



**HAL**  
open science

# Birds of a feather cannot always flock together : essays on the socio-economic impacts of local diversity

Camille Hémet

► **To cite this version:**

Camille Hémet. Birds of a feather cannot always flock together : essays on the socio-economic impacts of local diversity. Economics and Finance. Institut d'études politiques de paris - Sciences Po, 2013. English. NNT : 2013IEPP0041 . tel-03522695

**HAL Id: tel-03522695**

**<https://theses.hal.science/tel-03522695>**

Submitted on 12 Jan 2022

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

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

Institut d'Études Politiques de Paris  
ÉCOLE DOCTORALE DE SCIENCES PO  
Programme doctoral en économie  
Département d'économie de Sciences Po  
  
Doctorat en Sciences économiques

**Birds of a feather cannot always flock together.**  
*Essays on the socio-economic impacts of local diversity.*

Camille Hémet

*Thèse dirigée par Yann Algan, Professeur des Universités, IEP de Paris*

Soutenue à Paris le 20 novembre 2013

Jury:

M. Yann Algan	Professeur des Universités, IEP de Paris
M. Bruno Deceuse	Professeur des Universités, Université d'Aix-Marseille, <i>Rapporteur</i>
M. Alan Manning	Professor of Economics, London School of Economics and Political Science
M. Thierry Mayer	Professeur des Universités, IEP de Paris
M. Yves Zenou	Professor of Economics, Stockholm University, <i>Rapporteur</i>



Institut d'Études Politiques de Paris  
ÉCOLE DOCTORALE DE SCIENCES PO  
Programme doctoral en économie  
Département d'économie de Sciences Po  
  
Doctorat en Sciences économiques

Qui se ressemble ne s'assemble pas toujours.  
*Essais sur les effets socio-économiques de la diversité locale.*

Camille Hémet

*Thèse dirigée par Yann Algan, Professeur des Universités, IEP de Paris*

Soutenue à Paris le 20 novembre 2013

Jury:

M. Yann Algan	Professeur des Universités, IEP de Paris
M. Bruno Deceuse	Professeur des Universités, Université d'Aix-Marseille, <i>Rapporteur</i>
M. Alan Manning	Professor of Economics, London School of Economics and Political Science
M. Thierry Mayer	Professeur des Universités, IEP de Paris
M. Yves Zenou	Professor of Economics, Stockholm University, <i>Rapporteur</i>



# Remerciements

Avec la fin de cette thèse vient le moment de remercier tous celles et ceux qui, de près ou de loin, y ont contribué, et m'ont permis d'aller au bout de cette aventure.

Je tiens en premier lieu à exprimer ma gratitude à mon directeur, Yann Algan, qui m'a accompagnée et encouragée durant ces quatre années de doctorat. J'ose croire que son enthousiasme pour la recherche, sa clairvoyance et sa capacité à mettre en lumière des idées passionnantes continueront de m'influencer - mes *punchlines* finiront peut-être même par être assez *sharp*. Je le remercie également de m'avoir laissé une grande liberté dans ma recherche.

Je suis sincèrement reconnaissante à Bruno Decreuse et Yves Zenou d'avoir accepté d'être rapporteurs de cette thèse, ainsi qu'à Alan Manning et Thierry Mayer qui ont accepté de faire partie de mon jury. Je sais que leurs remarques et conseils me seront précieux en ce début de carrière.

Mes remerciements s'adressent également à mes co-auteurs, Yann Algan et David Laitin, qui ont directement contribué à cette thèse, et dont l'influence sur ma recherche de façon plus générale est immense. Travailler à leurs côtés m'aura énormément appris, tant du point de vue de la méthode que des idées, et j'espère que les fruits de notre collaboration seront bientôt cueillis. Je suis aussi reconnaissante à David de m'avoir ouvert les portes de son département à Stanford.

