95	0.52	0.00	2.53	261,668
$w(20)_{jt}$	0.96	0.52	0.00	2.47	261,858

Notes: $w(r)_{jt}$, with $r = \{5,10,15,20\}$, computed according to (2.18).

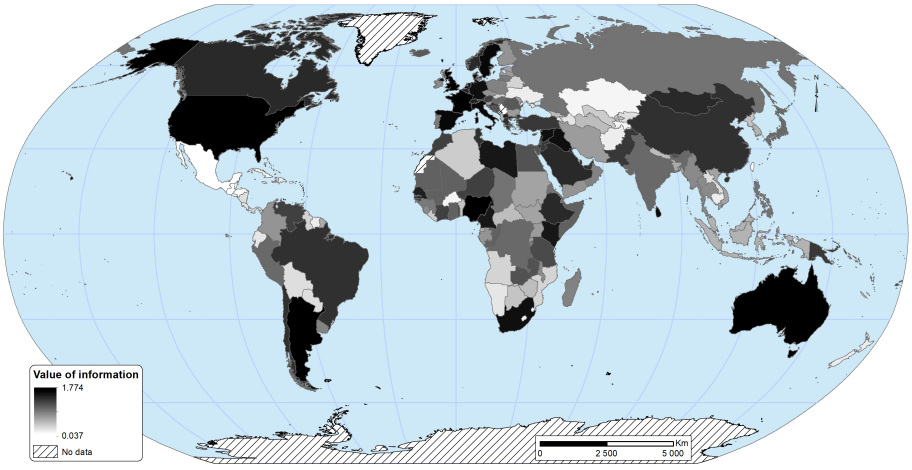
Source: Authors' elaboration on Abel (2018).

Going beyond specific examples, Table 2.1 reports the descriptive statistics for $w(r)_{jt}$, with $r = \{5,10,15,20\}$. The average value of the empirical counterpart of the value of information monotonically increases with r, from 0.86 for $w(5)_{jt}$ to 0.96 for $w(20)_{jt}$, as the share of migrants from j directed to the main destination declines with the length of the period over which we measure past migration flows. When we increase the length r of the time period over which we measure past migration flows, we get closer to the objective of obtaining a proxy for the *un*conditional probability of selecting the main alternative, but we also run the risk of introducing noise that is due to changes in the attractiveness of the various destinations; hence, it is important to test the robustness of our empirical evidence when cumulating past flows over different periods. Notice that, when r increases, the ensuing variation in $p(r)_{jt}$ can also reflect the change in the main destination: when we move from 5 to 10 years, we observe such

[32] This specific functional form is immaterial for the evidence that we present in Sections 2.4.2 and 2.4.3, which is robust to interacting GDP per capita at destination with $p(r)_{jt}$, or with $1/p(r)_{jt}$; results are available from the Authors upon request.

a switch for 331 out of 1,347 origin-period pairs, and the corresponding figures for 15 and 20 years stand at 459 and 521 origin-period pairs. Nevertheless, the four variants of the empirical counterparts of the value of information are closely correlated: the correlations range between 0.58 (between $w(5)_{jt}$ and $w(20)_{jt}$) and 0.93 ($w(15)_{jt}$ and $w(20)_{jt}$). For $w(5)_{jt}$, the observed values for $w(5)_{jt}$ range between 0 and 2.49, as reported in Table 2.1, thus covering a substantial portion of the range of values that are theoretically feasible.[33] The variability in $w(5)_{jt}$ reflects both time-invariant differences across origins, as well as within-origin differences over time. More precisely, a regression of $w(5)_{jt}$ on a set of origin dummies explains 40.4 percent of its variability. Beyond differences in λ, time-invariant heterogeneity across origins in $w(r)_{jt}$ might also capture the effects of geography, e.g., proximity to a high-income country increases the concentration of migration flows, while its within-origin variability might reflect as well variations in (observed or unobserved) determinants of the attractiveness or accessibility of major destinations.

Figure 2.2: Origin-specific average of the value of information $w(5)_{jt}$



Source: Authors' elaboration on Abel (2018).

Figure 2.2 plots the origin-specific average of the value of information $w(5)_{jt}$ between 1980 and 2015 on a world map, revealing that there is no clear geographical pattern in the data, with a substantial variability in the value of $w(5)_{jt}$ within, say, Latin America or Sub-Saharan Africa. Figure 2.2 also reveals that high-income countries in Western Europe, North America and Oceania are typically characterized by a high average value of $w(5)_{jt}$, a pattern that will

[33] The upper bound of the value of information stands at $-\ln 1/N = \ln 184 \approx 5.2$ when $N = 184$.

be taken into account in the econometric analysis.

2.4 Econometric analysis

Our objective is to test the empirical relevance of information frictions in shaping migration decisions. To this end, we bring to the data a theory-based specification of the migration gravity equation where we introduce an interaction between the empirical counterpart of the value of information and income per capita at destination.

2.4.1 Gravity equation with rational inattention

We can write the migration flows m_{jkt} between an origin j and a destination k in the five-year period starting in year t as:

$$m_{jkt} = \mathcal{P}_{jkt} \times n_{jt} \times \zeta_{jkt} \tag{2.19}$$

where $n_{jt} = \sum_{k \in A} m_{jkt}$, $\zeta_{jkt} > 0$ is an error term, and the probability $\mathcal{P}_{jkt}$ that destination k represents the utility-maximizing alternative for a migrant from j in period t is given by (2.12). Replacing $\mathcal{P}_{jkt}$, we can then rewrite equation (2.19) as:

$$m_{jkt} = \exp\left[\frac{1}{\lambda_{jt}}\epsilon_{kt} + \frac{1}{\lambda_{jt}(1-\lambda_{jt})}\overline{v}_{kt} + \Omega_{jt} + \ln(\zeta_{jkt})\right] \tag{2.20}$$

where:

$$\Omega_{jt} \equiv \ln(n_{jt}) - \sum_{a \in A} e^{\frac{\epsilon_{jat}}{\lambda_{jt}} + \frac{\overline{v}_{jat}}{\lambda_{jt}(1-\lambda_{jt})}}$$

We assume that the destination-specific utility $v_{jkt} = \overline{v}_{jkt} + \epsilon_{jkt}$ follows:

$$v_{jkt} = \alpha \ln\left(\frac{y_{kt}}{\tau_{jkt}}\right) \tag{2.21}$$

where y_{kt} is real GDP per capita in destination k in year t, and $\tau_{jkt} \geq 1$ are dyadic and time-varying iceberg migration costs. The specification in (2.21) implies that the semi-elasticity of v_{jkt} with respect to y_{kt} is always equal to α, and independent of the value of the determinants of dyadic migration costs τ_{jkt}. We further assume that $\overline{v}_{jkt} = -\alpha \ln \tau_{jkt}$ and $\epsilon_{jkt} = \alpha \ln y_{kt}$, i.e., migrants can observe the determinants of the accessibility of destination k, but are unable to costlessly observe local economic conditions. These assumptions allow rewriting (2.20) as follows:

$$m_{jkt} = \exp\left[\frac{\alpha}{\lambda_{jt}} \ln y_{kt} - \frac{\alpha}{\lambda_{jt}(1-\lambda_{jt})} \ln \tau_{jkt} + \Omega_{jt} + \ln(\zeta_{jkt})\right] \tag{2.22}$$

The specification that we bring to the data is given by:

$$m_{jkt} = \exp\left[\beta\left(w(r)_{jt} \times \ln y_{kt}\right) + d_{kt} + d_{jt} + d_{jk} + \varepsilon_{jkt}\right] \tag{2.23}$$

where $w(r)_{jt}$, with $r = \{5, 10, 15, 20\}$, represents, as discussed in Section 2.3.2, an empirical proxy for $1/\lambda_{jt}$; $\ln y_{kt}$ is the logarithm of GDP per capita in 2010 USD from World Bank (2018);[34] d_{kt}, d_{jt}, and d_{jk} represent destination-time, origin-time and origin-destination (dyadic) dummies; and ε_{jkt} is the error term. Since we have a large share of zeros (61.2 percent) in our dependent variable m_{jkt}, we estimate (2.23) using a Poisson pseudo-maximum-likelihood estimator, following Santos Silva and Tenreyro (2006). More precisely, we employ the Stata command `ppmlhdfe` developed by Correia et al. (2019, 2020), which allows handling in a computationally efficient way the large number of fixed effects in (2.23). Standard errors are clustered at the origin level following Bertrand et al. (2004).

The inclusion of origin-time dummies in (2.23) perfectly controls for Ω_{jt} in (2.20). The rich structure of fixed effects allows controlling for the dependence of the iceberg-type migration costs τ_{jkt} on dyadic time-invariant factors such as geographical distance, linguistic and cultural proximity, or on destination-time specific factors, such as policy-induced barriers to migrations. However, (2.22) reveals that the effect of $\ln \tau_{jkt}$ on bilateral migration flows is also mediated by λ_{jt}, so this confounding effect is potentially specific to each origin-destination-time triplet. We pursue two different and not mutually exclusive approaches to control for it: first, we augment the specification in (2.23) by interacting typical correlates of dyadic migration costs from Mayer and Zignago (2011) with $w(r)_{jt}$; second, we also control for $\ln(s_{jkt} + 1)$, i.e., the logarithm of (one plus) the stock of j-born migrants residing in destination k in year t, as in Beine et al. (2011).[35,36]

Our estimate for β will be consistent as long as (*i*) the effect of immigration flows from one particular origin on one particular destination is close to zero, and (*ii*) our proxy for information costs is both predetermined and persistent enough. Under (*i*) and (*ii*), there will be no simultaneity between our dependent variable, migration flows, GDP per capita and our empirical value of information. Condition (*i*) is likely to be satisfied. For example, the median migration flow in our dataset amounts to 0.003 per cent of the destination country

[34] The average and standard deviation of $\ln y_{kt}$ over our sample stand at 8.24 and 1.53 respectively.

[35] The data on the bilateral stock s_{jkt} comes from Özden et al. (2011) between 1960 and 2000, with interpolated values in between census years, and from United Nations Population Division (2015a) since 2005; the average and standard deviation of $\ln(s_{jkt} + 1)$ over our sample stand at 2.25 and 2.95 respectively.

[36] The econometric evidence is fully robust when relying on the inverse hyperbolic sine transformation of s_{jkt} to account for zeros in bilateral migrant stocks, or when also interacting the measure of networks with $w(r)_{jt}$; results are available from the Authors upon request.

population in a particular five-year period. According to the findings in Ortega and Peri (2014), this would translate into an increase in the GDP per capita of the typical destination country of 0.02 per cent over five years, that is, barely 0.004 per cent per year. As far as (*ii*) is concerned, the condition is clearly satisfied in theory, as λ is a parameter that determines migration flows. When it comes to our empirical proxy, $w(r)_{jt}$ is calculated on past migration flows and we experiment with different values of r precisely to make sure that our results hold under different notions of persistence. Furthermore, it must be emphasized that the empirical value of information $w(r)_{jt}$ is not a lagged version of the dependent variable. Recall that our dependent variable includes variation at the origin-destination-time level (jkt) while the value of information is origin-time specific (jt). All past flows out of an origin enter into the computation of $w(r)_{jt}$ but most of its variation corresponds to the main destination out of 189 in our dataset. This is why we show below that our results are robust to dropping the main destination, for which the lagged dependent variable would create a problem. Still, any remaining auto-correlation should be taken into account by our clustering of standard errors at the origin level.

2.4.2 Main results

Table 2.2 reports our baseline results for the gravity equation described in (2.23). Each column corresponds to one of the four variants of the empirical counterpart for the value of information $w(r)_{jt}$, with $r = \{5,10,15,20\}$, for the origin country j in the five-year period starting in year t. The estimates reveal that the coefficient $\widehat{\beta}$ of the interaction between GDP per capita at destination and the time-varying origin-specific value of information is always positive and significant at conventional confidence levels.[37] A one standard deviation increase in the value of $w(r)_{jt}$ is associated with an increase in the elasticity of the bilateral migration rate with respect to GDP per capita at destination ranging between 0.072, in column (1), and 0.093, in column (4). Going back to the example of China and Mexico that we introduced in Section 2.3.2, the estimates in Table 2.2 imply that the elasticity for migration from China to any destination between 1995 and 2000 was 0.182-0.241 higher than the corresponding elasticity for migration from Mexico over the same time period. Similarly, the estimates also imply a substantial variability over time for a given origin; for instance, the elasticity of migration out of Ecuador increased by 0.078-0.104 between the early 1980s and the early 2000s,[38] following a substantial diversification of the main destinations for Ecuadorian migrants (Bertoli et al.,

[37] Our analysis is fully robust to using gender-specific bilateral migration flows from Abel (2018); results are available from the Authors upon request.

[38] The value of $w(5)_{\text{ECU1980}}$ stood at 0.168, increasing to $w(5)_{\text{ECU2000}} = 0.744$.

2011).

Table 2.2: Baseline results

	Dependent variable: m_{jkt}			
	(1)	(2)	(3)	(4)
Value of r	5	10	15	20
$\ln(y_{kt}) \times w(r)_{jt}$	0.136^{**}	0.166^{**}	0.163^{**}	0.180^{*}
	(0.058)	(0.066)	(0.080)	(0.101)
Observations	221,342	224,184	225,327	225,458
Pseudo-R^2	0.962	0.962	0.962	0.962
$w(r)_{jt}$ (mean)	0.863	0.922	0.950	0.965
$w(r)_{jt}$ (s.d.)	0.533	0.524	0.519	0.517
d_{jt}, d_{kt} and d_{jk}	Yes	Yes	Yes	Yes

Notes: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$; clustered standard errors at the origin level in parentheses.
Source: Authors' elaboration on Abel (2018) and World Bank (2018).

We next show how our main results are affected when we perform some particular variations in the estimated specifications, and how they consolidate the interpretation that the value of information picks up information frictions in the way that our simple model describes.

2.4.2.1 Dropping the main destination from the estimation sample

First, we address the concern described at the end of Section 2.4.1. Since the value of information is constructed using lags of the dependent variable for all destinations, we check whether our results are robust to dropping the main origin-time specific destination from the estimation sample, as past flows to this specific country pick up, by construction, most of the variation in the value of information. The exercise is performed in Table 2.3. We can see that our estimate for β decreases in size for all four definitions of the value of information. Still, the coefficients remain significant at conventional levels (and more precisely estimated), and statistically identical to our main results in Table 2.2.

Table 2.3: Results excluding the main destination from the sample

	Dependent variable: m_{jkt}			
	(1)	(2)	(3)	(4)
Value of r	5	10	15	20
$\ln(y_{kt}) \times w(r)_{jt}$	0.094^{***}	0.117^{***}	0.122^{***}	0.108^{*}
	(0.033)	(0.034)	(0.042)	(0.055)
Observations	220,088	222,912	224,046	224,180
Pseudo-R^2	0.954	0.954	0.955	0.956
$w(r)_{jt}$ (mean)	0.863	0.922	0.950	0.965
$w(r)_{jt}$ (s.d.)	0.533	0.524	0.519	0.517
d_{jt}, d_{kt} and d_{jk}	Yes	Yes	Yes	Yes

Notes: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$; clustered standard errors at the origin level in parentheses.

Source: Authors' elaboration on Abel (2018) and World Bank (2018).

Besides showing that the results are not mechanically generated by lagged migration flows, the estimates in Table 2.3 prove that ours is not a story about Mexican migration flows to the United States being less responsive to economic conditions in the United States. Mexican migration flows are less responsive to economic conditions also in other destinations than, for example, Chinese emigration flows.

2.4.2.2 Using the past share of flows to the the second destination

The analysis of the theoretical model has revealed that only the unconditional probability of opting for the main destination is monotonically related to $1/\lambda_{jt}$, while the relationship of this key parameter of the model with the unconditional probabilities for other destinations is ambiguous. Thus, we define an alternative measure $w_2(r)_{jt} \equiv -\ln[p_2(r)_{jt}]$, where $p_2(r)_{jt}$ is the share of migrants to the second main destination rather than the share of migrants to the top destination in the past r years. We interact this alternative measure with the log of GDP per capita at destination in Table 2.4. This change in our variable of interest renders all of our estimated coefficients statistically insignificant. These results (or, rather, this lack of results) is fully consistent with our theoretical model.[39]

[39] Similar evidence is obtained when using the past share of flows directed to the third, fourth or fifth destination; results are available from the Authors upon request.

Table 2.4: Measuring the value of information with the second main destination

	Dependent variable: m_{jkt}			
	(1)	(2)	(3)	(4)
Value of r	5	10	15	20
$\ln(y_{kt}) \times w_2(r)_{jt}$	-0.019	-0.109	-0.061	-0.045
	(0.030)	(0.067)	(0.084)	(0.085)
Observations	219,079	223,547	224,669	225,144
Pseudo-R^2	0.963	0.962	0.962	0.962
$w(r)_{jt}$ (mean)	2.061	2.034	2.035	2.032
$w(r)_{jt}$ (s.d.)	0.848	0.689	0.625	0.598
d_{jt}, d_{kt} and d_{jk}	Yes	Yes	Yes	Yes

Notes: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$; clustered standard errors at the origin level in parentheses. $w_2(r)_{jt}$ is equal to minus the logarithm of the share of past flows directed towards the second main destination.
Source: Authors' elaboration on Abel (2018) and World Bank (2018).