L'initiative de cette thèse doit beaucoup à ceux dont les enseignements et les conseils ont jalonné mes études d'économie. C'est d'abord les cours suivis sur les bancs de l'Université de Cergy-Pontoise qui m'ont donné le goût de cette discipline. Je remercie notamment Régis Renault de m'avoir orientée vers la recherche à la fin de la licence. J'ai également une pensée particulière pour Jeanne Hagenbach, qui m'a donné envie de l'imiter alors qu'elle n'était encore "que" chargée de TD. L'exigence et la qualité des enseignements reçus pendant mes années de master à l'École d'Économie de Paris ont bien entendu confirmé mon envie de faire de la recherche.

Au cours de mes années de doctorat, j'ai eu la chance d'être accueillie dans deux laboratoires de recherche, dont le dynamisme et les conditions de travail exceptionnelles m'ont permis de réaliser cette thèse dans un cadre privilégié. J'ai pu y bénéficier des con-

seils et de l'écoute de nombreux chercheurs, qui m'ont aidé à progresser dans mon travail. Je remercie ainsi l'ensemble des chercheurs du département d'économie de Sciences Po, et notamment Yann Algan, Ruben Durante, Roberto Galbiati, Alfred Galichon, Emeric Henry, Elise Huillery, Koen Jochmans, Guy Laroque, Philippe Martin, Thierry Mayer, Jean-Marc Robin et Etienne Wasmer pour avoir pris le temps de me recevoir et de répondre à mes questions. La bonne humeur et les encouragements de Ruben ont largement contribué à rendre la vie au département agréable. Je suis par ailleurs très reconnaissante à Thierry qui n'a jamais compté le temps passé à prodiguer d'innombrables conseils en termes d'orientation professionnelle. Je tiens également à remercier Yves Zenou pour ses conseils, ses recommandations, et pour ses cours sur les réseaux sociaux et sur l'économie urbaine qui ont été une véritable source d'inspiration.

À l'École d'Économie d'Aix-Marseille, je suis infiniment reconnaissante à Yann Bramoullé, Pierre-Philippe Combes, Bruno Deceuse, Cécilia Garcia-Peñalosa, Christian Schluter et Tanguy Van Ypersele qui ont su faire preuve d'une grande écoute, réfléchissant à mes problèmes, proposant des solutions et élargissant le champs de mes réflexions. Les discussions animées sur le fonctionnement du monde académique m'ont également beaucoup appris et me seront sans doute fort utiles dans les années à venir. Je remercie aussi Patricia Augier, Olivier Bargain, Stephen Bazen, Sebastian Bervoets, Raouf Boucekkine, Renaud Bourlès, Marion Dovis, Alice Fabre, Nicolas Gravel, Charles Lai Tong, Michel Lubrano, Christophe Muller, Carine Nourry, Anne Peguin-Feissolle, Patrick Pintus, Thomas Seegmuller, Antoine Soubeyran, Alain Trannoy et Federico Trionfetti qui ont contribué à faire de cette année à Aix-Marseille une expérience enrichissante.

Je tiens également à remercier les chercheurs rencontrés au gré des séminaires et lors du trimestre passé à Stanford avec lesquels j'ai pu avoir des échanges constructifs. Je pense notamment à Alberto Alesina, Eliana La Ferrara, Giacomo De Giorgi, Jamie Goodwin-White, John Huber, Matthew Jackson, Ann-Sofie Kolm, Rafael Lalive, Birthe Larsen, David Neumark, Leah Platt Boustan, Daniel Posner et Mahtias Thoenig.

Bien entendu, ces quatre dernières années auraient été bien plus compliquées sans l'aide, l'efficacité et la gentillesse de Cathy Bénard, Pilar Calvo, Claudine Lamaze, Cécile Markarian et Michelle N'Cho à Sciences Po, Eliana Vasquez et Jackie Sargent à Stanford, et Diane Djenderedjian (dont les services et les petites attentions sont allés bien au-delà du cadre professionnel), Elisabeth L'Huillier, Isabelle Mauduech et Mathilde Martelli à l'AMSE. Je les en remercie chaleureusement.