2.4.2.3 Controlling more thoroughly for migration costs

We advanced above, when discussing equation (2.22), two different strategies to control for the influence of migration costs τ_{jkt} on migration flows, as their effect was also mediated by information costs λ_{jt}. First, Table 2.5 expands our preferred specification by controlling for the interaction between the value of information and the classic dyadic time-invariant gravity determinants of migration flows: contiguity between j and k, the existence of a common language between j and k, whether j and k ever had a common colonial link, and the logarithm of the geodesic distance between j and k. Most of the interactions of these added variables with the value of information turn out not to be significant. We have one significant positive interaction of distance out of four and two marginally significant negative interactions of the colony variable. In contrast, our interaction of interest between the value of information and GDP per capita at destination remains positive and significant and, while all the coefficients go down in size with respect to our baseline in Table 2.2, the differences between both sets of coefficients are not statistically significant.

Table 2.5: Interactions of dyadic variables with $w(r)_{jt}$

	Dependent variable: m_{jkt}			
	(1)	(2)	(3)	(4)
Value of r	5	10	15	20
$\ln(y_{kt}) \times w(r)_{jt}$	0.092**	0.131***	0.119*	0.167**
	(0.045)	(0.050)	(0.063)	(0.081)
$w(r)_{jt} \times$ Contiguity$_{jk}$	-0.107	-0.212	-0.238	0.025
	(0.153)	(0.154)	(0.189)	(0.279)
$w(r)_{jt} \times$ Common language$_{jk}$	0.098	-0.093	0.035	0.033
	(0.091)	(0.101)	(0.103)	(0.146)
$w(r)_{jt} \times$ Colony$_{jk}$	-0.192*	-0.157	-0.256*	-0.230
	(0.111)	(0.134)	(0.141)	(0.162)
$w(r)_{jt} \times \ln($distance$_{jk})$	0.109***	0.033	0.061	0.059
	(0.035)	(0.041)	(0.050)	(0.073)
Observations	214,838	217,518	218,654	218,785
Pseudo-R^2	0.963	0.962	0.962	0.962
$w(r)_{jt}$ (mean)	0.867	0.928	0.957	0.972
$w(r)_{jt}$ (s.d.)	0.531	0.524	0.520	0.517
d_{jt}, d_{kt} and d_{jk}	Yes	Yes	Yes	Yes

Notes: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$; clustered standard errors at the origin level in parentheses.

Source: Authors' elaboration on Abel (2018), World Bank (2018) and Mayer and Zignago (2011).

Second, in Table 2.6, we augment our baseline specification with the variable $\ln(s_{jkt} + 1)$, the logarithm of (one plus) the stock of j-born migrants residing in destination k in year t, as in Beine et al. (2011). This serves two purposes. On the one hand, it allows us to control directly for an observable factor that has been shown to be relevant in affecting migration costs (McKenzie and Rapoport, 2010). On the other hand, it shows that our value of information is not picking up omitted network effects. While the coefficient for the stock of previous migrants from the same destination is positive and highly significant in all specifications, our estimated $\widehat{\beta}$ also remains positive and significant, and close to our baseline estimates in Table 2.2.

Table 2.6: Baseline results on the value of information with networks

	Dependent variable: m_{jkt}			
	(1)	(2)	(3)	(4)
Value of r	5	10	15	20
$\ln(y_{kt}) \times w(r)_{jt}$	0.119^{**}	0.145^{**}	0.144^{**}	0.161^{*}
	(0.051)	(0.057)	(0.070)	(0.089)
$\ln(s_{jkt}+1)$	0.193^{***}	0.192^{***}	0.195^{***}	0.196^{***}
	(0.032)	(0.031)	(0.031)	(0.031)
Observations	220,627	223,469	224,612	224,743
Pseudo-R^2	0.963	0.963	0.963	0.963
$w(r)_{jt}$ (mean)	0.866	0.925	0.953	0.968
$w(r)_{jt}$ (s.d.)	0.532	0.523	0.518	0.515
d_{jt}, d_{kt} and d_{jk}	Yes	Yes	Yes	Yes

Notes: *** $p<0.01$, ** $p<0.05$, * $p<0.10$; clustered standard errors at the origin level in parentheses.

Source: Authors' elaboration on Abel (2018), World Bank (2018), Özden et al. (2011) and United Nations Population Division (2015a).

In Table 2.7, we also interact the network variable with our variable for the value of information, as equation (2.22) suggests that any component of migration costs will have its effect on migration flows mediated through information costs. Table 2.7 shows that the interaction between the network variable and the value of information is not significantly different from zero. This is not surprising considering that the elasticity of migration flows with respect to migration costs, while depending on λ_{jt}, is not monotonic in λ_{jt}. On the contrary, the elasticity of migration flows with respect to GDP per capita at destination is monotonically related to λ_{jt} and this is reflected in the positive and statistically significant coefficient $\widehat{\beta}$ in all specifications in Table 2.7. Again, these coefficients are not statistically different from those reported in the baseline. Our results are fully robust when we put together both strategies for more thoroughly controlling for migration costs, that is, when we combine Tables 2.5 and 2.7 and include both the network variable, its interaction with the value of information and the interactions of the value of information with time-invariant dyadic variables.[40]

[40] Results are available from the Authors upon request.

Table 2.7: Interacting networks with the value of information

	Dependent variable: m_{jkt}			
	(1)	(2)	(3)	(4)
Value of r	5	10	15	20
$\ln(y_{kt}) \times w(r)_{jt}$	0.122**	0.146**	0.144**	0.157*
	(0.051)	(0.058)	(0.072)	(0.091)
$\ln(s_{jkt}+1) \times w(r)_{jt}$	-0.017	-0.011	-0.001	0.023
	(0.017)	(0.021)	(0.024)	(0.028)
$\ln(s_{jkt}+1)$	0.207***	0.201***	0.196***	0.175***
	(0.038)	(0.042)	(0.044)	(0.047)
Observations	220,627	223,469	224,612	224,743
Pseudo-R^2	0.964	0.963	0.963	0.963
$w(r)_{jt}$ (mean)	0.866	0.925	0.953	0.968
$w(r)_{jt}$ (s.d.)	0.532	0.523	0.518	0.515
d_{jt}, d_{kt} and d_{jk}	Yes	Yes	Yes	Yes

Notes: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$; clustered standard errors at the origin level in parentheses.

Source: Authors' elaboration on Abel (2018), World Bank (2018), Özden et al. (2011) and United Nations Population Division (2015a).

2.4.2.4 Do migrants form consideration sets?

Our theoretical model implies that all choice probabilities should be strictly positive. Under the assumptions that we needed to invoke to provide an analytical solution for our model there would be no zero flows. However, empirically we observe that 61.2 per cent of observations correspond to zero migration flows over a five-year period. Discrete-choice models of rational inattention can lead, under alternative distributional assumptions, to the formation of consideration sets that are strictly smaller than the choice set (Caplin et al., 2019).

In this spirit, let $d^{\text{zero}}(5)_{jkt}$ be a dummy signaling a zero migration from j to k in the five years up to year t. We have that 60.4 percent of the observations in our sample correspond to origin-destination dyads with a zero flow in the recent past. Notice that we do not even use 37 per cent of these for identification since they correspond to origin-destination pairs where the flows are always zero in our baseline sample. Still, we would not want our result on the value of information, derived from a model where zero flows are not possible, to be affected by these zero-flow observations. Intuitively, the migration flows for these dyads could be less sensitive to variations in economic conditions in the various destination countries, as migrants from j could exclude destination k from their (time-varying) consideration sets when $d^{\text{zero}}(5)_{jkt} = 1$.

Table 2.8: Zero past flows reduce current responsiveness

	Dependent variable: m_{jkt}		
	(1)	(2)	(3)
Value of r	5	5	5
$\ln(y_{kt}) \times d^{\text{zero}}(r)_{jkt}$	-0.039***	-0.037***	
	(0.007)	(0.007)	
$\ln(y_{kt}) \times w(r)_{jt}$		0.132**	0.136**
		(0.059)	(0.058)
Observations	221,342	221,342	221,342
Pseudo-R^2	0.962	0.963	0.962
$w(r)_{jt}$ (mean)	0.863	0.863	0.863
$w(r)_{jt}$ (s.d.)	0.533	0.533	0.533
d_{jt}, d_{kt} and d_{jk}	Yes	Yes	Yes

Notes: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$; clustered standard errors at the origin level in parentheses.
Source: Authors' elaboration on Abel (2018) and World Bank (2018).

Table 2.8 confirms that this is indeed the case: the elasticity with respect to GDP per capita at destination is 0.039 points lower for origin-destination dyads characterized by zero flows over the previous five years. However, this does not explain the role played by the value of information in our baseline results, as our coefficient of interest is only marginally reduced when introducing the additional interaction between $d^{\text{zero}}(5)_{jkt}$ and $\ln(y_{kt})$, as a comparison of the second and of the third data columns in Table 2.8 reveals. This also applies when using data over the previous 10, 15 or 20 years to identify origin-destination pairs with past zero flows, or when we define a relative or an absolute threshold higher than zero to identify minor destinations.[41]

2.4.3 Threats to our interpretation

The econometric evidence presented in Section 2.4.2 above is consistent with the testable implications laid out in Section 2.2.6, but we need to understand whether they could also be generated by a canonical full-information model, or by a full-information model with a richer and more flexible specification of location-specific utility.

What would it happen if migrants were able to costlessly observe location-specific utilities before deciding where to move? A random utility maximization model with distributional

[41] Results are available from the Authors upon request.

assumptions *à la* McFadden, and where the variance of the stochastic component of utility is origin-specific, also implies a systematic relationship between the distribution of migrants across destinations and the responsiveness of bilateral migration flows with respect to variations in economic conditions of the various destinations.[42] More precisely, origin countries with a greater preference heterogeneity will have migration flows that are both (*i*) more dispersed across destinations, and (*ii*) less responsive to changes in economic conditions. This, in turn, implies that a canonical full information model generates the testable implication that the coefficient of the interaction between $w(r)_{jt}$ and $\ln y_{kt}$ should be negative, a prediction that is clearly rejected by the data.

The pattern that we uncover in the data might be explained by a more flexible version of the full-information model. For instance, migration decisions could be subject to binding liquidity constraints, which could influence migrants' ability to respond to variations in economic conditions even though they are able to costlessly observe them. Furthermore, location-specific utility might not be additively separable in y_{kt} and in τ_{jkt} (an assumption that we have retained so far), so that the semi-elasticity of v_{jkt} with respect to y_{kt} could be a function of the determinants of dyadic migration costs τ_{jkt}, e.g., the marginal utility of income might be a function of dyadic migration costs, or it might depend on migrants' individual characteristics such as education.[43]

2.4.3.1 Liquidity constraints

The empirical counterpart $w(5)_{jt}$ for the value of information is higher in some geographical areas where most high-income countries are concentrated (see Figure 2.2 in Section 2.3.2). If we rely on the classification by income groups from the World Bank, we have that the average value of $w(5)_{jt}$ is equal to 1.088 for high-income origin countries, and to 0.805 for the other origin countries.[44] Migration decisions can be subject to binding liquidity constraints, as shown notably by Clemens (2014), Angelucci (2015), Djajic et al. (2016), Bazzi (2017) or Dao et al. (2018).

[42] The full analysis of this model is presented in the Appendix A.2.1.

[43] Porcher (2019) provides empirical evidence that bilateral migration flows respond more to economic conditions in the various destinations for closer origin-destination pairs.

[44] The classification by the World Bank is available on an yearly basis since 1989; we use the classification for year t since 1990, and the earliest available classification for previous years for each origin; source: datahelpdesk.worldbank.org/knowledgebase/articles/906519-world-bank-country-and-lending-groups (last accessed on January 22, 2019).

Table 2.9: Heterogeneity by income group

	Dependent variable: m_{jkt}			
	(1)	(2)	(3)	(4)
Value of r	5	10	15	20
$\ln(y_{kt}) \times w(r)_{jt}$	0.136**	0.162**	0.159*	0.180*
	(0.059)	(0.067)	(0.081)	(0.104)
$\ln(y_{kt}) \times d_{jt}^{\text{low}}$	-0.163	-0.136	-0.137	-0.124
	(0.124)	(0.124)	(0.128)	(0.133)
$\ln(y_{kt}) \times d_{jt}^{\text{l. middle}}$	-0.206*	-0.180*	-0.194*	-0.189*
	(0.110)	(0.109)	(0.112)	(0.115)
$\ln(y_{kt}) \times d_{jt}^{\text{u. middle}}$	-0.051	-0.039	-0.057	-0.047
	(0.085)	(0.085)	(0.086)	(0.090)
Observations	216,742	219,584	220,727	220,858
Pseudo-R^2	0.962	0.962	0.962	0.962
$w(r)_{jt}$ (mean)	0.864	0.929	0.962	0.978
$w(r)_{jt}$ (s.d.)	0.528	0.521	0.516	0.513
d_{jt}, d_{kt} and d_{jk}	Yes	Yes	Yes	Yes

Notes: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$; clustered standard errors at the origin level in parentheses.

Source: Authors' elaboration on Abel (2018) and World Bank (2018).

Liquidity constraints imply that the set of affordable destinations is smaller than the choice set (Marchal and Naiditch, 2020), and hence this pattern in the data poses a threat to our interpretation of the results in Table 2.2. Migrants from lower-income countries might not value information less, but they might be less able to react to variations in economic conditions, and their past distribution could be more concentrated in the main (affordable) destination. We thus estimate an extended version of the gravity equation in (2.23), where we allow for a heterogeneous effect of $\ln(y_{kt})$ across groups of origins with a different level of income. Table 2.9 reveals that the elasticity of the migration rate with respect to y_{kt} is higher for origins classified as high-income countries in year t (the omitted category), albeit these differences are not precisely identified.[45] However, this does not influence either the size or the significance of the coefficient for our interaction effect, thus dismissing the concern that the values of $\widehat{\beta}$ in Table 2.2 were picking up a spurious correlation between $w(r)_{jt}$ and the income group to which the origin j belonged in year t.[46]

[45] Notice that liquidity constraints can hinder the ability of migrants to react to an increase in the attractiveness of a country, but they do not limit their ability to react to worsening economic conditions.

[46] We obtain similar results when considering a time-invariant income classification, or when introducing an interaction between $\ln(y_{kt})$ and $\ln(y_{jt})$; results are available from the Authors upon request.

2.4.3.2 More flexible responsiveness to economic conditions

Do the results presented in Table 2.2 survive once we allow for a more general functional form of the deterministic component of utility v_{jkt}, or for differences across destinations or at the dyadic level in the cost of acquiring information, thus relaxing the assumption that λ does not vary across alternatives in the choice set? For instance, one could plausibly imagine that migrants from countries with larger past migration flows, with stronger networks at destination or facing lower moving costs could more easily acquire information on the attractiveness of the alternative destinations.[47] We address this relevant empirical concern introducing an additional interaction term, between $\ln(y_{kt})$ and the logarithm of the total emigration rate for the origin j in the r years up to year t, with r taking the same value that is used to measure the value of information $w(r)_{jt}$. The estimated coefficient for this additional interaction term is always positive, and significant in three out of four data columns in Table 2.10, in line with the idea that larger past migration flows reduce the cost of acquiring information on the attractiveness of the alternative destinations.[48] However, the inclusion of the additional interaction term only marginally influences the size of the estimated value of $\widehat{\beta}$, and it leaves its significance unchanged. The estimated coefficients for the interaction between economic conditions at destination and value of information at origin range between 0.136 and 0.180 in Table 2.2, and between 0.112 and 0.193 in Table 2.10.

[47] The empirical counterparts for λ are insensitive to the scale of past migration flows.

[48] An alternative, but not mutually exclusive, explanation is that migrants' remittances help relaxing liquidity constraints at origin, thus increasing, as suggested by Table 2.10, the responsiveness of bilateral migration flows with respect to varying economic conditions.

Table 2.10: Interaction with the past emigration rate

	Dependent variable: m_{jkt}			
	(1)	(2)	(3)	(4)
Value of r	5	10	15	20
$\ln(y_{kt}) \times w(r)_{jt}$	0.112**	0.158**	0.151**	0.193**
	(0.050)	(0.062)	(0.073)	(0.097)
$\ln(y_{kt}) \times \ln[\text{emigration rate}(r)_{jt}]$	0.021**	0.017	0.048**	0.077***
	(0.009)	(0.024)	(0.020)	(0.024)
Observations	214,260	216,390	216,805	216,570
Pseudo-R^2	0.963	0.963	0.963	0.963
$w(r)_{jt}$ (mean)	0.866	0.931	0.966	0.982
$w(r)_{jt}$ (s.d.)	0.528	0.520	0.516	0.513
d_{jt}, d_{kt} and d_{jk}	Yes	Yes	Yes	Yes

Notes: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$; clustered standard errors at the origin level in parentheses.

Source: Authors' elaboration on Abel (2018) and World Bank (2018).

Similarly, our empirical evidence is robust when interacting $\ln(y_{kt})$ with the (logarithm of the) size of the network of j-born migrants residing in destination k in year t, as shown in Table 2.11. Interestingly, the coefficient of this additional interaction term is negative and significant, suggesting that migration flows directed to destinations with larger diasporas from a given origin are less responsive to the varying attractiveness of those destinations. This might reflect the relevance of flows related to family reunification provisions, which are likely to be less responsive to business cycle conditions at destination.

Table 2.12 similarly extends the gravity equation in (2.23) by introducing (either separately or jointly) interactions between the canonical dyadic controls from Mayer and Zignago (2011) and $\ln y_{kt}$: origin-destination pairs with lower dyadic migration costs, e.g., contiguous countries, are characterized by a greater responsiveness of bilateral migration flows with respect to economic conditions.[49] In particular, the interaction between $\ln y_{kt}$ and the geodesic distance between the origin j and the destination k is negative and significant. However, this does not influence the estimated coefficient for $\ln(y_{kt}) \times w(r)_{jt}$, which ranges between 0.141 and 0.181, perfectly in line with the 0.136-0.180 range for $\widehat{\beta}$ from Table 2.2.[50]

[49] This result could be of independent interest with respect to the reliance on the estimation of a zero-stage gravity equation with dyadic time-invariant correlates of migration costs to generate an instrument for observed immigration (see, for instance, Ortega and Peri, 2014 and Alesina et al., 2016).

[50] Additional results, which are available from the Authors upon request, reveal that our empirical evidence is also robust to interacting $\ln y_{kt}$ with various measures of cultural and linguistic proximity between the origin j and the destination k from Spolaore and Wacziarg (2016) and Adserà and Pytliková (2015).

Table 2.11: Heterogeneity with respect to the size of bilateral networks

	Dependent variable: m_{jkt}			
	(1)	(2)	(3)	(4)
Value of r	5	10	15	20
$\ln(y_{kt}) \times w(r)_{jt}$	0.106**	0.125***	0.120**	0.129*
	(0.042)	(0.047)	(0.060)	(0.075)
$\ln(y_{kt}) \times \ln(s_{jkt}+1)$	-0.070***	-0.066***	-0.066***	-0.066***
	(0.018)	(0.018)	(0.018)	(0.017)
$\ln(s_{jkt}+1)$	0.845***	0.802***	0.808***	0.807***
	(0.186)	(0.181)	(0.184)	(0.180)
Observations	220,627	223,469	224,612	224,743
Pseudo-R^2	0.964	0.964	0.964	0.963
$w(r)_{jt}$ (mean)	0.866	0.925	0.953	0.968
$w(r)_{jt}$ (s.d.)	0.532	0.523	0.518	0.515
d_{jt}, d_{kt} and d_{jk}	Yes	Yes	Yes	Yes

Notes: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$; clustered standard errors at the origin level in parentheses.

Source: Authors' elaboration on Abel (2018), World Bank (2018), Özden et al. (2011) and United Nations Population Division (2015a).

The stability of the coefficient $\widehat{\beta}$ for our main interaction term when we allow for the elasticity to vary across groups of origins or across origin-destination pairs is also reassuring with respect to the concern that the value of information might be picking up differences across origins in the composition of international migration flows that are associated with a differential responsiveness to economic conditions at destination. For instance, tertiary educated migrants might react differently to changing economic conditions at destination, but the inclusion of additional interactions of $\ln y_{kt}$ with main origin-specific, i.e., income, or bilateral, e.g., networks, correlates of the educational composition of migration flows (see, for instance, Beine et al., 2011) allows, at least partially, to downplay this concern.

Table 2.12: Heterogeneity with respect to dyadic determinants of migration costs

	Dependent variable: m_{jkt}			
	(1)	(2)	(3)	(4)
Value of r	5	10	15	20
$\ln(y_{kt}) \times w(r)_{jt}$	0.141**	0.173***	0.170**	0.183*
	(0.058)	(0.065)	(0.078)	(0.100)
$\ln(y_{kt}) \times$ Contiguity$_{jk}$	0.276	0.261	0.220	0.208
	(0.186)	(0.186)	(0.183)	(0.184)
$\ln(y_{kt}) \times$ Common language$_{jk}$	0.163	0.171	0.198	0.195
	(0.211)	(0.206)	(0.205)	(0.206)
$\ln(y_{kt}) \times$ Colony$_{jk}$	-0.347	-0.354	-0.371	-0.357
	(0.296)	(0.295)	(0.296)	(0.299)
$\ln(y_{kt}) \times \ln($distance$_{jk})$	-0.203**	-0.213**	-0.225***	-0.222**
	(0.084)	(0.085)	(0.086)	(0.088)
Observations	214,838	217,518	218,654	218,785
Pseudo-R^2	0.963	0.963	0.962	0.962
$w(r)_{jt}$ (mean)	0.867	0.928	0.957	0.972
$w(r)_{jt}$ (s.d.)	0.531	0.524	0.520	0.517
d_{jt}, d_{kt} and d_{jk}	Yes	Yes	Yes	Yes

Notes: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$; clustered standard errors at the origin level in parentheses.

Source: Authors' elaboration on Abel (2018), World Bank (2018) and Mayer and Zignago (2011).

2.5 Concluding remarks

The insights obtained from applying the theory of rational inattention to the location-decision problem that migrants face are relevant to enhance our understanding of the determinants of international migration flows. The model delivers clear testable implications with respect to the role played by economic conditions in the various destinations in shaping incoming flows from origins that differ with respect to the value that migrants (rationally) attach to information acquisition. The theory reveals that the past distribution of origin-specific migration flows across destinations is informative about the (unknown) value of information. The econometric evidence is consistent with this testable prediction, and robust to alternative explanations derived from a model without information frictions.

Bibliography

ABEL, G. J. (2018): "Estimates of Global Bilateral Migration Flows by Gender between 1960 and 2015," *International Migration Review*, 52, 809–852.

ABEL, G. J. AND N. SANDER (2014): "Quantifying Global International Migration Flows," *Science*, 343, 1520–1522.

ADSERÀ, A. AND M. PYTLIKOVÁ (2015): "The role of language in shaping international migration," *Economic Journal*, 125, F49–F81.

ALESINA, A., J. HARNOSS, AND H. RAPOPORT (2016): "Birthplace diversity and economic prosperity," *Journal of Economic Growth*, 21, 101–138.

ANGELUCCI, M. (2015): "Migration and financial constraints: Evidence from Mexico," *Review of Economics and Statistics*, 97, 224–228.

BARTOŠ, V., M. BAUER, J. CHYTILOVÁ, AND F. MATĚJKA (2016): "Attention Discrimination: Theory and Field Experiments with Monitoring Information Acquisition," *American Economic Review*, 106, 1437–1475.

BATISTA, C. AND D. MCKENZIE (2018): "Testing Classic Theories of Migration in the Lab," Mimeo.

BAZZI, S. (2017): "Wealth Heterogeneity and the Income Elasticity of Migration," *American Economic Journal: Applied Economics*, 9, 219–255.

BEINE, M., F. DOCQUIER, AND Ç. ÖZDEN (2011): "Diasporas," *Journal of Development Economics*, 95, 30–41.

BERTOLI, S. AND J. FERNÁNDEZ-HUERTAS MORAGA (2013): "Multilateral Resistance to Migration," *Journal of Development Economics*, 102, 79–100.

BERTOLI, S., J. FERNÁNDEZ-HUERTAS MORAGA, AND F. ORTEGA (2011): "Immigration Policies and the Ecuadorian Exodus," *World Bank Economic Review*, 25, 57–76.

BERTRAND, M., E. DUFLO, AND S. MULLAINATHAN (2004): "How much should we trust differences-in-differences estimates?" *Quarterly Journal of Economics*, 119, 249–275.

BORJAS, G. J. (1987): "Self-Selection and the Earnings of Immigrants," *American Economic Review*, 77, 531–553.

BROWN, Z. Y. AND J. JEON (2020): "Endogenous Information and Simplifying Insurance Choice," Mimeo, University of Michigan.

BUNCH, D. S. AND D. M. ROCKE (2016): "Variance-component-based nested logit specifications: Improved formulation, and practical microsimulation of random disturbance terms," *Journal of Choice Modelling*, 21, 30–35.

CAPLIN, A., M. DEAN, AND J. LEAHY (2019): "Rational Inattention, Optimal Consideration Sets, and Stochastic Choice," *Review of Economic Studies*, 86, 1061–1094.

CAPLIN, A., J. LEAHY, AND F. MATĚJKA (2016): "Rational Inattention and Inference from Market Share Data," mimeo.

CARDELL, N. S. (1997): "Variance Components Structures for the Extreme-Value and Logistic Distributions with Application to Models of Heterogeneity," *Econometric Theory*, 13, 185–213.

CLEMENS, M. (2014): "Does development reduce migration?" in *International Handbook on Migration and Economic Development*, ed. by R. Lucas, Cheltenham: Edward Elgar Publishing, 152–185.

CLEMENS, M. A., C. MONTENEGRO, AND L. PRITCHETT (2019): "The Place Premium: Bounding the Price Equivalent of Migration Barriers," *Review of Economics and Statistics*, 101, 201–213.

CORREIA, S., P. GUIMARÃES, AND T. ZYLKIN (2020): "ppmlhdfe: Fast Poisson Estimation with High-Dimensional Fixed Effects," *Stata Journal*, 20, 95–115.

——— (2019): "Verifying the existence of maximum likelihood estimates for generalized linear models," ArXiv:1903.01633.

DAO, T. H., F. DOCQUIER, C. PARSONS, AND G. PERI (2018): "Migration and development: Dissecting the anatomy of the mobility transition," *Journal of Development Economics*, 132, 88–101.

DASGUPTA, K. AND J. MONDRIA (2018): "Inattentive Importers," *Journal of International Economics*, 112, 150–165.

DJAJIC, S., M. G. KIRDAR, AND A. VINOGRADOVA (2016): "Source-country earnings and emigration," *Journal of International Economics*, 99, 46–67.

DOMENCICH, T. AND D. L. MCFADDEN (1975): *Urban Travel Demand: A Behavioral Analysis*, Oxford: North-Holland Publishing.

EATON, J. AND S. KORTUM (2002): "Technology, geography, and trade," *Econometrica*, 70, 1741–1779.

FOSGERAU, M., E. MELO, A. DE PALMA, AND M. SHUM (2020): "Discrete Choice and Rational Inattention: a General Equivalence Result," *International Economic Review*, https://doi.org/10.1111/iere.12469.

GIORDANI, P. E. AND M. RUTA (2013): "Coordination failures in immigration policy," *Journal of International Economics*, 89, 55–67.

GROGGER, J. AND G. H. HANSON (2011): "Income maximization and the selection and sorting of international migrants," *Journal of Development Economics*, 95, 42–57.

MAĆKOWIAK, B., F. MATĚJKA, AND M. WIEDERHOLT (2018): "Rational Inattention, a Disciplined Behavioral Model," CEPR Discussion Papers 13243, London, CEPR.

MARCHAL, L. AND C. NAIDITCH (2020): "How Borrowing Constraints Hinder Migration: Theoretical Insights from a RUM Model," *Scandinavian Journal of Economics*, 122, 732–761.

MATĚJKA, F. AND A. MCKAY (2015): "Rational Inattention to Discrete Choices: A New Foundation for the Multinomial Logit Model," *American Economic Review*, 105, 272–298.

MAYDA, A. M. (2010): "International migration: A panel data analysis of the determinants of bilateral flows," *Journal of Population Economics*, 23, 1249–1274.

MAYER, T. AND S. ZIGNAGO (2011): "Notes on CEPII's distances measures: The GeoDist database," CEPII Working Paper No. 2011-25, Paris.

MCFADDEN, D. (1974): "Conditional logit analysis of qualitative choice behavior," in *Frontiers in Econometrics*, ed. by P. Zarembka, New York: Academic Press, 105–142.

——— (1978): "Modeling the Choice of Residential Location," in *Spatial interaction theory and planning models*, ed. by A. Karlqvist, L. Lundqvist, F. Snickars, and J. Weibull, Amsterdam: North-Holland, 75–96.

MCKENZIE, D., J. GIBSON, AND S. STILLMAN (2013): "A land of milk and honey with streets paved with gold: Do emigrants have over-optimistic expectations about incomes abroad?" *Journal of Development Economics*, 102, 116–127.

MCKENZIE, D. AND H. RAPOPORT (2010): "Self-Selection Patterns in Mexico-U.S. Migration: The Role of Migration Networks," *The Review of Economics and Statistics*, 92, 811–821.

McKenzie, D., S. Stillman, and J. Gibson (2010): "How important is selection? Experimental vs. non-experimental measures of the income gains from migration," *Journal of the European Economic Association*, 8, 913–945.

Munshi, K. (2003): "Networks in the modern economy: Mexican migrants in the US labor market," *The Quarterly Journal of Economics*, 118, 549–599.

Natenzon, P. (2019): "Random Choice and Learning," *Journal of Political Economy*, 127, 419–457.

Ortega, F. and G. Peri (2013): "The Role of Income and Immigration Policies in Attracting International Migrants," *Migration Studies*, 1, 47–74.

——— (2014): "Openness and income: The roles of trade and migration," *Journal of International Economics*, 92, 231–251.

Özden, Ç., C. R. Parsons, M. Schiff, and T. L. Walmsley (2011): "Where on Earth is Everybody? The Evolution of Global Bilateral Migration 1960-2000," *World Bank Economic Review*, 25, 12–56.

Porcher, C. (2019): "Migration with Costly Information," Mimeo, Princeton University.

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

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

Shannon, C. E. (1948): "A mathematical theory of communication," *Bell System Technical Journal*, 27, 379–423.

Shrestha, M. (2020): "Get Rich or Die Tryin': Perceived Earnings, Perceived Mortality Rates, and Migration Decisions of Potential Work Migrants from Nepal," *The World Bank Economic Review*, 34, 1–27.

Sims, C. A. (1998): "Stickiness," *Carnegie-Rochester Conference Series on Public Policy*, 49, 317–356.

——— (2003): "Implications of rational inattention," *Journal of Monetary Economics*, 50, 665–690.

Sjaastad, L. A. (1962): "The Costs and Returns of Human Migration," *Journal of Political Economy*, 70, 80–93.

SPOLAORE, E. AND R. WACZIARG (2016): "Ancestry, language and culture," in *The Palgrave Handbook of Economics and Language*, ed. by V. Ginsburgh and S. Weber, London: Palgrave McMillan, 174–211.

STEINER, J., C. STEWART, AND F. MATĚJKA (2017): "Rational Inattention Dynamics: Inertia and Delay in Decision-Making," *Econometrica*, 85, 521–553.

UNITED NATIONS POPULATION DIVISION (2015a): *Trends in International Migrant Stock: The 2015 Revision*, New York, NY: United Nations, Department of Economic and Social Affairs.

——— (2015b): *World Population Prospects: The 2015 Revision*, New York, NY: United Nations, Department of Economic and Social Affairs.

WORLD BANK (2018): *World Development Indicators*, Washington DC: the World Bank.

Appendix

A.1 Proofs

A.1.1 Simplifying the maximization problem

The objective function in the constrained maximization problem that identifies the optimal choice probabilities within the set B_k is given by:

$$\int_{\boldsymbol{v}} \lambda \ln \left[\sum_{a \in B_k} \mathcal{P}_a e^{v_a/\lambda} \right] f(\boldsymbol{v}) \, \mathrm{d}\boldsymbol{v} \tag{.1}$$

The key of the proof, which draws on Brown and Jeon (2020), rests on a result established by Domencich and McFadden (1975): in RUM models with full information and where the stochastic component of utility is i.i.d. EVT-1, we have that the expected value from the choice situation is equal to the logarithm of the sum of the exponentials of the expected value of utility in each alternative. Rewrite the objective function:

$$\begin{aligned}
\int_{\boldsymbol{v}} \lambda \ln \left[\sum_{a \in B_k} \mathcal{P}_a e^{v_a/\lambda} \right] f(\boldsymbol{v}) \, \mathrm{d}\boldsymbol{v} &= \int_{\boldsymbol{v}} \lambda \ln \left[\sum_{a \in B_k} e^{v_a/\lambda + \ln(\mathcal{P}_a)} \right] f(\boldsymbol{v}) \, \mathrm{d}\boldsymbol{v} \\
&= \lambda \mathbb{E}_{\boldsymbol{v}} \left[\ln \left(\sum_{a \in B_k} e^{v_a/\lambda + \ln(\mathcal{P}_a)} \right) \right] \\
&= \lambda \mathbb{E}_{\boldsymbol{\epsilon}} \left[\ln \left(\sum_{a \in B_k} e^{\overline{v}_a/\lambda + \ln(\mathcal{P}_a) + \epsilon_a/\lambda} \right) \right] \\
&= \lambda \mathbb{E}_{\boldsymbol{\epsilon},\boldsymbol{\eta}} \left[\max_{a \in B_k} \left(\overline{v}_a/\lambda + \ln(\mathcal{P}_a) + \epsilon_a/\lambda + \eta_a \right) \right]
\end{aligned}$$

where η_a is i.i.d EVT-1. If ϵ_a follows a $C(\lambda)$ distribution, then $\epsilon'_a \equiv \epsilon_a + \lambda\eta_a$ follows an EVT-1 distribution with scale parameter equal to 1. This entails that:

$$\begin{aligned}
\int_{\boldsymbol{v}} \lambda \ln \left[\sum_{a \in B_k} \mathcal{P}_a e^{v_a/\lambda} \right] f(\boldsymbol{v}) \, \mathrm{d}\boldsymbol{v} &= \lambda \mathbb{E}_{\boldsymbol{\epsilon}'} \left[\max_{a \in B_k} \left(\overline{v}_a + \lambda \ln(\mathcal{P}_a) + \epsilon'_a \right) \right] \\
&= \ln \left[\sum_{a \in B_k} e^{\overline{v}_a + \lambda \ln(\mathcal{P}_a)} \right]
\end{aligned}$$