Cette thèse a par ailleurs été pour moi l'occasion de me familiariser avec l'enseignement. J'ai ainsi pu prendre pleinement conscience de l'importance de cette activité, complémentaire à la recherche, et tout aussi passionnante et gratifiante. Je remercie donc Yann Algan, Ruben Durante et Lionel Fontagné de m'avoir confié des groupes de travaux dirigés. Je

n'oublie pas Etienne Wasmer, qui m'a de plus donné l'opportunité de faire face à mon premier "amphi" dans le cadre de révisions de partiels. Mes anciens élèves ne sont pas étrangers non plus à mon envie de poursuivre dans cette voie. Qu'ils aient été sérieux ou dissipés, motivés ou désintéressés, ils ont tous contribué à me faire aimer l'enseignement avec les satisfactions et les défis que cela implique.

Mes pensées se tournent maintenant vers les doctorants et anciens doctorants, ATER, postdoc, assistants de recherche et informaticiens que j'ai côtoyés pendant toutes ces années à Sciences Po: Gabriel, Jérôme, Gong, Lucile, Thomas, Camilo (le roi du chocolat), Nina, Marc, Jules, Alexis, Guillaume, Meradj, Jamil, Dilan, Marlène, Liza, et plus récemment Guillaume, Valéria, Elisa, "Mazzini", Marta, Ivan. J'ai une pensée particulière pour Sarra, Arthur, Xavier, Adrien et bien sûr Maxime, Marion et Matthieu qui approchent de la fin aussi. Courage... Merci à Florian pour les moments de détente dans les dernières semaines. Un grand merci à Elizabeth pour son aide avec l'anglais. Merci à Maxime pour son entrain à répondre à toutes les questions d'économétrie et pour sa pédagogie. Merci à Anne et Elise pour leur bonne humeur et pour de bonnes parties de rigolade, pour ne pas dire autre chose. Merci aux Marseillais, Morgane et Clément pour avoir apporté un peu de mer et de soleil au département. A François pour sa gentillesse et ses bons mots, à Marion pour son oreille attentive à tous les ragots. À tous les quatre, un grand merci pour toutes ces soirées qui ne se terminaient qu'à l'approche du dernier bus de nuit - parce qu'on a bien le droit de s'amuser un peu !

Je pense aussi aux doctorants (et assimilés) d'Aix-Marseille: Antoine, Karine, Gilles, Thomas, Cyril, Kalila, Lise, Nicolas, Paolo, Nick, Martha, Qays, Anne-Sarah, Marion, Anastasia, Natalya, Antoine, Vivien et tous les autres. Merci pour votre accueil. Je tiens enfin à remercier Greg, Matthieu et Marc sans qui cette année à Aix n'aurait pas eu la même allure. Nos parties de pétanque et de mölkky me manquent déjà, de même que les dimanches au château (même si je n'étais pas la plus assidue là-dessus). Greg, merci encore pour toutes les relectures et corrections de papiers et lettres de motivation, j'ai une sacrée dette envers lui ! Merci aussi à Matthieu pour les gâteaux à la broche, farçous et autres découvertes culinaires. Marc, mon "grand frère de thèse", merci infiniment pour tous ses conseils et son soutien. Merci aussi à Julien, fin gourmet et teufeur fou, dont le court passage à Aix m'aura néanmoins laissé de nombreux et bons souvenirs. Merci enfin à Aurore pour sa gentillesse et son accueil.

On l'aura compris, une thèse ne se résume pas à une vie de labo, rythmée par les séminaires et les cours. C'est aussi des pauses café, des soirées, des discussions, des mots de soutien et d'encouragement entre collègues, qui deviennent ainsi parfois plus que de simples collègues. Anne, Elise, Maxime, Morgane, Clément, Marion, François, Matthieu,



Greg et Marc sont évidemment de ceux-là. Je tiens en plus à remercier Etienne et Lisa particulièrement pour leur amitié. Merci à Etienne pour d'innombrables soirées à refaire le monde autour d'un (un ?) verre de bon vin, de la rue d'Ulm à Asnières, en passant par San Francisco, Los Angeles et Notre Dame de Lorette. Affaire à suivre à Bristol. Merci à Lisa pour son soutien permanent et ses encouragements qui ont été particulièrement précieux pendant cette dernière année. Risotto aux cèpes, me voilà !