A.1.2 Solving for optimal unconditional probabilities

The maximization problem can thus be rewritten as follows:

$$\max_{\mathcal{P}_1,\ldots,\mathcal{P}_k} \ln\left[\sum_{a\in B_k} e^{\overline{v}_a+\lambda\ln(\mathcal{P}_a)}\right]$$

under the constraints that $\sum_{a\in B_k}\mathcal{P}_a = 1$, and $\mathcal{P}_a \geq 0, \forall a \in B_k$. Exponentiating the objective function, the Lagrangian of is given by:

$$\mathcal{L}(\mathcal{P}) = \sum_{a\in B_k}\mathcal{P}_a^{\lambda}e^{\overline{v}_a} - \psi\left(\sum_{a\in B_k}\mathcal{P}_a - 1\right) + \sum_{a\in B_k}\phi_a\mathcal{P}_a$$

The complementary slackness condition is $\phi_a P_a^0 = 0$ with $\phi_a \geq 0$. The first order condition with respect to $\mathcal{P}_h$ is:

$$\lambda(\mathcal{P}_h^{B_k})^{\lambda-1}e^{\overline{v}_h} - \psi + \phi_h = 0$$

As we have restricted the alternatives so that $P_h > 0, \forall h \in B_k$, the first order condition can be simplified to:

$$\mathcal{P}_h^{B_k} = \left(\frac{\psi}{\lambda}e^{-\overline{v}_h}\right)^{\frac{1}{\lambda-1}}$$

Summing over alternatives:

$$\sum_{a\in B_k}\mathcal{P}_a^{B_k} = \sum_{a\in B_k}\left(\frac{\psi}{\lambda}e^{-\overline{v}_a}\right)^{\frac{1}{\lambda-1}} = 1$$

This can be rewritten as:

$$\psi^{\frac{1}{\lambda-1}}\sum_{a\in B_k}\left(\frac{e^{-\overline{v}_a}}{\lambda}\right)^{\frac{1}{\lambda-1}} = 1$$

Thus the Lagrangian multiplier ψ is equal to:

$$\psi = \left[\frac{1}{\sum_{a\in B_k}\left(\frac{e^{-\overline{v}_a}}{\lambda}\right)^{\frac{1}{\lambda-1}}}\right]^{\lambda-1}$$

Replacing this value of the Lagrangian multiplier in the expression for $\mathcal{P}_h^{B_k}$:

$$\begin{aligned}\mathcal{P}_h^{B_k} &= \frac{\left(\frac{e^{-\overline{v}_h}}{\lambda}\right)^{\frac{1}{\lambda-1}}}{\sum_{a\in B_k}\left(\frac{e^{-\overline{v}_a}}{\lambda}\right)^{\frac{1}{\lambda-1}}} \\ &= \frac{e^{\overline{v}_h/(1-\lambda)}}{\sum_{a\in B_k} e^{\overline{v}_a/(1-\lambda)}}\end{aligned} \tag{.2}$$

A.1.3 Optimal consideration set

If we plug in the expression for the optimal unconditional choice probabilities in (.2) into the objective function in (.1), we obtain the expected value from optimally choosing from an arbitrary set B_k:

$$\begin{aligned}\mathbb{E}_{B_k} &= \ln\left[\sum_{a=1}^{k} e^{\overline{v}_a+\lambda\ln\left(\mathcal{P}_a^{B_k}\right)}\right] \\ &= \ln\left[\sum_{a=1}^{k} e^{\overline{v}_a}\left(\mathcal{P}_a^{B_k}\right)^{\lambda}\right] \\ &= \ln\left[\frac{\sum_{a=1}^{k} e^{\overline{v}_a/(1-\lambda)}}{\left(\sum_{l=1}^{k} e^{\overline{v}_l/(1-\lambda)}\right)^{\lambda}}\right]\end{aligned} \tag{.3}$$

We have that $\mathbb{E}_{B_{k+1}} > \mathbb{E}_{B_k}$, for $k \leq N-1$, if and only if:

$$\frac{\sum_{a=1}^{k+1} e^{\overline{v}_a/(1-\lambda)}}{\left(\sum_{l=1}^{k+1} e^{\overline{v}_l/(1-\lambda)}\right)^{\lambda}} > \frac{\sum_{a=1}^{k} e^{\overline{v}_a/(1-\lambda)}}{\left(\sum_{l=1}^{k} e^{\overline{v}_l/(1-\lambda)}\right)^{\lambda}}$$

Moving terms around:

$$\frac{\sum_{a=1}^{k+1} e^{\overline{v}_a/(1-\lambda)}}{\sum_{a=1}^{k} e^{\overline{v}_a/(1-\lambda)}} > \left(\frac{\sum_{l=1}^{k+1} e^{\overline{v}_l/(1-\lambda)}}{\sum_{l=1}^{k} e^{\overline{v}_l/(1-\lambda)}}\right)^{\lambda}$$

which is always satisfied as the ratio that appears on both sides is always greater than 1, and $\lambda \in (0,1)$. Hence, $\mathbb{E}_{B_{k+1}} > \mathbb{E}_{B_k}$, $\forall k \leq N-1$, and the consideration set is thus $B_N = A$.

A.1.4 Analytical results of $C_1(\mathcal{P}_1)$ and $C_1(\mathcal{P}_1)/C(\mathcal{P}_1)$

If the take the partial derivative of $C_1(\mathcal{P}_1)$ with respect to $\mathcal{P}_1$, we obtain:

$$\begin{aligned}\frac{\partial C_1(\mathcal{P}_1)}{\partial \mathcal{P}_1} &= -\lambda\left[\ln \mathcal{P}_1 + 1\right] + \lambda \int_{-\infty}^{+\infty} \frac{\partial \mathcal{P}_1(x)}{\partial \mathcal{P}_1}\left(\ln \mathcal{P}_1(x) + 1\right) g(x)\,\mathrm{d}x \\ &= -\lambda\left[\ln \mathcal{P}_1 + 1\right] + \\ &\quad + \frac{1}{\mathcal{P}_1(1-\mathcal{P}_1)}\left[\int_{-\infty}^{+\infty}\left[\ln \mathcal{P}_1(x) + 1\right]\mathcal{P}_1(x)[1-\mathcal{P}_1(x)]g(x)\,\mathrm{d}x\right]\end{aligned} \tag{.4}$$

as:

$$\frac{\partial \mathcal{P}_1(x)}{\partial \mathcal{P}_1} = \frac{1}{\lambda}\frac{\mathcal{P}_1(x)}{\mathcal{P}_1}\frac{1-\mathcal{P}_1(x)}{1-\mathcal{P}_1}$$

When alternatives are *ex ante* identical, i.e., $\mathcal{P}_1 = 1/2$, we have that:

$$\mathcal{P}_1(x) = \frac{1}{1+e^{-x/\lambda}}$$

We can thus rewrite (.4) as follows:

$$\frac{\partial C_1(\mathcal{P}_1)}{\partial \mathcal{P}_1}\Big|_{\mathcal{P}_1=1/2} = \lambda\left[\ln(2) - 1\right] - 4\int_{-\infty}^{+\infty} k(x)h(x)\,\mathrm{d}x \tag{.5}$$

where we have defined:

$$h(x) \equiv \frac{1}{1+e^{-x/\lambda}}\left(1 - \frac{1}{1+e^{-x/\lambda}}\right) g(x)$$

and:

$$k(x) \equiv \ln(1+e^{-x/\lambda}) - 1$$

As the function $h(x)$ is symmetric around zero, i.e., $h(x) = h(-x)$, while the function $k(z)$ is such that $k(x) + k(-x) \geq 0$, we can conclude that the integral appearing in (.5) is positive, and thus:

$$\frac{\partial C_1(\mathcal{P}_1)}{\partial \mathcal{P}_1}\Big|_{\mathcal{P}_1=1/2} < \lambda\left[\ln(2) - 1\right] < 0 \tag{.6}$$

This also entails that, when $\mathcal{P}_1 = 1/2$, $\partial C_2(\mathcal{P}_1)/\partial \mathcal{P}_1 > 0$, $\partial C(\mathcal{P}_1)/\partial \mathcal{P}_1 = 0$, and the share of the total cost devoted to alternative 1 is decreasing.[51]

[51] Following the same steps, we can also extend the results about the slope of $C_1(\mathcal{P}_1)$ to any $\mathcal{P}_1 \geq 1/2$.

A.2 Full-information RUM model

A.2.1 Unobserved heterogeneity and $\mathcal{P}_1$

Consider a full-information RUM model with unobserved heterogeneity describing the location-decision problem that migrants from a given origin face.[52] Let $v_k = \alpha(\ln y_k - \ln \tau_k)$ represent the deterministic component of utility associated with migrating to k, with k denoting one of the N alternatives belonging to the choice set A. Let us introduce the canonical assumption that the individual-specific stochastic component of utility ϵ_{ik} is i.i.d. EVT-1, with a scale parameter $\sigma > 0$. The variance of this distribution is equal to $(\pi^2/6)\sigma^2$, so that a greater value of σ reflects a greater unobserved heterogeneity in location-specific utility. The probability that a migrant finds optimal to opt for destination $k \in A$ is given by (McFadden, 1978):

$$\mathcal{P}_k = \frac{e^{v_k/\sigma}}{\sum_{a\in A} e^{v_a/\sigma}} \tag{.1}$$

A key property of this discrete choice model is the independence from irrelevant alternatives, i.e., $\ln(P_k/P_l) = (v_k - v_l)/\sigma$, $\forall k, l \in A$. An implication of this fundamental property is that the marginal effect of a variation in the deterministic component of utility on the log odds ratio $\ln(P_k/P_l)$ is independent from $\mathcal{P}_k$ and $\mathcal{P}_l$, i.e., from v_k and v_l. The partial derivative of $\ln \mathcal{P}_k$ in (.1) with respect to v_k is given by:

$$\frac{\partial \ln \mathcal{P}_k}{\partial v_k} = \frac{1}{\sigma}(1 - \mathcal{P}_k) \geq 0 \tag{.2}$$

Without loss of generality, let us assume that $v_1 \geq v_2 \geq ... \geq v_N$, so that $P_1 \geq P_k$, $\forall k \in A \setminus \{1\}$. If we compute the partial derivative of $\ln P_1$ with respect to σ, we obtain:

$$\frac{\partial \ln \mathcal{P}_1}{\partial \sigma} = -\frac{1}{\sigma^2}\left(v_1 - \sum_{a\in A} v_a \mathcal{P}_a\right) \leq 0 \tag{.3}$$

with the inequality in (.3) holding strictly whenever $\mathcal{P}_1 > 1/N$. Thus, when σ is lower, then the probability $\mathcal{P}_1$ of opting for the alternative that is, on average, most attractive increases, and the responsiveness of the choice probabilities $\mathcal{P}_k$ with respect to variations in the deterministic component of utility v_k gets magnified. This, in turn, entails that even in a full-information RUM model the share of migration flows in the main destination is correlated with the size of the estimated coefficients, but in a way that is opposite to the one that characterizes a model with costly information acquisition.

[52] We avoid, as in Section 2.2 introducing origin and time subscripts to avoid cluttering the notation.

Chapter III*

*This chapter is under revise and resubmit at the *Journal of Economic Geography*.

Structural Gravity Model of Migration: Direct and Diversion Effects of an Asylum Policy*

Lucas Guichard[a,b]

[a] *Université Clermont Auvergne, CNRS, IRD, CERDI*
[b] *IAB*†

Abstract

Does the adoption of a list of safe countries of origin influence the asylum applications lodged in OECD member states? We draw on a structural gravity model to derive an empirical migration equation that is brought to the data to estimate the direct effect of the list on the bilateral number of asylum claims. This, in turn, allows us to solve the structural model to quantify the externalities arising from a counter-factual experiment about the safe country policy. The empirical analysis reveals that the introduction of a list of safe source countries leads to a decrease of around 30% in the bilateral volume of asylum applications. The simulation exercise under an hypothetical change of the asylum policy suggests the presence of diversion effects on the sheer scale of asylum claims across *both* origin and destination countries.

Keywords: refugees; asylum seekers; safe country.

JEL codes: F22.

*I would like to thank Simone Bertoli, Sekou Keita, Steffen Sirries, Ignat Stepanok, the participants to the COMPIE Conference (Berlin, September 2018), to the ZEW Workshop on the Economics of Immigration (Mannheim, October 2018) and to the 3rd Conference on Understanding Voluntary and Forced Migration (Lille, April 2019) for their comments and suggestions. Lucas Guichard acknowledges the support received from the *Agence Nationale de la Recherche* of the French government through the program "*Investissements d'avenir*" (ANR-10-LABX-14-01). The usual disclaimers apply.
†Regensburger Strasse, 104, D-90478 Nuremberg; E-mail: Lucas.Guichard2@iab.de.

3.1 Introduction

The evolution of migration policy modifies the attractiveness of the destination where the change has been implemented. Depending on its nature, the latter will alter migration costs that, in turn, can reduce, increase or leave bilateral migration flows unaffected, i.e., it can trigger a *direct* effect on the number of migrants observed between two countries. At the same time, the policy innovation will also generate externalities on other migration flows, i.e., it will lead to *diversion* effects across countries. Bertoli and Fernández-Huertas Moraga (2015) have shown that the requirement of a visa in one country affects the migration flows going to alternative locations, thus establishing evidence and a possible magnitude for redirection effects across destinations. More recently, Beverelli and Orefice (2019) have built on the fact that the attractiveness of a destination is endogenous with respect to the scale of migration flows to highlight the occurrence of deflection effects on migration flows across (similiar) source countries when lower bilateral costs reduce employment from third origins in the same host country. The literature on structural gravity model of trade, which has been extended to migration by Anderson (2011), provides us with a framework to evaluate the direct and the (two types of) indirect effects on migration flows.

We draw on the model to examine the consequences of the adoption of a list of safe source countries[1] on the number of asylum applications lodged in OECD countries. The recent large arrival of asylum seekers in Europe has fed the debate on the failures of the Dublin regulations and the related turmoil about identifying the countries that should process asylum applications. A corollary feature has been to focus on the soundness of asylum claims originating from countries where individuals faced different security and socio-economic conditions. Under the asylum context, the decision to migrate is often expected to be forced and less driven by economic motives, which implies that the debate is regularly shifted towards granting the refugee status only to the well-founded cases while dismissing other asylum applications. By definition, the safe country concept is one way to address the above issue as it makes an explicit distinction across origins where asylum claims from countries considered as being safe are almost exclusively specified as manifestly unfounded by the destinations relying on the safe country policy. The latter has recently received attention in the public debate due to the proposal (September 2015) by the European Commission of a regulation towards an EU

[1]The concept of safe source country is a presumption that certain countries can be defined as being safe for their nationals to the extent that "it can be shown that there is generally and consistently no persecution as defined in Article 9 of Directive 2004/83/EC, no torture or inhuman or degrading treatment or punishment and no threat by reason of indiscriminate violence in situations of international or internal armed conflict." Source: Annex II of the Asylum Procedure Directive on the "Designation of safe countries of origin for the purposes of Articles 29 and 30(1)."

common list of safe countries of origin[2].

The contribution of this paper is twofold. First, we propose a methodology to estimate the direct and indirect externalities associated to migration policy building upon the structural gravity model of Anderson (2011). Second, we collect fine-grained information about the evolution of the lists of safe source countries in OECD countries to analyse their effects on the number of asylum applications. We estimate first a migration gravity equation to obtain the direct effect of the asylum policy. The retrieved coefficient is then used to solve the structural gravity model and quantify the diversion effects arising under a counter-factual experiment on the list of safe source countries. The implication of the above procedure with available data is to document the deterrent effect of the asylum policy on the bilateral number of asylum claims and the presence of redirection of asylum seekers occurring *both* across destination and origin countries.

The analysis is carried out in two steps. We first derive a migration gravity equation from the model introduced by Anderson (2011) that is brought to the data to determine the direct effect of the list of safe source countries on the bilateral sheer scale of asylum applications. We rely on monthly asylum data from the UNHCR over the period 2000-2017 and unique information about the evolution of origin countries registered as being safe in 19 OECD countries. The high-frequency pattern of the data is exploited to estimate the gravity specification with a rich structure of fixed effects. The results reveal that the destination-specific inclusion of one origin in the list of safe countries leads to a decrease of around 30% in the volume of asylum applications between the two countries. The second stage uses the empirical coefficient to solve the gravity model and identify the redirection effects arising from a counter-factual experiment that implements the introduction of a list of safe source countries in four destinations (Finland, Italy, Norway, and Sweden) that do not have such a list. The comparison between the baseline and the hypothetical scenario suggests a diversion on the number of asylum claims from safe origins to third host countries (i.e., redirection across destinations) and from unlisted source countries to Finland, Italy, Norway, and Sweden (i.e., redirection across origins).

Our paper is related to two different strands of literature: (*i*) estimation of the effects related to migration policy with gravity equations, and (*ii*) determinants of asylum migration.

With respect to the estimation of the effects associated to migration policy with gravity equations, our work is related to papers that have dealt with the methodological challenges raised

[2]http://www.europarl.europa.eu/legislative-train/theme-towards-a-new-policy-on-migration/file-european-list-of-safe-countries-of-origin

by the influence of multilateral resistance to migration on the empirical derivations of the gravity model (Bertoli and Fernández-Huertas Moraga, 2013; Beine *et al.*, 2016). More specifically, Bertoli *et al.* (2011) and Bertoli and Fernández-Huertas Moraga (2015) have studied how visa policy affects international migration flows, whereas Ortega and Peri (2013) have considered more general variables, e.g. an index capturing the direction of the change in entry tightness. The presence of policy externalities has been acknowledged in the migration literature, but evidence on their amplitude remains scarce. Two notable exceptions are Bertoli and Fernández-Huertas Moraga (2015) who derived bounds associated to the indirect effects of visa policy on migration flows across destinations, and Beverelli and Orefice (2019) who document the occurrence of redirection across economically similar origins with some degree of cultural affinity (e.g., sharing a common language).

As far as the determinants of asylum migration are concerned, our study is close to papers that have included asylum policy as drivers explaining asylum applications beyond other push and pull factors. The main idea has revolved around the elaboration of indexes that could track the stance in terms of migration/asylum policy both over time and across countries. Thielemann (2004, 2006) has built a deterrence index related to asylum policies, whereas Neumayer (2004) has used proxy variables (e.g., the percentage of cabinet portfolios held by left-wing parties) to account for expected trends in the laws enacted by host countries towards asylum seekers. More closely related to our paper, Hatton (2016) has estimated a gravity equation with different sets of fixed effects to assess the effect of three categories of policies depending on their priorities (access, processing, and welfare) with respect to asylum applications. The variables are based on data collection described first in Hatton (2004), and then extended in Hatton (2009).

The rest of this paper is structured as follows: Section 3.2 presents the migration gravity framework of Anderson (2011), and the steps required to (*i*) derive an empirical gravity equation and (*ii*) run a counter-factual experiment after the model has been solved. Section 3.3 applies the procedure to identify the direct and indirect effects of the list of safe source countries on the number of asylum applications lodged in OECD countries. Finally, Section 3.4 concludes.

3.2 Structural gravity model of migration

3.2.1 Description of the model

Consider an individual i, born in country j, who decides to migrate to locations included in a choice set D that encompasses N alternatives corresponding to all possible destination countries k. Individual i is a utility-maximizer, and she will thus choose the alternative in the choice set associated with the highest level of utility. As usual in random utility maximization models, the utility is defined by two components, i.e. a deterministic part common to all potential migrants and an individual-specific feature that captures unobserved heterogeneity in preferences among decision makers.

The deterministic component of utility is determined by the characteristics v_k of destination $k \in D$ and by the costs of migrating from j to k, i.e. δ_{jk}. Suppose that the location-decision problem is solved after individual i has observed all the realizations of the stochastic part of utility φ_{ijk}, which is identically and independently distributed according to an EVT-1 distribution, then McFadden (1974) gives us the probability p_{jk} that destination k is the utility-maximizing alternative for a j-born individual i:

$$p_{jk} = \frac{e^{v_k - \delta_{jk}}}{\sum_{l \in D} e^{v_l - \delta_{jl}}} \tag{3.1}$$

Anderson (2011) assumes that the labour demand at destination k is totally rigid. This, in turn, implies that the deterministic component of utility v_k, which corresponds to wages in his setting, is endogenous to the scale of migration flows[3]. The implication is that the number of individuals moving from j to k depends on the entire matrix of bilateral migration costs, thus resulting in an increased interdependency in migration flows across countries.

We define $W^j \equiv \sum_l w_l / \delta_{jl}$ and the labour force $L^k \equiv \sum_j m_{jk}$ supplied to destination k from all origins, where $m_{jk} = p_{jk} N^j$. The world labour supply $N \equiv \sum_j N^j = \sum_k L^k$ follows. This

[3] It is often emphasized that asylum seekers are forced to leave their home countries, thus being less driven by economic motives (e.g., wages at destination) in their migration decision. We argue that the conditions at origin can indeed push people out of a given country, and they can then decide where would be the best location to go to based on the socio-economic situation in potential alternatives forming their choice set. The model is therefore consistent with explaining how asylum seekers choose the utility-maximizing destination, while being relatively agnostic about the causes underlying their departure. We relax this assumption in the empirical analysis, where we control for origin-specific push factors of asylum applications through origin-time fixed effects.

notation allows us to write the labour market clearance equation used in Anderson (2011):

$$L^k = w_k \sum_j \frac{1/\delta_{jk}}{W^j} N^j \tag{3.2}$$

Combining (3.1) with the market-clearing condition in (3.2) for each alternative in the choice set, one can express the expected scale of migration flows from j to k as:

$$E(m_{jk}) = \underbrace{\frac{L^k}{N} N^j}_{\text{Frictionless migration}} \times \underbrace{\frac{1/\delta_{jk}}{\Omega^k W^j}}_{\text{Migration frictions}} \tag{3.3}$$

with:

$$\underbrace{\Omega^k = \sum_{l \in D} \frac{1/\delta_{lk}}{W^l} \frac{N^l}{N}}_{\text{Inward multilateral resistance}} \quad \text{and} \quad \underbrace{W^j = \sum_{l \in D} \frac{1/\delta_{jl}}{\Omega^l} \frac{L^l}{N}}_{\text{Outward multilateral resistance}}$$

The expected migration flows are driven by the size of the origin population N^j and the labour force L^k supplied to destination k, relative to the world population N. They both increase bilateral migration flows, and the first component of equation (3.3) corresponds to the hypothetical number of migrants that would be observed in a frictionless world. However, migration flows are hindered by the moving barriers δ_{jk} and the multilateral resistance terms Ω^k and W^j. The inward multilateral resistance Ω^k is the weighted average of the probabilities that k is the utility-maximizing destination for j-born individuals, with weights given by the share of the l-born population in the world population N. The outward multilateral resistance W^j is the weighted average of the accessibility of destination l for j-born migrants, with weights given by the share of country j in the total world demand for labour.

3.2.2 From the theory to the empirical analysis

The migration gravity system proposed in (3.3) can be translated into an empirical gravity specification. Two steps are required. First, we add an error term ε_{jk} to capture the stochastic component of utility. Second, the estimated equation will rely on migration flows, whereas the framework of Anderson (2011) refers to migration stocks m_{jk}. The latter reflects the fact that the structural model is a stationary equilibrium in which the labour forces L^k are the result

of migration m_{jk} completely adjusting the labour supplied at each location to its equilibrium value given the initial stocks of labour N^j and the set of migration frictions δ_{jk}.

The switch from the model to the empirical setting is done through the observation that actual data provides us with the above variables at points in time; more precisely, the information is linked over time as we work with panel data. Moving from stocks to flows is then the outcome if we assume that the sequence of observations can be regarded as reaching the static equilibrium at each data points, with migration flows being the corresponding figure needed to get to that equilibrium (Anderson, 2011). In that context, the model involves a time dimension and its components can now be written with a time superscript t. Rearranging (3.3), we can then specify the migration flows from origin j to destination k as:

$$m_{jkt} = \exp\left[\ln(L^{kt}) + \ln(N^{jt}) - \ln(N^t) + \ln(1/\delta_{jkt}) - \ln(\Omega^{kt}) - \ln(W^{jt})\right] + \varepsilon_{jkt} \quad (3.4)$$

The above equation can be simplified through the inclusion of fixed effects that can control for most of the variables. Multilateral resistance terms are enclosed in origin-time fixed effects μ_{jt} and destination-time fixed effects μ_{kt}, which also directly account for origin- and destination-specific attributes (i.e., N^t, L^{kt}, and N^{jt}) driving migration flows. Country-pair dummies μ_{jk} capture *only* the migration costs δ_{jk} that do not vary over time, i.e. standard gravity controls such as physical distance or common language. The large set of fixed effects also implies that several of the classic push and pull factors (e.g., political terror scale or freedom indexes) associated to the sheer scale of asylum applications between countries are taken into account. We can thus rewrite (3.4) as follows:

$$m_{jkt} = \exp\left[\beta \boldsymbol{X}_{jkt} + \mu_{jt} + \mu_{kt} + \mu_{jk}\right] + \varepsilon_{jkt} \quad (3.5)$$

The specification presented in (3.5) is the empirical gravity equation that is brought to the data to identify the coefficient β related to the vector $\boldsymbol{X}$ of characteristics that are part of the migration costs and vary both across country-pairs and over time (i.e., δ_{jkt}). We use this setting in Section 3.3.2.1, where we estimate the direct effect of the list of safe source countries on the number of asylum applications lodged in OECD destination countries.

3.2.3 Estimating externalities through model simulation

We now describe how the migration gravity system can be used to determine the redirection effects of a policy change through counter-factual experiment on the migration costs. The

procedure relies on solving the model proposed in (3.3). To do so, we have to compute explicit values of the multilateral resistance terms Ω^k and W^j, which can be obtained from (available) data on L^k, N^j and N and from the estimation of the migration costs δ_{jk}. The latter are empirically recovered following the two-step strategy implemented in Anderson and Yotov (2016). The first stage consists of estimating (3.5) to derive values of the bilateral fixed effects for country-pairs with non-missing (or non-zero) migration flows:

$$m_{jkt} = \exp\left[\widehat{\beta}\boldsymbol{X}_{jkt} + \widehat{\mu}_{jt} + \widehat{\mu}_{kt} + \widehat{\mu}_{jk}\right] + \varepsilon_{jkt} \tag{3.6}$$

The second stage involves using the estimates of the dyadic fixed effects $\widehat{\mu}_{jk}$ from (3.6) as the dependent variable in a specification where the covariates include the set of standard gravity control variables along with origin and destination fixed effects:

$$\exp(\widehat{\mu}_{jk}) = \exp\left[\widehat{\theta}_1 \ln(\text{distance}_{jk}) + \widehat{\theta}_2\, \text{colony}_{jk} + \widehat{\theta}_3\, \text{language}_{jk}\right] + \varepsilon_{jk} \tag{3.7}$$

The predicted values from the estimation of (3.7) are then exploited to fill in missing figures of the migration costs. This, in turn, entails that the model can be solved to get the baseline migration flows. The next step is carried out by defining an hypothetical scenario that modifies the current state of world, i.e. that alters one component entering the migration costs, and solving again the migration gravity framework. We then compare the baseline results with the simulated ones to assess the diversion effects of the counter-factual policy change on migration flows between countries.

3.3 Asylum applications and the list of safe source countries

In this section, our objective is to test the empirical relevance of the adoption of a list of safe countries of origin on the number of asylum applications. To this end, we first introduce our data, before turning to the analysis of the direct and diversion effects arising from the asylum policy.

3.3.1 Data

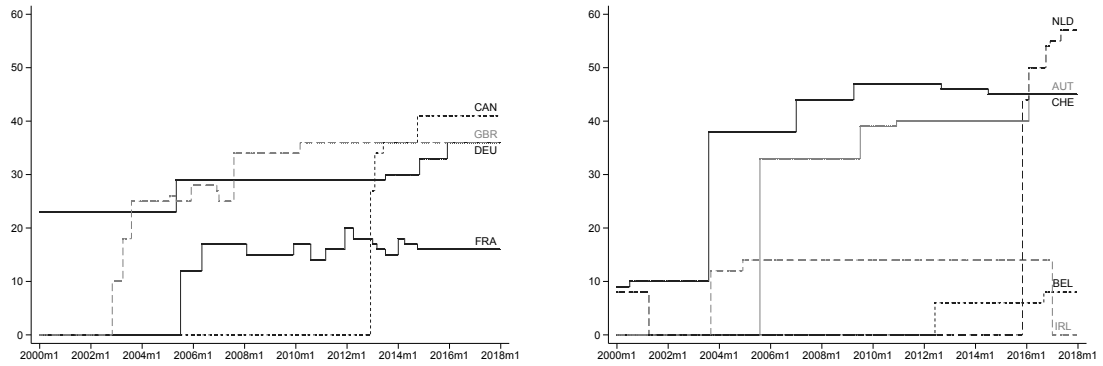
We describe here the source of our panel data on the list of safe source countries and on the bilateral number of asylum applications. We also present basic descriptive statistics.

3.3.1.1 List of safe countries of origin

The list of safe source countries is a policy that restricts potential asylum seekers from claiming international protection in some destinations where individuals are considered coming from a safe country of origin. The concept can be implemented and changed at the discretion of the host countries and several of them have relied on this tool within their asylum policy to deter asylum application originatings from specific source countries in the last thirty years. In principle, the list can be settled provided that the registered country can be considered as safe for its own nationals, mainly according to the security and socio-economic conditions that prevail there. In fact, only a subset of destinations have decided to enact a list of safe countries of origin.

This paper builds upon original (self-collected) information covering nineteen OECD countries over the period 2000-2017 with policy changes occurring at the monthly level. The data has been combined into a binary variable that takes the value 1 if an origin j is in the list of destination k at month t, and 0 otherwise. In our sample, ten countries (i.e., Australia, Finland, Greece, Italy, New Zealand, Norway, Portugal, Spain, Sweden, United States of America) do not have a list of safe source countries, while the remaining nine (i.e., Austria, Belgium, Canada, France, Germany, Ireland, the Netherlands, Switzerland, the United Kingdom) have classified some origins as safe, but not necessarily the same countries[4].

Figure 3.1: Evolution of the number of safe source countries in the nine relevant destinations



Source: Author's elaboration based on self-collected data.

Detailed information about the source countries notified as being safe and the date of their inclusion in (or, in some instances, removal from) the list are presented in Appendix A.1. An

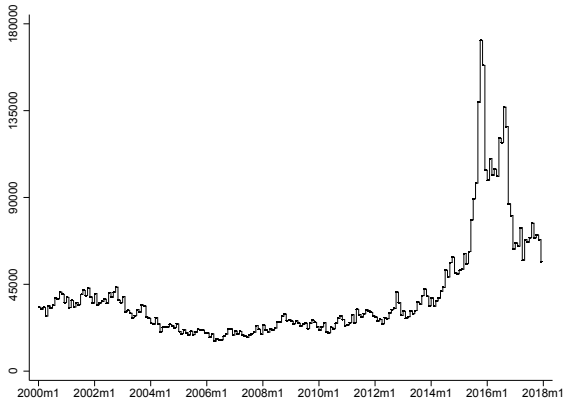
[4]Denmark is not in the sample as the list is likely to exist but is confidential, and we were only able to gather information on the safe countries in the early 2000s. The same seems to be true for Italy, but no clear evidence on the presence of the list entails that we have decided to keep Italy in the analysis.

interesting feature is represented by the heterogeneity across destinations with respect to (*i*) the identity of the safe countries of origin and (*ii*) the timing of their addition in the list. For instance, Canada mainly define high-income countries as being safe, while France has only registered low- and middle-income countries in the list with several of them being located in the Balkan region. Belgium has had a list of safe countries only since June 2012, while Germany has started to classify some countries as being safe since July 1993. Figure 3.1 displays the above patterns graphically, highlighting the evolution of the total number of safe countries that have been listed by the nine relevant destinations. The general trend exhibits an increase in the number of safe source countries over time, even though Ireland removed its list at the end of 2016 and others (i.e., France and the United Kingdom) have withdrawn some origins (e.g., Bangladesh and Sri Lanka) from their own respective lists.

3.3.1.2 Asylum applications

We rely on information about monthly asylum applications from the UNHCR[5]. The evolution of the number of asylum claims lodged in the 19 OECD countries is depicted on Figure 3.2. The total amount is relatively constant over time, until the onset of the civil conflict in Syria and the related surge in the arrival of asylum seekers to Europe. This, in turn, entails that Syria is the main origin country of asylum claims recorded in the sample, followed by Afghanistan and Iraq (Table 3.1). A large fraction of the applications has been received by Germany, with more than twice the number of claims counted in the second top destination (France).

Figure 3.2: Evolution of the total number of asylum applications



Source: Author's elaboration based on UNHCR monthly asylum applications.

Table 3.1: Main asylum origin and destination countries

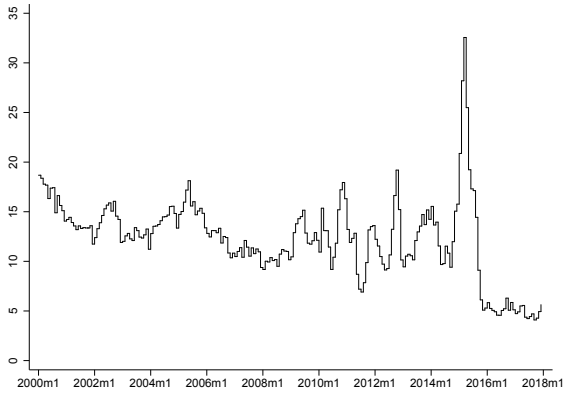
Origin		Destination	
Syria	888,203	Germany	2,114,635
Afghanistan	627,907	France	923,723
Iraq	598,238	UK	717,254
Serbia	436,929	Sweden	640,344
Pakistan	283,676	Italy	475,173
Nigeria	277,697	Canada	459,457
Russia	259,141	USA	442,583
Somalia	256,243	Austria	431,674
Iran	248,127	Belgium	333,567
Eritrea	245,669	Switzerland	319,619
Total	8,034,640	Total	8,034,640

Source: Author's elaboration based on UNHCR monthly asylum applications.

[5]Data is available from 1999 and can be downloaded at: http://popstats.unhcr.org/en/asylum_seekers_monthly. Notes about data availability can be found at: http://popstats.unhcr.org/download_notes.

Figure 3.3 and Table 3.2 mirror the above descriptives for asylum applications originating from countries listed as being safe. The only difference between the two plots is that Figure 3.3 reports the share (rather than the sheer scale) of asylum claims from safe countries relative to the total number of applications. The values range from 10 to 25% from 2000 to 2014, followed by a sharp and short-lived increase simultaneous to the peak in the level of asylum claims observed in OECD countries. Several Balkan countries (Serbia, Albania, Macedonia, or Bosnia-Herzegovina) are the main safe source countries, and their claims have mostly been lodged in Germany and France.

Figure 3.3: Evolution of the share of applications from safe countries relative to the total number of claims



Note: The reported share has been computed for a set of (safe) countries that is consistent over time, i.e. we have excluded those that were added and then removed from the list within our time frame.
Source: Author's elaboration based on UNHCR monthly asylum applications, and self-collected information on the destination-specific list of safe source countries.