Mes remerciements s'adressent aussi à tous les amis qui m'ont soutenue pendant ces années. Je pense en particulier à mes amies de toujours, Caroline et Solène, les meilleures pour passer des après-midis entre filles à se changer les idées, ainsi qu'à Nico et Melaine. Je pense aussi à la grande Camille et Etienne, à Laure et Guillaume, à Catherine et Rémi, à Barbara et Arnaud, à Camille et Mathieu que la distance et les années ne m'ont jamais empêchée de retrouver avec grand plaisir. Je remercie Bastien, Ekédi et Johann, Liliana, et Séverine pour leurs encouragements et pour la force qu'ils dégagent.

I also want to express my gratitude to David and Delia for welcoming me into their home and for taking me around when I first went to the US. Obviously, I am also grateful to my American family, the Griffiths. Thank you Julye, Gary, Tiffany, little Griff for showing me the American way of life, taking me to concert, hockey game and Halloween party, and for sharing your thanksgiving meal with me.

Merci aussi à Manu, Nico, Gil et Zouzou qui m'ont accueillie chez eux, facilitant mes multiples périodes de transition.

Je suis bien sûre reconnaissante à la famille Winter, Catherine, Philippe, Sandra, Nico, Camille, et aussi Betty et Jean qui m'ont si bien accueillie, sans jamais s'offusquer de tous ces week-ends passés chez eux à travailler.

Mes pensées les plus chères se tournent à présent vers ma famille qui m'a soutenue depuis le début sans toujours bien savoir ce dont il retournait. J'ai une pensée émue pour ma soeur Joanne et mes frères Simon et Pierre qui ont su devenir les personnes qu'ils avaient envie d'être. Un immense merci à Maman, pour son amour indéfectible, et pour avoir toujours cru en nous, et à Isa qui partage nos vies depuis tant d'années.

Enfin, je ne saurais oublier de remercier Zouzou qui partage ma vie depuis bientôt quatre ans, et qui aura supporté mes sautes d'humeur sans jamais s'en plaindre. Merci aussi pour son aide en cette fin de thèse, qui m'aura fait gagner un temps précieux. Surtout, je lui suis infiniment reconnaissante de m'avoir toujours soutenue et encouragée, y compris lorsqu'il s'agissait de me laisser partir, aux Etats-Unis, puis à Aix, et maintenant à Barcelone. Merci d'avoir compris l'importance de ces choix, merci pour tout.





# Abstract

Diversity reflects the extent to which members of a given community share different characteristics, usually pertaining to ethnicity, socio-economic status or even culture. As a result of trade development and economic integration, modern societies have to cope with increasing levels of diversity, both in terms of origins and social statuses. The purpose of this dissertation is to assess the social and economic impacts of local diversity. More precisely, this thesis shows how neighborhood level diversity affects individuals' living conditions and employment prospects. This work contributes to the existing literature in three ways: it examines unexplored issues at a very local level, gives new insights about the underlying mechanisms and provides new methods to address the endogeneity issue. The first chapter shows that diversity in terms of origins has a negative effect on the quality of local public goods. This not only due to vandalism, not deterred by social policing, but also due to collective action failure to ensure effective property management. However, diversity has no robust effect on public safety. Chapter two reveals that the effect of unemployment on crime has a spatial dimension: for economic crimes the effect of unemployment rate in surrounding neighborhoods is stronger than that of the immediate neighborhood, while the reverse holds for vandalism. The third chapter shows that people living in more diverse neighborhoods face lower employment prospects than those living in more homogeneous areas. An additional result is that this effect is more related to cultural (e.g. language) diversity than to ethnic diversity. Finally, chapter four develops a model rationalizing the fact that ethnic minorities turn to the informal economy in response to adverse labor market conditions.