Table 3.2: Main safe origin countries and their respective destination

Origin		Destination	
Serbia	128,103	Germany	130,438
Albania	59,248	France	83,852
Macedonia	27,354	UK	47,014
Bosnia-Herz.	18,385	Switzerland	25,215
India	15,135	Austria	10,333
Nigeria	14,877	Belgium	5,119
Ghana	14,648	Netherlands	4,937
Armenia	11,138		
Senegal	10,023		
Gambia	7,997		
Total	306,908	Total	306,908

Note: Canada and Ireland did not receive asylum applications from the listed ten origin countries.
Source: Author's elaboration based on UNHCR monthly asylum applications.

3.3.2 Safe countries of origin and asylum applications: direct effect

The above data are now exploited to estimate the direct effect of the list of safe source countries on the number of asylum applications. We also highlight some channels that could explain the negative pattern observed in the results.

3.3.2.1 Empirical strategy and results

We first analyse the consequences of the list of safe source countries on the number of asylum claims originating from country j and lodged in destination k at time t, i.e. the direct effect

of the asylum policy. The empirical strategy follows a staggered difference-in-difference framework, similar to the one used by Autor (2003). The specifications control for the influence of multilateral resistance to migration through the rich structure of fixed effects, which also take into account most of the push and pull factors driving the decision to migrate of asylum seekers. We rely on OLS and Poisson Pseudo-Maximum Likelihood (PPML) estimators to identify the coefficient β of the following equation:

$$m_{jkt} = \exp\left[\beta SCO_{jkt} + \mu_{jt} + \mu_{kt} + \begin{pmatrix}\mu_{jk}\\ \mu_{jky}\end{pmatrix}\right] + \varepsilon_{jkt} \tag{3.8}$$

with m_{jkt}, the (log) number of asylum applications from origin j to destination k at month t; SCO_{jkt} refers to the list of safe countries of origin; μ_{jt} corresponds to origin-month fixed effects; μ_{kt} represents destination-month fixed effects; μ_{jk} is a set of dyadic fixed effects, possibly varying across years (μ_{jky}), and ε_{jkt} is the error term.

The large set of fixed effects involved in the estimation of (3.8) is key to ensure that the identification of the coefficient β is purged from the influence of confounding factors. More specifically, allowing the dyadic fixed effects to vary over time allows us to control for unobserved dyad-specific time-varying (at the year level) determinants of bilateral asylum applications (Bertoli and Fernández-Huertas Moraga, 2013)[6].

However, the decision to register a specific origin in the list of safe countries might also be endogenous. A destination is likely to change its own list of safe source countries if (*i*) economic and/or safety conditions at origin have improved, or if (*ii*) it faces a high number of asylum applications coming from this origin[7]. The origin-time dummies μ_{jt} should attenuate the influence of (*i*) as they capture improvements in security or economic conditions at origin that could lead several destinations to consider that country as being safe. On the other hand, extending (3.8) with dyadic-year dummies should mitigate the effect of (*ii*), given that we would use variability at the bilateral level and within year to identify β. In case the above attempts are insufficient to (completely) deal with endogeneity concerns, the direction of the

[6]They are also crucial to rule out potential bias coming from factors that are imperfectly observable. In particular, receiving countries seem to become tougher with respect to undocumented migrants close to the date of the policy change. Prospective asylum seekers could use the information as a proxy for an increase of the risk of being sent back to the home country, thus deterring them from lodging an asylum claim in the respective destination. Results on the effectiveness of the fixed effects to account for the above pattern are available from the Author upon request.

[7]Note that point (*i*) might explain the observed diversity in terms of origins listed as safe across host countries. Point (*ii*) is consistent with the fact that some destinations (e.g., Ireland or the United Kingdom) have several English-speaking countries in their own lists, which are not included by other destinations.

related bias is unclear. We would be over-estimating the effect of interest if the inclusion in the list of safe countries of origin mostly reflected an improvement in conditions at origin, while we would be under-estimating the size of β if the policy change was mostly driven by a reaction to an increase in asylum applications from a specific source country.

Table 3.3: Direct effect of the list of safe source countries on asylum applications

	(1)	(2)	(3)	(4)
	OLS	PPML	OLS	PPML
Dependent variable	$\ln(m_{jkt})$	m_{jkt}	$\ln(m_{jkt})$	m_{jkt}
SCO_{jkt}	-0.262***	-0.304**	-0.243***	-0.343***
	(0.079)	(0.148)	(0.055)	(0.068)
μ_{jt}	✓	✓	✓	✓
μ_{kt}	✓	✓	✓	✓
μ_{jk}	✓	✓	✗	✗
μ_{jky}	✗	✗	✓	✓
Total sample	231,109	231,109	231,109	231,109
Zero dropped	26	0	26	0
Singletons dropped	6,180	6,175	12,309	12,317
Estimation sample	224,903	224,934	218,774	218,792
R^2	0.809		0.909	
Pseudo-R^2		0.902		0.953

Notes: OLS is estimated with the command `reghdfe` and PPML is based on the command `ppmlhdfe`. Clustered standard errors by country-pair in parentheses. ***, **, and * denote significance at the 1, 5, and 10 percent levels, respectively.
Source: Author's elaboration based on UNHCR monthly asylum applications and self-collected information about the destination-specific list of safe source countries.

Turning to the estimation, the results are presented in Table 3.3. They reveal a significant and negative effect of the list of safe countries of origin on the number of asylum applications lodged in the host country. In terms of magnitude, the coefficient (-0.304) reported in column (2) translates into an average percent partial impact of $e^{-0.304} - 1 \simeq -26\%$[8]. This entails that, *ceteris paribus*, the registration of a country as being safe reduces, on average, the number of asylum claims from that country to the corresponding destination by 26%. Adding country-pair-year fixed effects into the specification to control for dyadic-specific confounding factors varying across time implies a slightly higher effect, whereby the number

[8]The empirical estimates discussed here correspond to partial effects, meaning that they neglect adjustments that could arise from general equilibrium effects.

of asylum applications decreases on average by 29% given the estimate in column (4)[9].

The specification in (3.8) is useful to pinpoint how modifications in the list of safe countries of origin affect the bilateral number of asylum applications[10]. However, it does not help shedding light on the dynamics of the impact, i.e. we can not evaluate whether anticipatory responses arised due to the past number of asylum claims lodged in a given destination. If individuals seeked asylum before the list was actually modified and, in turn, triggered the policy change rather than vice versa, the estimates would obscure this reverse causality. This potential issue is explored in the coefficients displayed on Table 3.4, which provides results from the estimation of (3.8) augmented with leads and lags of the variable of interest. Specifically, we add an indicator variable for 1 and 2 months before the list has been changed, months 0-3 after the policy change, and month 4 forward. Of these seven indicator variables, the first six are equal to one in the relevant month, while the last one is equal to one in each month, starting with the fourth month after the evolution of the list. The related equation can be written as follows:

$$m_{jkt} = \exp\left[\sum_{i=-2}^{4} \beta_i SCO_{jkt+i} + \mu_{jt} + \mu_{kt} + \begin{pmatrix} \mu_{jk} \\ \mu_{jky} \end{pmatrix}\right] + \varepsilon_{jkt} \tag{3.9}$$

The first two data columns of Table 3.4 only control for time-invariant dyadic covariates (e.g., physical distance or common language) through the inclusion of country-pair fixed effects. The last data two columns provide the most conservative estimation, where we account for dyadic determinants of asylum applications that vary over time. The preferred specification in column (4) presents coefficients on the change leads close to zero (and insignificant), suggesting little evidence of an anticipatory response of the policy variable associated to the past number of asylum applications. In the month of change, the number of asylum claims from safe countries lodged at destination decreases by 24%, after which the effect peaks at 49% and averages at around 42% in month 4 forward.

[9] Several robustness checks on the direct negative effect are shown in Appendix A.2. They emphasize the robustness of the findings, especially with respect to potential issues related to sample selection. We have also carried out the estimation without high-income origin countries, and the results (available from the Author upon request) are identical to the coefficients presented in Table 3.3.

[10] So far, we have assumed that the effect of the policy change is symmetric, i.e. adding a source country to the list has the same (albeit opposite) effect as removing it from the list. We have relaxed the hypothesis by having two binary variables for the inclusion in and the removal of a country from the list, respectively. The estimates related to the former are identical to those reported in Table 3.3, whereas the coefficients of the latter are insignificant. Results are available from the Author upon request.

Table 3.4: Dynamic effect of the policy change on asylum applications

	(1)	(2)	(3)	(4)
	OLS	PPML	OLS	PPML
Dependent variable	$\ln(m_{jkt})$	m_{jkt}	$\ln(m_{jkt})$	m_{jkt}
2 months prior	0.386***	0.609***	0.122*	-0.011
	(0.114)	(0.124)	(0.072)	(0.088)
1 month prior	0.421***	0.630***	0.123	-0.013
	(0.126)	(0.123)	(0.084)	(0.108)
Month of change	0.314**	0.421***	-0.018	-0.273**
	(0.125)	(0.128)	(0.089)	(0.115)
1 month after	0.074	0.192	-0.219**	-0.169
	(0.124)	(0.143)	(0.095)	(0.123)
2 months after	0.124	0.022	-0.238**	-0.679***
	(0.129)	(0.142)	(0.095)	(0.155)
3 months after	0.038	0.162	-0.379***	-0.564***
	(0.127)	(0.156)	(0.087)	(0.144)
4 or more months after	-0.334***	-0.353**	-0.487***	-0.541***
	(0.121)	(0.161)	(0.084)	(0.101)
μ_{jt}	✓	✓	✓	✓
μ_{kt}	✓	✓	✓	✓
μ_{jk}	✓	✓	✗	✗
μ_{jky}	✗	✗	✓	✓
Total sample	133,385	133,385	133,385	133,385
Singletons dropped	5,742	5,742	6,538	6,538
Estimation sample	127,643	127,643	126,847	126,847
R^2	0.799		0.912	
Pseudo-R^2		0.901		0.955

Notes: OLS is estimated with the command `reghdfe` and PPML is based on the command `ppmlhdfe`. Clustered standard errors by country-pair in parentheses. ***, **, and * denote significance at the 1, 5, and 10 percent levels, respectively.

Source: Author's elaboration based on UNHCR monthly asylum applications and self-collected information about the destination-specific list of safe source countries.

3.3.2.2 Potential mechanisms

Asylum destinations are likely to decide on whether some countries can be defined as being safe based on the origin-specific safety and economic situation. If most nationals are assumed not to be endangered in their home country, seeking asylum abroad is often deemed as being manifestly unfounded by the corresponding host countries. They, in turn, can choose to register the origin in the list of safe source countries with the aim of (possibly implicitly)

deterring the arrival of further asylum seekers from the safe country. However, the list is, *per se*, immaterial to asylum seekers, and the mechanisms underlying the negative effect uncovered in Section 3.3.2.1 remain unclear.

On the one hand, the policy change might entail more restrictive decision with respect to the determination of the status of individuals coming from safe source countries. This would lower the origin-specific recognition rate of the respective applications[11]. We explore this channel with the empirical setting already used in Section 3.3.2.1. The main difference is that we now consider the recognition rate of asylum applications as the outcome variable[12]. We also look at the dynamics of the effect, i.e. whether it grows, stabilises or mean reverts over time. We do so by augmenting the baseline specification with leads and lags of the variable of interest. More specifically, we add binary indicators for 1 and 2 quarters before and quarters 0-4 after a policy change[13]. The results are reported in Table 3.5, where the different specifications have been estimated with PPML.

The coefficient obtained in column (1) does not provide evidence of a significant effect of the list of safe source countries on the overall recognition rate of asylum applications lodged at destination. However, the estimates depicted in column (2) might suggest an increase of the recognition rate some time (around 6 months) after the policy modification was implemented in a given host country. This finding may be consistent with variations in the composition of the pool of asylum seekers who decide to migrate after the policy hase been modified. If individuals who assess their expected probability to be granted protection as being high is the only group claiming asylum abroad, the related (self-)selection could reduce the total number of asylum applications, while keeping the probability of acceptance unchanged.

We investigate the above mechanism in columns (3) to (6) of Table 3.5, which replicate the analysis with disaggregated data on the type of protection that was granted in the host country. The estimates in columns (5) and (6) document that the potential positive effect uncovered with the total recognition rate could be explained by an increase (after the policy change has

[11] The recognition rate is defined as the share of first instance positive decisions relative to the total number of decisions made on asylum applications.

[12] Due to the (un)availability of data, the analysis covers only recent years (2008-2016) with quarterly information, and the sample is restricted to European countries, i.e. Australia, Canada, New Zealand and the United States are not included.

[13] Of the seven variables, the first six are equal to one only in the relevant quarter, while the last dummy takes the value one in each quarter starting with the fourth quarter after the policy change.

been adopted) in the probability of being granted subsidiary protection[14]. In line with the previous results, we also find in columns (3) and (4) that a change in the listed safe countries seems not to alter the probability of being granted refugee/humanitarian protection[15].

On the other hand, the safe country concept could be used to influence the processing of asylum applications, so that the inclusion of one source country in the list would trigger a decrease in the expected time required to process the claims coming from some origins[16]. We propose an empirical test of this assumption with a specification where the expected processing time is set as dependent variable. The related results are presented in the last two columns of Table 3.5. The estimates provide support for a negative significant effect of the list of safe source countries on the expected duration of stay in the host country for individuals coming from safe origins, which could deter future asylum arrivals from these countries.

[14]Subsidiary protection is defined as "a status as defined in Art.2(f) of Directive 2004/83/EC. According to Art.2(e) of Directive 2004/83/EC a person eligible for subsidiary protection means a third country national or a stateless person who does not qualify as a refugee but in respect of whom substantial grounds have been shown for believing that the person concerned, if returned to his or her country of origin, or in the case of a stateless person, to his or her country of former habitual residence, would face a real risk of suffering serious harm and is unable, or, owing to such risk, unwilling to avail himself or herself of the protection of that country."

[15]Refugee protection is defined by Eurostat as "a status as defined in Art.2(d) of Directive 2004/83/EC within the meaning of Art.1 of the Geneva Convention relating to the Status of Refugees of 28 July 1951, as amended by the New York Protocol of 31 January 1967." Humanitarian protection encompasses "persons who are not eligible for international protection as currently defined in the first stage legal instruments, but are nonetheless protected against removal under the obligations that are imposed on all Member States by international refugee or human rights instruments or on the basis of principles flowing from such instruments."

[16]The expected processing time is defined by comparing the number of origin-specific pending applications at the end of month $t-1$ with the number of months (from 1 to 36) over which it is necessary to cumulate applications (from $t-1$) in order to reach the number of pending applications.

Table 3.5: Effect of the list of safe source countries on the recognition rate and expected processing time

Dependent variable	Recognition rate		Recognition rate Ref./Hum.		Recognition rate Subsidiary		Expected processing time	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
SCO_{jkt}	0.283		-0.015		1.394**		-0.331***	
	(0.228)		(0.176)		(0.583)		(0.114)	
1 quarter prior		-0.505*		-0.630**		0.316		-0.140
		(0.264)		(0.320)		(0.554)		(0.143)
Quarter of change		0.077		-0.445		-0.761		-0.195
		(0.475)		(0.381)		(1.594)		(0.153)
1 quarter after		-0.497		-0.687		0.367		-0.365**
		(0.340)		(0.424)		(0.742)		(0.173)
2 quarters after		0.588**		-0.020		1.645**		-0.655***
		(0.274)		(0.264)		(0.721)		(0.174)
3+ quarters after		0.371		-0.084		2.195***		-0.304**
		(0.277)		(0.183)		(0.635)		(0.123)
μ_{jt}	✓	✓	✓	✓	✓	✓	✓	✓
μ_{kt}	✓	✓	✓	✓	✓	✓	✓	✓
μ_{jk}	✓	✓	✓	✓	✓	✓	✓	✓
Total sample	104,400	93,960	104,400	93,960	104,400	93,960	104,400	93,960
Singletons dropped	69,152	62,554	72,548	65,854	88,224	79,553	13,644	9,963
Estimation sample	35,248	31,406	31,852	28,106	16,176	14,407	90,678	83,940
Pseudo-R^2	0.518	0.522	0.500	0.504	0.606	0.609	0.497	0.489

Notes: All regressions have been estimated with PPML using the command `ppmlhdfe`. Clustered standard errors by country-pair in parentheses. ***, **, and * denote significance at the 1, 5, and 10 percent levels, respectively.
Source: Author's elaboration based on Eurostat data and self-collected information about the destination-specific list of safe source countries.

3.3.3 Safe countries of origin and asylum applications: redirection effects

The recent surge in the arrival of asylum seekers in European countries has fuelled the debate on the failures of the Dublin regulations and the related turmoil about determining the countries that should process asylum applications. The evolution of the list of safe source countries in a given destination can entail externalities in other receiving countries, thus affecting the distribution of asylum applications in Europe. More specifically, a policy change in country k is likely to trigger, beyond the direct bilateral effect uncovered in Section 3.3, a redirection of asylum seekers towards other destinations l. In technical terms, the latter implies that a fraction of the claims that would have been lodged in k if no modification of the

list had occurred there is now diverted to countries l[17].

Another externality can arise across origin countries, i.e. a destination k is now attractive for some individuals seeking asylum from origin countries q who decide to migrate there after the policy change. Anderson (2011) proposes a conditional general equilibrium model in which the labour market clears in each country, but the labour demand at destination is exogenous. An alteration of the list of safe source countries in k leads theoretically to a reduction in the bilateral migration (stock) from j to k, thus increasing wages in the host country and attracting more migrants coming from origins q.