# Résumé

La notion de diversité reflète le fait que les membres d'une communauté diffèrent selon certaines caractéristiques, liées à l'origine ethnique, au statut socio-économique ou à la culture. L'essor du commerce et l'intégration économique placent les sociétés modernes face à des niveaux de diversité croissants. L'objectif de cette thèse est d'évaluer l'impact social et économique de la diversité au niveau local. Plus précisément, ce travail montre comment la diversité d'un quartier influe sur les conditions de vie et les perspectives d'emploi de ses habitants. Cette thèse contribue à la littérature existante de trois façons: en étudiant des questions inexplorées à un niveau très local, en révélant les mécanismes sous-jacents et en élaborant de nouvelles méthodes permettant de contourner les problèmes d'endogénéité. Le premier chapitre montre que la diversité des origines a un effet négatif sur la qualité des biens publics locaux. Cela s'explique non seulement par des actes de vandalisme liés à un manque de pression des pairs, mais aussi par une gestion inefficace des immeubles d'habitation du fait de l'échec de l'action collective dans les quartiers les plus diversifiés. Cependant, aucun effet robuste de la diversité sur la sécurité publique n'est à noter. Le chapitre deux révèle que l'effet du chômage sur la criminalité a une dimension spatiale. Pour les crimes économiques, le taux de chômage des quartiers environnants a un effet plus fort que celui du voisinage immédiat, alors qu'on observe l'effet inverse pour le vandalisme. Le troisième chapitre montre que les personnes vivant dans des quartiers où la diversité est plus élevée ont des perspectives d'emploi inférieures à celles des personnes vivant dans des environnements plus homogènes. Il révèle en outre que cet effet est davantage lié à la dimension culturelle et non ethnique de la diversité. Enfin, le chapitre quatre développe un modèle rationalisant le recours des minorités ethniques à l'économie informelle en réponse à des conditions défavorables sur le marché du travail.



## Notice

The chapters of this dissertation are self-containing research articles. This explains that the terms "paper" or "article" are used, and why some information may be repeated. Chapter 1 is co-authored by Yann Algan and David Laitin. This explains the use of the pronoun "we". Access to sensitive data used in this chapter was granted through a convention between the French statistical institute (INSEE) and Sciences Po. Access to sensitive data used in chapters 2 and 3 was granted by the French Committee for Statistical Secret, through the use of a Secure Remote Center of Access to the Data (*Centre d'Accès Sécurisé Distant*, CASD).

## Avertissement

Les chapitres de cette thèse sont issus d'articles de recherche indépendants. Ceci explique la présence des termes "paper" ou "article", ainsi que l'éventuelle répétition de certaines informations. Le chapitre 1 est co-écrit avec Yann Algan et David Laitin. Ceci explique l'emploi du pronom "we". L'accès aux données sensibles utilisées dans ce chapitre a été accordé par le biais d'une convention entre l'INSEE et Sciences Po. L'accès aux données sensibles utilisées dans les chapitres 2 et 3 a été accordé par le Comité du Secret Statistique, moyennant l'utilisation des données via le Centre d'Accès Sécurisé Distant (CASD).





# Contents

<b>Remerciements Acknowledgements</b>	<b>v</b>
<b>Abstract</b>	<b>xi</b>
<b>Résumé</b>	<b>xiii</b>
<b>Notice Avertissement</b>	<b>xv</b>
<b>Contents</b>	<b>xvii</b>
<b>Introduction</b>	<b>1</b>
<b>I Diversity and Living Conditions</b>	<b>15</b>
<b>1 The Social Effects of Ethnic Diversity at the Local Level: A Natural Experiment with Exogenous Residential Allocation</b>	<b>17</b>
1 Introduction . . . . .	17
2 Presentation of the data . . . . .	21
2.1 Data sets . . . . .	21
2.2 Fractionalization indexes . . . . .	23
2.3 Sample characteristics . . . . .	23
3 The exogeneity of diversity in the public housing sector . . . . .	24
3.1 An ethnically-blind allocation process built into law . . . . .	25
3.2 Tests on quasi-random distribution of ethnic groups shares across public housing blocks . . . . .	28
4 Results . . . . .	30
4.1 Specification . . . . .	30
4.2 The effect of fractionalization on the opinion about housing conditions . . . . .	32
4.3 The various effects of diversity . . . . .	32
5 Robustness checks on self-reported quality of public goods . . . . .	39

6	Further tests: Fractionalization and ethnic shares . . . . .	41
7	Additional tests on the exogeneity of residential allocation in the public housing sector . . . . .	42
7.1	Absence of self-sorting on ethnic backgrounds . . . . .	42
7.2	Tests on the refusal rate of public housing offers . . . . .	46
8	Conclusion . . . . .	49
	Tables . . . . .	50
1.A	Fractionalization index and French blocks . . . . .	71
1.B	Principal component analysis . . . . .	72
1.C	Alternative summary indices - Mean effect analysis . . . . .	75
<b>2</b>	<b>The Local Determinants of Victimization</b>	<b>81</b>
1	Introduction . . . . .	81
2	Data overview . . . . .	85
3	Preliminary results . . . . .	88
3.1	Empirical methodology . . . . .	88
3.2	Contextual versus individual determinants . . . . .	90
4	The issue of location selection . . . . .	92
5	A spatial approach . . . . .	95
6	Conclusion . . . . .	98
	Tables . . . . .	100
<b>II</b>	<b>Diversity and Employment</b>	<b>113</b>
<b>3</b>	<b>Diversity and Employment Prospects: Do Neighbors Matter?</b>	<b>115</b>
1	Introduction . . . . .	115
2	Diversity and employment prospects . . . . .	118
3	Data and descriptive statistics . . . . .	120
4	Analysis . . . . .	123
5	Results with endogenous diversity . . . . .	126
5.1	Local neighborhood diversity: a local approach . . . . .	126
5.2	Employment zone diversity: an instrumental variable approach . . . . .	128
6	Interpretation of the results . . . . .	130
7	Conclusion . . . . .	133
	Tables . . . . .	134
3.A	Construction of the predicted level of employment zone diversity . . . . .	145

---

<b>4</b>	<b>Ethnic Networks and the Informal Labor Market</b>	<b>147</b>
1	Introduction . . . . .	147
2	The model . . . . .	151
2.1	Social interactions and job information transmission . . . . .	152
2.2	Flows of dyads between states . . . . .	154
2.3	Incentives . . . . .	155
3	Steady-state analysis . . . . .	156
4	Equilibria . . . . .	159
4.1	Unemployed workers decisions . . . . .	159
4.2	Equilibria . . . . .	160
4.3	Intuitions and comments . . . . .	162
5	Further results and comparative statics . . . . .	164
6	Conclusion and discussion . . . . .	166
4.A	Characterization of the steady-states . . . . .	168
4.B	Equilibria . . . . .	171
4.C	The minority group turns informal . . . . .	176
4.D	Comparison of majority and minority employment rates . . . . .	177
4.E	Comparative statics . . . . .	178
	<b>Conclusion</b>	<b>179</b>
	<b>Bibliography</b>	<b>181</b>



# Introduction

*"One of the most important challenges facing modern societies, and at the same time one of our most significant opportunities, is the increase in ethnic and social heterogeneity in virtually all advanced countries." Putnam (2007)<sup>1</sup>*

Post-colonial migrations, followed by trade development and economic integration have led to growing international migration flows. Besides, immigrants to a given country come from an increasingly wide range of countries.<sup>2</sup> Modern countries are therefore facing increasingly mixed populations, made of individuals from various origins, sharing different cultures and speaking different languages. These changes spark debate within the public and the academic spheres alike. Recent value surveys reveal for instance that a significant proportion of public opinion is hostile to immigrants: they are often perceived as a threat to job security and wages, as a burden in terms of welfare spending, or as a factor of crime. This resentment towards immigration is reinforced by high profile examples of failed integration such as urban riots in French *banlieues* in 2005, in London in 2011, and in Stockholm's suburbs a few months ago. This view contrasts sharply with the idea of diversity as enhancing people's welfare, which is promoted by governments and international organizations. An eloquent illustration is the 1<sup>st</sup> article of the 2001 Universal Declaration on Cultural Diversity of the UNESCO, according to which "cultural diversity is as necessary for humankind as biodiversity is for nature". No consensus seems to emerge among scholars either. In an article entitled *The clash of civilizations*, Huntington (1993) argued that "the great divisions among humankind and the dominating source of conflict will be cultural", and not ideological or economic as in the past. Yet, there is some evidence of a culture club rather than this predicted culture clash, to paraphrase Manning and Roy (2010). These authors indeed show that the longer immigrants stay in the UK, the more likely they are to think of themselves as British, this assimilation process being faster for immigrants coming from the countries the most dissimilar to the UK.