We rely on model simulation based on the methodology described in Section 3.2.3 to evaluate the potential redirection effects generated by a (counter-factual) modification in the list of safe source countries. The hypothetical change consists of assigning a list of safe origins to four destinations (i.e., Finland, Italy, Norway and Sweden) that actually did not use the safe country concept in their own respective asylum policy. We define the same source countries, i.e., Albania, Bosnia-Herzegovina, Bulgaria, Macedonia, Montenegro, and Serbia, as being safe given that they have commonly been listed by several host countries (Table A.1). We build upon the estimated coefficient for the direct effect of the policy to compute the migration costs using equations (3.6) and (3.7):

$$m_{jkt} = \exp\left[-0.343 \cdot SCO_{jkt} + \widehat{\mu}_{jt} + \widehat{\mu}_{kt} + \widehat{\mu}_{jk}\right] + \varepsilon_{jkt} \tag{3.10}$$

$$\exp(\widehat{\mu}_{jk}) = \exp\left[\widehat{\theta}_1 \ln(distance_{jk}) + \widehat{\theta}_2\, colony_{jk} + \widehat{\theta}_3\, language_{jk}\right] + \varepsilon_{jk} \tag{3.11}$$

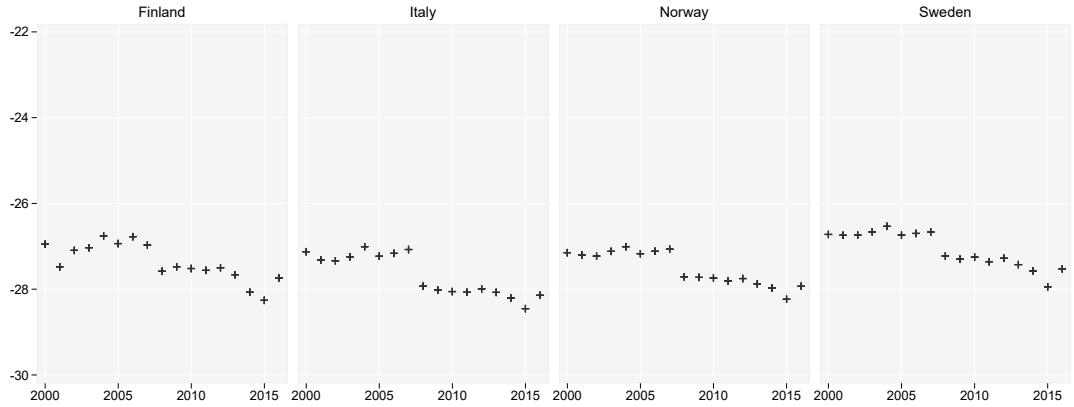
The above first stage gives us the dyadic baseline migration costs and their counterpart under the counter-factual scenario. The implication of the policy change is an increase of the migration barriers faced by asylum seekers from the six safe source countries attributed to Finland, Italy, Norway and Sweden. We use the estimated costs in the second stage, where we derive the baseline and hypothetical migration flows in line with the gravity system introduced by Anderson (2011). We can then simply compare the difference between the two values to quantify the potential magnitude of the redirection effects implied by a modification in the list of safe countries of origin.

[17] The current analysis represents one way to evaluate the redirection effects arising from a policy change at destination. Looking at the response of countries l affected by the change in k is more challenging and would require the availability of specific data (e.g., on the evolution, by citizenship of asylum seekers, of Dublin requests submitted by country k to country l after the policy change). Therefore, we are forced to leave this question open for further research.

Some comments are worth mentioning before turning out to the results. First, the structural analysis is carried out on a year-by-year basis assuming that the (static) equilibrium is reached at each data point. Second, the identification of the multilateral resistance terms is only possible up to a scalar, calling for the use of a normalisation procedure. We decide to set Canada as the country of choice for the normalisation given that this destination is likely not to be strongly affected by the counter-factual policy shock. The idea underlying this choice is the following: if the benchmark country is not impacted by the policy evolution, then the "relative" counter-factual changes in the multilateral resistance terms should be relatively close to their "absolute" counterparts.

The results are presented graphically in four figures where each point estimate corresponds to the yearly-specific percentage change between the baseline and the counter-factual expected number of asylum applications lodged in the host country. Except on Figure 3.6 where we highlight the pattern for some specific source countries, every symbol represents the average change across origins for which we only disentangle whether they are considered as being safe. We start the discussion of the findings by looking at how the model performs to assess the direct bilateral effect of the counter-factual list of safe source countries on the number of asylum applications (Figure 3.4). The retrieved estimates match closely with the coefficients obtained in Section 3.3.2.1, i.e. the policy change leads to a decrease in the number of asylum claims coming from safe origins and lodged in Finland, Italy, Norway and Sweden that ranges between 25% and 30%. This outcome provides support for the validity of the procedure in order to get consistent simulated effects from the model.

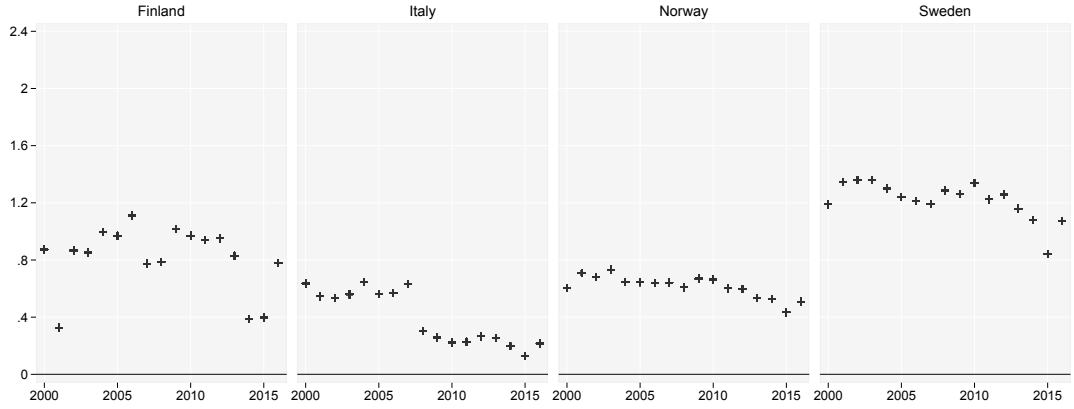
Figure 3.4: Direct effect of the counter-factual policy scenario



Source: Author's elaboration based on simulation of the gravity model.

The main contribution of the paper hinges on the analysis, in a unified framework, of the externalities both arising across origin and destination countries due to changes in asylum policy. On the one hand, the model derived by Anderson (2011) entails that a modification of the list of safe source countries would lead to the redirection of asylum applications from origins not (virtually) considered as being safe to Finland, Italy, Norway, and Sweden. The simulation results averaged across all origin countries (except Albania, Bosnia-Herzegovina, Bulgaria, Macedonia, Montenegro, and Serbia) are presented on Figure 3.5.

Figure 3.5: Redirection effects across origins under the counter-factual scenario

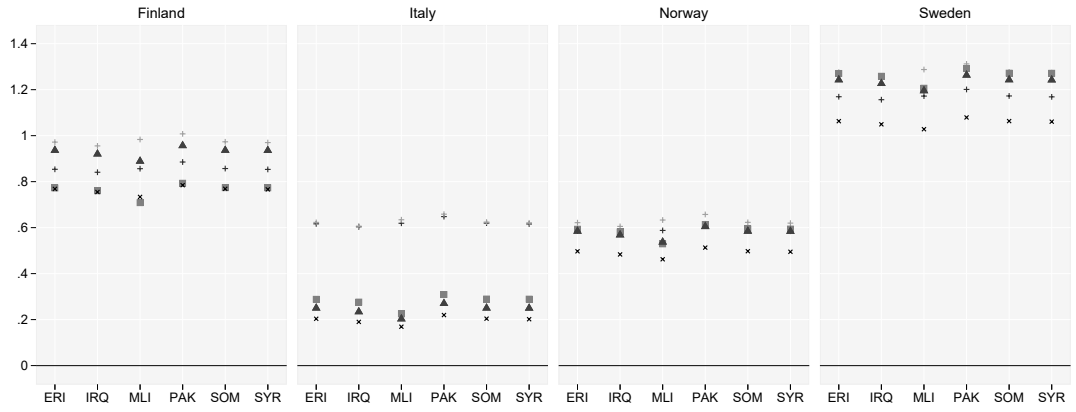


Source: Author's elaboration based on simulation of the gravity model.

They reveal the presence of indirect effects, whereby the number of asylum claims would increase in a range going from 0.2 to 1.5% (compared to the baseline total) in countries that faced the policy change. Moreover, the amplitude of the redirection is heterogeneous over time and across destinations. For instance, Sweden will potentially face 1.5% more applications in 2010, compared with the baseline scenario, while Norway would experience an increase of 0.7%.

The above figure hides the heterogeneity of the estimates among the different origins of the sample. We disaggregate them for a set of specific countries and years on Figure 3.6. The reported coefficients are disentangled by country in each column, while the pattern allow differentiating them by time. They document that the redirection effects are mainly homogeneous across source countries conditional on the host country and the respective year. This outcome is the corollary of the change in the migration costs: the implementation of the hypothetical list of safe source countries results in a similar reduction of the bilateral costs for all sending countries which, in turn, implies that the variation in the attractiveness of each destination is not origin-specific.

Figure 3.6: Redirection effects across specific origins under the counter-factual scenario

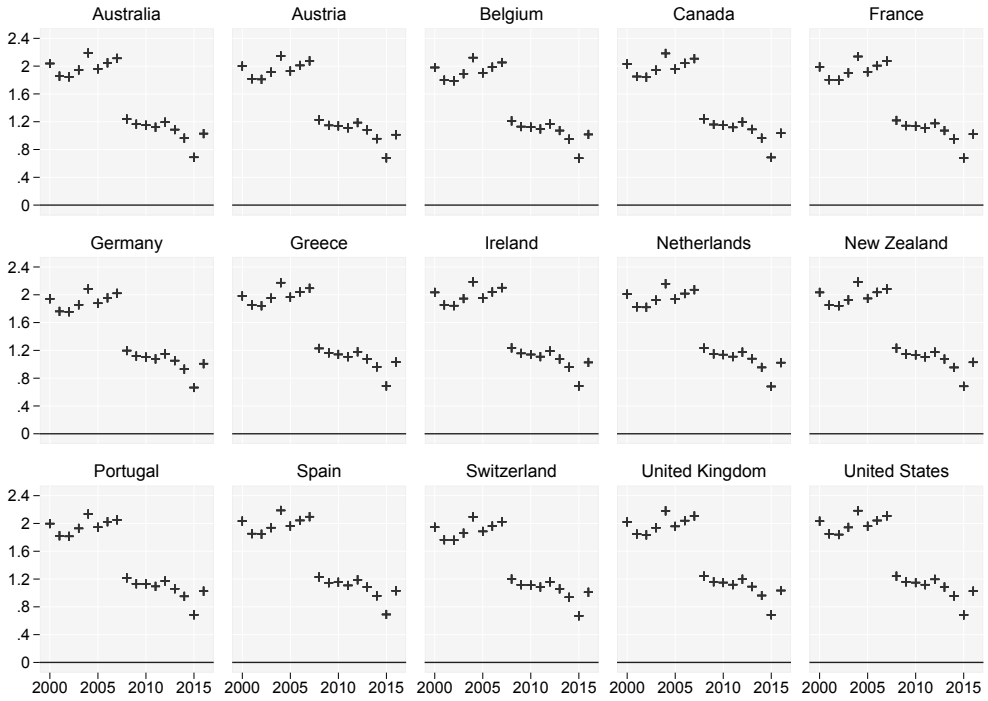


Notes: Black plus, gray plus, square, triangle and cross symbols correspond to the year 2000, 2004, 2008, 2012, and 2016, respectively.
Source: Author's elaboration based on simulation of the gravity model proposed.

Finally, changing the list of safe source countries has also consequences on the redirection of asylum applications across destinations, as already shown for a different migration policy in Bertoli and Fernández-Huertas Moraga (2015). The simulation results are displayed on Figure 3.7, i.e. the counterpart of Figure 3.5 for asylum host countries. They reveal that a fraction of asylum seekers experiencing the new restrictive policy in Finland, Italy, Norway or Sweden is diverted to other OECD member states. The size of the externality effects ranges from 0.5 to 2.3% and does not vary across receiving countries. The explanation follows the argument described in the previous paragraph: the simulated scenario affects homogeneously the bilateral migration costs of host countries not involved in the policy adjustment, so that their relative attractiveness is not altered[18].

[18] In theoretical terms, the similarity of the findings across origins (Figure 3.6) and, respectively, across destinations (Figure 3.7) corresponds to an implication of the well-known Independence of Irrelevant Alternatives (IIA).

Figure 3.7: Redirection effects across destinations under the counter-factual scenario



Source: Author's elaboration based on simulation of the gravity model.

3.4 Conclusion

The gravity model introduced by Anderson (2011) is relevant to evaluate the direct and indirect effects of migration policy that aim at influencing internation flows. We have proposed a methodology to quantify the related externalities in a single framework, which has been applied to assess the consequences of the list of safe source countries on the number of asylum applications lodged in OECD destinations. The econometric study have pointed out the deterrent effect of the safe country policy on the bilateral number of asylum claims, and we have highlighted potential mechanisms that could explain the observed negative coefficient. The latter has been used to solve the gravity model of migration and structurally estimate the redirection effects obtained under a counter-factual experiment associated to the list of safe source countries. The analysis has provided evidence of diversion effects arising across *both* origin and destination countries.

Bibliography

Anderson, J. E. (2011): "The Gravity Model," *Annual Review of Economics*, 3(1), 133–160.

Anderson, J. E., and Y. V. Yotov (2016): "Terms of Trade and Global Efficiency Effects of Free Trade Agreements, 1990-2002," *Journal of International Economics*, 99, 279–298.

Autor, D. H. (2003): "Outsourcing at Will: The Contribution of Unjust Dismissal Doctrine to the Growth of Employment Outsourcing," *Journal of Labor Economics*, 21(1), 1–42.

Beine, M., S. Bertoli, and J. Fernández-Huertas Moraga (2016): "A Practitioners' Guide to Gravity Models of International Migration," *The World Economy*, 39(4), 496–512.

Bertoli, S., and J. Fernández-Huertas Moraga (2013): "Multilateral resistance to migration," *Journal of Development Economics*, 102, 79–100.

Bertoli, S., and J. Fernández-Huertas Moraga (2015): "The Size of the Cliff at the Border," *Regional Science and Urban Economics*, 51, 1–6.

Bertoli, S., J. Fernández-Huertas Moraga, and F. Ortega (2011): "Immigration Policies and the Ecuadorian Exodus," *World Bank Economic Review*, 25(1), 57–76.

Beverelli, C., and G. Orefice (2019): "Migration Deflection: The Role of Preferential Trade Agreements," *Regional Science and Urban Economics*, 79, 103469.

Hatton, T. J. (2004): "Seeking Asylum in Europe," *Economic Policy*, 19(38), 5–62.

——— (2009): "The Rise and Fall of Asylum: What Happened and Why?," *The Economic Journal*, 119(535), F183–F213.

——— (2016): "Refugees, Asylum Seekers, and Policy in OECD Countries," *American Economic Review*, 106(5), 441–45.

McFadden, D. (1974): "Conditional Logit Analysis of Qualitative Choice Behavior," in *Frontiers in Econometrics*, ed. by P. Zarembka, pp. 105–142. New York: Academic Press.

Meyers, E. (2004): *International Immigration Policy: A Theoretical and Comparative Analysis*. Palgrave Macmillan US.

Neumayer, E. (2004): "Asylum Destination Choice: What Makes Some West European Countries More Attractive Than Others?," *European Union Politics*, 5(2), 155–180.

Ortega, F., and G. Peri (2013): "The Role of Income and Immigration Policies in Attracting International Migrants," *Migration Studies*, 1(1), 47–74.

Thielemann, E. (2004): "Why Asylum Policy Harmonisation Undermines Refugee Burden-Sharing," *European Journal of Migration and Law*, 6(1), 47–65.

——— (2006): "The Effectiveness of Governments' Attempts to Control Unwanted Migration," in *Immigration and the Transformation of Europe*, ed. by C. Parsons, and T. Smeeding. Cambridge University Press.