---

<sup>1</sup>This sentence was the opening of the lecture Putnam gave when he received the Johan Skytte Prize, rewarding Political Scientists.

<sup>2</sup>The example of the Chinese diaspora in France illustrates these two facts: there was virtually no Chinese immigrant in France at the beginning of the 20<sup>th</sup> century (less than 300 in 1912, contrasting with the 5,000 immigrants from Algeria at the same date), but they were more than 85,000 in 2009.

Besides, modern economies are facing growing social and economic heterogeneity, due to various historical events and economic changes. The Industrial Revolution first amplified spatial economic inequalities, not only between but also within countries, as emphasized by [Combes et al. \(2008\)](#) in the first chapter of their book. The 1980's were later characterized by an expansion of wage inequality and educational wage differentials, mostly driven by skill-biased technological change (*e.g.* [Bound and Johnson, 1992](#)). In the meantime, urbanization and social mix policies, embodied in the spread of social housing in the 1960's and 1970's led families from different social backgrounds to live in common urban spaces (*e.g.* Section 8 program in the US, *Grands Ensembles* in France). Yet, this spatial proximity did not reduce social distance, as analysed by French sociologists ([Chamboredon and Lemaire, 1970](#)). In addition, in the French case at least, social housing initially favored social mobility for middle class households who gained access to home-ownership, but soon became a poverty trap for low-income - often immigrant - families. However, recent US studies suggest that social diversity can have positive spillovers, and in particular in terms of mental and physical health (see [Ludwig et al., 2013](#)).

The previous discussion makes it easier to define diversity as it is understood in this dissertation. Diversity reflects the extent to which members of a given community, whether a country or a village, share different characteristics, usually pertaining to ethnicity, socio-economic status or even culture. Two points are worth stressing from what precedes. First, modern societies are facing increasing levels of diversity, which are not only ethnic and cultural, but also economic and social. Second, diversity can be an opportunity, but it can also bring about some difficulties. Assessing the impact of diversity on countries, firms and individuals is therefore not trivial, and deserves careful consideration. The purpose of this dissertation is to bring some answers to this broad question. More precisely, this thesis answers to the following question. Do individuals living in more diverse neighborhoods fare better or worse than those living in less diverse neighborhoods? This work deals mainly with diversity of origins, showing how it affects individuals' living conditions and employment prospects. To a lesser extent, it also studies the role of spatial disparities in unemployment. In particular, this factor turns out to be an important predictor of crime victimization, by opposition to origin diversity. Before getting to the heart of the matter, the remainder of this introduction will introduce central concepts and provide relevant contextual elements. I will first explain how diversity can impact economic and social outcomes, based on existing theories and evidence in the various social sciences. I will then present and discuss the results of the existing economic literature regarding the effects of diversity. I will finally provide a brief overview of what could be viewed as solutions to counteract the potential negative effects of diversity.

## How can diversity impact socio-economic outcomes?

The purpose of this section is to explain how diversity can affect social or economic outcomes, synthesizing the main theoretical mechanisms proposed by sociologists, political scientists, social psychologists and economists.

Admittedly, members of different groups do not necessarily share common tastes or the same goals, and as such have diverging preferences. A first channel through which diversity may be influential is precisely that of *preferences*. It can be particularly relevant when looking at public goods provision, as different groups may not agree on which type of public good should be provided. For instance, groups speaking different languages may disagree on the language that should be used in public schools. Similarly, if each group lives in a different place of a given jurisdiction, dispute about the location of public investments (e.g. a bridge) may arise. In the same vein, rich and poor households may wish for opposite types of public goods, for instance highway versus public transportation. Diversity may consequently lead to an under-provision of public goods, as suggested in [Alesina et al. \(1999\)](#). On a different matter, [Page \(2007\)](#) argues that divergent preferences may also erode trust because they are a "potential for disagreement [that] may create incentives to misrepresent how we feel. We may try to manipulate process and agenda, creating distrust and dislike".