Appendix

A.1 Details on the list of safe countries of origin

Table A1.1: List of safe source countries in the nine relevant receiving countries

	Austria	Belgium	Canada	France	Germany	Ireland	Netherlands	Switzerland	UK
Albania	12-2012	09-2016		05-2006	12-2015		11-2015	10-1993	04-2003
				(02-2008)					
				03-2011					
				(04-2012)					
				01-2014					
Algeria	02-2016						10-2016		
Armenia				12-2009					
				(08-2010)					
				12-2011					
Benin				07-2005				01-2007	
Burkina Faso								04-2009	
Bangladesh				12-2011					08-2003
				(03-2013)					(05-2005)
Bulgaria	08-2005				07-1993	09-2003[b]	(03-2001)	03-1991	04-2003
							11-2015		(01-2007)
Bosnia-Herz.	07-2009	06-2012		07-2005	11-2014		11-2015	08-2003	08-2007
Bolivia									08-2003
Cape Verde				07-2005					
Ecuador									08-2003
Georgia	02-2016	09-2016		07-2005			10-2016		
				(12-2009)					
				01-2014					
Ghana	02-2016			07-2005	07-1993		(03-2001)	10-1993	12-2005[c]
							02-2016		
Gambia								10-1993	08-2007[c]
Croatia	07-2009		12-2012	07-2005	07-2013	09-2003[b]	11-2015	01-2007	
				(07-2013)					
Hungary	08-2005		12-2012		07-1993	12-2004[b]	(03-2001)	08-2003	11-2002
							11-2015		
India		06-2012		07-2005			02-2016	03-1991	02-2005
Jamaica							02-2016		04-2003
Kosovo	07-2009	06-2012		03-2011	12-2015		11-2015	04-2009	03-2010
				(04-2012)					
				01-2014					
				(10-2014)					
				10-2015					
Kenya									08-2007[c]
Liberia									08-2007[c]
Morocco	02-2016						02-2016		
Moldova				12-2011				01-2007	04-2003
Madagascar				05-2006					
				(08-2010)					
Macedonia	07-2009	06-2012		05-2006	11-2014		11-2015	08-2003	04-2003
Mali				07-2005[a]				01-2007	08-2007[c]
				(01-2013)				(03-2012)	
Mauritius									08-2007
Montenegro	07-2009	06-2012		12-2011	12-2015		11-2015	01-2007	04-2003
Mongolia				07-2005			02-2016	07-2000	12-2005
Malawi									08-2007[c]
Niger				05-2006					
				(02-2008)					
Nigeria									12-2005[c]
Peru									08-2007
Romania	08-2005		10-2014		07-1993	09-2003[b]	(03-2001)	11-1991	04-2003
							11-2015		(01-2007)
Senegal				07-2005	07-1993		(03-2001)	10-1993	
							02-2016		
Sierra Leone									08-2007[c]
Serbia	07-2009	06-2012		12-2009	11-2014		11-2015	04-2009	04-2003
Sri Lanka									08-2003
									(12-2006)
Togo							12-2016		
Tunisia	02-2016						10-2016		
Tanzania				05-2006					
				(10-2015)					
Trinidad-Tobago							05-2017		
Ukraine				07-2005			10-2016	01-2007	08-2003
				(04-2014)				(07-2014)	

Table A1.1: List of safe source countries in the nine relevant receiving countries (cont.)

	Austria	Belgium	Canada	France	Germany	Ireland	Netherlands	Switzerland	UK
Andorra			10-2014				11-2015		
Australia			02-2013				11-2015		
Austria			12-2012		01-2000[d]		11-2015	08-2003	
Belgium	08-2005		12-2012		01-2000[d]		11-2015	08-2003	
Brazil							05-2017		08-2003
Canada	08-2005						11-2015		
Chile			06-2013						
Cyprus	08-2005		12-2012		05-2005	09-2003[b]	11-2015	08-2003	11-2002
Czech Republic	08-2005		12-2012		07-1993	09-2003[b]	(03-2001)	08-2003	11-2002
							11-2015		
Denmark	08-2005		12-2012		01-2000[d]		11-2015	08-2003	
Estonia	08-2005		12-2012		05-2005	09-2003[b]	11-2015	08-2003	11-2002
Finland	08-2005		12-2012		01-2000[d]		11-2015	10-1993	
France	08-2005		12-2012		01-2000[d]		11-2015	08-2003	
Germany	08-2005		12-2012				11-2015	08-2003	
Greece	08-2005		12-2012		01-2000[d]		11-2015	08-2003	
Iceland	08-2005		02-2013				11-2015	08-2003	
Ireland	08-2005		12-2012		01-2000[d]		11-2015	08-2003	
Israel			02-2013						
Italy	08-2005		12-2012		01-2000[d]		11-2015	08-2003	
Japan			02-2013				11-2015		
Latvia	08-2005		12-2012		05-2005	09-2003[b]	11-2015	08-2003	11-2002
Liechtenstein	08-2005		10-2014				11-2015	08-2003	
Lithuania	08-2005		12-2012		05-2005	09-2003[b]	11-2015	06-1998	11-2002
Luxembourg	08-2005		12-2012		01-2000[d]		11-2015	08-2003	
Malta	08-2005		12-2012		05-2005	09-2003[b]	11-2015	08-2003	11-2002
Mexico			02-2013						
Monaco			10-2014				11-2015		
Netherlands	08-2005		12-2012		01-2000[d]			08-2003	
Norway	08-2005		02-2013				11-2015	08-2003	
New Zealand	08-2005		02-2013				11-2015		
Poland	08-2005		12-2012		07-1993	09-2003[b]	(03-2001)	08-2003	11-2002
							11-2015		
Portugal	08-2005		12-2012		01-2000[d]		11-2015	08-2003	
San Marino			10-2014				11-2015		
Slovakia	08-2005		12-2012		07-1993	09-2003[b]	(03-2001)	08-2003	11-2002
							11-2015		
Slovenia	08-2005		12-2012		05-2005	09-2003[b]	11-2015	08-2003	11-2002
South Africa						12-2004[b]			08-2003
South Korea			06-2013						03-2010
Spain	08-2005		12-2012		01-2000[d]		11-2015	08-2003	
Sweden	08-2005		12-2012		01-2000[d]		11-2015	08-2003	
Switzerland	08-2005		02-2013				11-2015		
Turkey				12-2009					
				(08-2010)					
United Kingdom	08-2005		12-2012		01-2000[d]		11-2015	08-2003	
United States			12-2012				11-2015		

Notes: Dates in parentheses indicate that the country was dropped from the list of safe country of origin, otherwise the dates represent the time when the source country was added to the list of the respective destination country. *a* – on July 23, 2010 (i.e., 08-2010 in the data), Mali was maintained in the list but only for male asylum seekers. *b* – the International Protection Act that came into effect on December 31, 2016 revoked the list of safe source country, i.e. the safe country indicator is equal to 0 for Ireland from January 2017 onwards. *c* – these countries were included in the list of safe country of origin only for male asylum seekers. *d* – these countries were automatically added in the list of safe source countries because they are EU Member States. The date refers simply to the starting date of the time period considered, since they have been EU Member States before January 2000.
Source: Author's elaboration based on self-collected data.

A.1.2 Sources of information

The notes written below the figures and tables inserted in the main text refer to self-collected data as the source of information about the destination-specific list of safe countries of origin. This has been done for the sake of clarity and this section aims at clarifying the material that was gathered to extract information on the evolution of the safe country policy across countries and over time.

Table A1.2: List of safe countries of origin: sources of information

Destination	Source of information	Website
Multiple	Asylum Information Database (AIDA) country reports	https://www.asylumineurope.org/reports
	Asylum Procedures: Report on Policies and Practices in IGC Participating States	http://www.igc-publications.ch/
	European Parliament	http://aei.pitt.edu/4906/1/4906.pdf
	European Commission	https://ec.europa.eu/home-affairs/sites/homeaffairs/files/e-library/docs/pdf/safe_countries_2004_en_en.pdf
Canada	Government of Canada	https://www.canada.ca/en/immigration-refugees-citizenship/services/refugees/claim-protection-inside-canada/apply/designated-countries-policy.html
Ireland	Irish Statute Book	http://www.irishstatutebook.ie/eli/2003/si/422/made/en/print http://www.irishstatutebook.ie/eli/2004/si/714/made/en/print
UK	The National Archives	http://www.legislation.gov.uk/ukpga/2002/41/pdfs/ukpga_20020041_en.pdf http://www.legislation.gov.uk/uksi/2003/970/made http://www.legislation.gov.uk/uksi/2003/1919/made http://www.legislation.gov.uk/uksi/2005/330/article/2/made http://www.legislation.gov.uk/uksi/2005/1016/article/2/made http://www.legislation.gov.uk/uksi/2005/3306/made http://www.legislation.gov.uk/uksi/2006/3215/article/2/made http://www.legislation.gov.uk/uksi/2006/3275/article/2/made https://www.legislation.gov.uk/uksi/2007/2221/made http://www.legislation.gov.uk/uksi/2010/561/article/3/made
Switzerland	State Office for Migration	https://www.sem.admin.ch/sem/de/home/aktuell/news/2009/ref_2009-03-19.html https://www.sem.admin.ch/sem/fr/home/aktuell/news/2012/ref_2012-09-141.html https://www.sem.admin.ch/sem/fr/home/aktuell/news/2014/ref_2014-06-201.html
France	Wikipedia	https://fr.wikipedia.org/wiki/Pays_d%27origine_s%C3%BBr_en_droit_fran%C3%A7ais_de_l%27asile
	National website	http://www.rtw.fr/pays.htm
Germany	Federal Law Gazette	https://www.bgbl.de/xaver/bgbl/start.xav?start=%2F%2F*%5B%40attr_id%3D%27bgbl193s1062.pdf%27%5D#__bgbl__%2F%2F*%5B%40attr_id%3D%27bgbl193s1062.pdf%27%5D__1515413943342 https://www.bgbl.de/xaver/bgbl/start.xav?start=%2F%2F*%5B%40attr_id%3D%27bgbl195s0430.pdf%27%5D#__bgbl__%2F%2F*%5B%40attr_id%3D%27bgbl195s0430.pdf%27%5D__1515595956360
	National website	https://www.buzer.de/s1.htm?g=AsylVfG+27.08.2007&a=Anlage+II https://www.buzer.de/s1.htm?g=AsylVfG+5.11.2014&a=Anlage+II
	Wikipedia	https://de.wikipedia.org/wiki/Sicherer_Herkunftsstaat_(Deutschland)#2004/07:_Austragen_der_in_die_EU_beigetretenen_Staaten
Austria	Legal Information System	https://www.ris.bka.gv.at/Dokumente/BgblAuth/BGBLA_2005_I_100/BGBLA_2005_I_100.pdf https://www.ris.bka.gv.at/Dokumente/BgblAuth/BGBLA_2009_II_177/BGBLA_2009_II_177.pdf https://www.ris.bka.gv.at/Dokumente/BgblAuth/BGBLA_2010_II_428/BGBLA_2010_II_428.pdf https://www.ris.bka.gv.at/Dokumente/BgblAuth/BGBLA_2016_II_47/BGBLA_2016_II_47.pdf

Table A1.2: List of safe countries of origin: sources of information (cont.)

Destination	Source of information	Website
Netherlands	Central Government Information	https://www.rijksoverheid.nl/onderwerpen/asielbeleid/vraag-en-antwoord/lijst-van-veilige-landen-van-herkomst
		https://www.government.nl/topics/asylum-policy/question-and-answer/list-safe-countries-of-origin
	Meyers (2004)	https://www.palgrave.com/de/book/9780312231439#reviews
Belgium	Belgian Official Journal	http://www.etaamb.be/fr/arrete-royal-du-26-mai-2012_n2012000365.html
		http://www.etaamb.be/fr/arrete-royal-du-07-mai-2013_n2013000322.html
	Federal Public Service - Justice	http://www.ejustice.just.fgov.be/cgi/article_body.pl?language=fr&caller=summary&pub_date=14-05-15&numac=2014000310
		http://www.ejustice.just.fgov.be/cgi/article_body.pl?language=fr&caller=summary&pub_date=15-05-15&numac=2015000227
		http://www.ejustice.just.fgov.be/cgi/article_body.pl?language=fr&caller=summary&pub_date=16-08-29&numac=2016000489
		http://www.ejustice.just.fgov.be/cgi/article_body.pl?language=fr&caller=summary&pub_date=17-12-27&numac=2017031718
	Council of State	http://www.raadvst-consetat.be/Arrets/228000/900/228901.pdf
		http://www.raadvst-consetat.be/Arrets/228000/900/228902.pdf
		http://www.raadvst-consetat.be/Arrets/231000/100/231157.pdf
		http://www.raadvst-consetat.be/Arrets/235000/200/235211.pdf

Source: Author's elaboration based on self-collected information.

We have mostly relied on the Asylum Information Database (AIDA) country reports and the "Asylum Procedures: Report on Policies and Practices in IGC Participating States" published by the intergovernmental consultations (IGC) on migration, asylum and refugees. Beyond these two corpuses, some official websites tracking and reporting changes in migration law enacted in the respective countries (e.g., Canada and the United Kingdom) have been exploited. We have also drawn on some reports delivered by European institutions, in particular to get information on the safe country policy at the beginning of the time period covered by our sample. Finally, we have complemented the above data with some country-specific information, e.g. from Wikipedia. Details about the data collection are presented in Table A1.2.

A.2 Robustness checks - Direct effect of the list of safe source countries

The robustness of the negative effect of the list of safe countries of origin on the number of asylum applications is challenged through different tests. First, the sample is restricted to the ten countries (i.e., Germany, France, United Kingdom, Sweden, Italy, Canada, United States, Austria, Belgium and Switzerland) that received the highest number of asylum applications over the period 2000-2017. Second, non-European countries are further excluded from the set of destination countries (i.e., Canada and United States are dropped from the previous list). Third, the sample is constrained to the main source countries, i.e. origins for which the share of applications corresponds to more than 5% of the total number of asylum claims each year. Fourth, the analysis is replicated with information related only to 2012 and later years (column (4)), whereas the counterpart is done is done in column (5) for the time period 2000-2012. Finally, the estimation is done on a sample that drops data for 2017, since UNHCR figures between 1 and 4 have been replaced with an asterisk to protect the anonymity of individuals and are therefore represented by missing values in the baseline sample.

Table A2.1: Robustness tests - Direct effect of the list of safe countries of origin

	Dependent variable: m_{jkt}					
	(1)	(2)	(3)	(4)	(5)	(6)
SCO_{jkt}	-0.460***	-0.450***	-0.343***	-0.269**	-0.402***	-0.342***
	(0.061)	(0.063)	(0.086)	(0.136)	(0.054)	(0.068)
μ_{dt}	✓	✓	✓	✓	✓	✓
μ_{ot}	✓	✓	✓	✓	✓	✓
μ_{ody}	✓	✓	✓	✓	✓	✓
Total sample	231,109	231,109	231,109	231,109	231,109	231,109
Selection dropped	73,820	115,866	125,129	157,444	73,665	8,417
Singletons dropped	9,370	8,529	5,532	4,118	8,199	11,828
Estimation sample	147,919	106,714	100,448	69,547	149,245	210,864
Pseudo-R^2	0.961	0.964	0.964	0.970	0.914	0.952

Notes: All regressions have been estimated with PPML using the command `ppmlhdfe`. Clustered standard errors by country-pair in parentheses. ***, **, and * denote significance at the 1, 5, and 10 percent levels, respectively.

Source: Author's elaboration based on UNHCR monthly asylum applications, and self-collected information about the destination-specific list of safe source countries.

Conclusion

Le système migratoire mondial et les flux entre les différents agents de ce réseau global s'intensifient et se diversifient. Mus par la croissance des inégalités et des disparités économiques, démographiques et socio-politiques, de plus en plus d'individus décident de quitter leur pays d'origine pour chercher en d'autres lieux, au-delà des frontières nationales, des perspectives de vie plus sûres et/ou meilleures. À ces forces avivant l'expansion des flux migratoires, s'ajoutent les changements climatiques, en cours et à venir, qui seront probablement responsables de nouveaux déplacements à l'échelle mondiale. Les pressions migratoires iront de pair avec des évolutions dans les politiques associées aux migrations économiques et/ou forcées, à la fois dans les pays d'origine et de destination.

Si les travaux en lien avec l'économie des migrations ont déjà produit plusieurs résultats fondamentaux, les nombreux défis actuels ou futurs coïncident avec les multiples questions qui restent en suspens quant aux effets des migrations internationales sur les migrants eux-mêmes, leurs pays d'origine, les pays de destination et leurs résidents. La présente thèse développe trois chapitres sur des sujets variés, ayant chacun pour but de contribuer à la recherche sur les migrations, notamment ses causes et ses conséquences.

Le chapitre I permet d'améliorer les (rares) connaissances sur les caractéristiques, en particulier en termes d'éducation, des demandeurs d'asile qui ont récemment réussi à atteindre l'Europe en provenance de diverses régions, par rapport à la population restée dans le pays d'origine. Le chapitre II offre des contributions théoriques et empiriques à la littérature sur les migrations, et également sur la recherche concernant les modèles d'inattention rationnelle. La principale conclusion suggère que les migrants internationaux sont rationnellement inattentifs, et ce même si les enjeux liés à leurs décisions de localisation sont certainement très élevés. Enfin, le chapitre III apporte des éléments de réponse sur l'évolution et les effets des listes de pays d'origine sûrs entre les membres de l'OCDE. L'étude montre que cette politique d'asile conduit à une baisse des demandes d'asile bilatérales et à des effets de diversion entre les pays d'origine et les pays de destination.

Pour conclure, les trois essais de cette thèse ont tous en commun, non seulement le thème des migrations internationales, mais aussi et surtout la volonté de trouver des explications cohérentes afin de mieux comprendre les décisions de migrer, les caractéristiques individuelles des migrants et les répercussions politiques des flux migratoires. Il en va d'une amélioration de la compréhension générale autour de ces questions, susceptible de transférer des informations précieuses pour une juste appréciation de l'adaptation des migrants et de leurs effets dans les pays de destination, ainsi qu'une meilleure évaluation de leurs interactions avec les pays d'origine.

"Bien que la littérature ait beaucoup voyagé, l'étude de l'économie de l'immigration reste une entreprise intellectuellement dynamique : il y a encore beaucoup à apprendre."[1]

Georges Borjas, *Immigration Economics* (2014).

[1] Traduction du texte original: "Although the literature has traveled far, the study of the economics of immigration remains an intellectually vibrant endeavor: there is still much to learn."