A different yet related channel through which diversity may affect economic outcomes is the *other-regarding preferences* mechanism. The idea in this case is that individuals will derive a greater utility from the welfare of their co-ethnics than from the well-being of out-group individuals ([Tajfel et al., 1971](#)). As a consequence, individuals may be less inclined to contribute to a public good if they know it will benefit members from another group. This will, in turn, lead to an under-provision of public goods in more diverse communities (again, see [Alesina et al., 1999](#)). This is also the underlying idea of the [Tiebout \(1956\)](#) model, where rich individuals do not want to participate to redistributive policies and relocate accordingly. Another possible consequence of this bias against out-group members is that individuals will participate less to social activities because they prefer interacting with people who are similar to themselves. This is formalized in [Alesina and La Ferrara \(2000\)](#). Note that this paper also suggests that this channel can be at work in the context of ethnic diversity as well as in the case of wealth inequality.

Alternatively, diversity can have important effects through *social control*. Heterogeneous communities are more likely to form open social structure, that is with few connections between groups. As [Coleman \(1988\)](#) argues, this lack of closure prevents action that imposes external effects on others, thus hindering the emergence of effective social norms. In particular, such network structure may be a barrier to the implementation of sanctions to monitor and guide behavior. To put it differently, the threat of social sanction



is less credible across groups, so that the more diverse the community, the weaker the collective ability to discourage free-riders (Miguel and Gugerty, 2005). This idea is also behind the social disorganization theory of crime developed by Shaw and McKay (1942). They defend that identified poverty and ethnic heterogeneity undermine the ability and willingness of communities to exercise informal control over their members, hence facilitating criminal behavior.

Because diversity may imply sharing different cultural norms or speaking different languages, it can obviously lead to *communication* issues. This channel can also help predicting the impact of diversity on various outcomes. For instance, Lazear (1999b) states that multinational firms hiring workers speaking different languages may incur important costs (e.g. translation, coordination). Given the importance of social contacts in the process of finding jobs (see Ioannides and Datcher Loury, 2004, for a review), diversity may also reduce employment prospects by preventing job information transmission. On another ground, poor communication could make it more difficult to coordinate and to undertake collective action that would improve living condition. Finally, reduced information can prevent social control. As pointed out by Fearon and Laitin (1996), "it is more difficult to get information on a potential trading or social partner from "across the tracks." And if individuals are hard to identify or investigate across ethnic groups, then cooperation and trust across groups cannot be supported by punishment strategies that condition on individual behavior."

However, diversity can also lead to positive outcomes, due to *complementarity*. Workers from different origins are indeed more likely to have been exposed to diverse cultures and distinct school systems, acquiring various skills and learning different approaches to the same problem. In this case, diversity can increase productivity and facilitate innovation. More formally, Hong and Page (2001) develop a model showing that team work may benefit more from low-skilled but cognitively diverse workers than from homogeneous high-skilled workers. In a different theoretical setting, Lazear (1999b) shows that when multicultural workers are complementary, in the sense that they can exchange non-redundant and relevant information, the benefits from diversity offset its costs (e.g. barriers to communication).

In the same vein, social and economic heterogeneity can exert a positive impact through *peer effects*. High-income families can for instance act as positive role models for low-income families: as argued by Wilson (1987), blacks who live in poor neighborhoods are not exposed to "mainstream" role models, which hampers their economic mobility. Sociologists have for instance documented the positive relationship between children educational attainments and neighborhood characteristics such as average income or share of high skilled workers (Crane, 1991). More recently the economic literature has provided evidence of the existence of peer effects (Evans et al., 1992) and neighborhood

effects (Ioannides and Topa, 2010). Therefore, socio-economic heterogeneity could improve the outcomes of low-income or low-skilled families by putting them in contact with more successful individuals.

## Assessing the impact of diversity: findings and challenges

The previous section described the possible mechanisms through which ethnic diversity or social heterogeneity may influence socio-economic outcomes. This section now focuses on the literature devoted to assess the impact of diversity. It provides an overview of the results, and emphasizes that establishing causal inference can be challenging.

The question of the impact of diversity has gained increasing attention in the economic literature over the past two decades. Initially focusing on growth from a macroeconomic perspective, economists have also turned toward more microeconomic approaches, aiming their attention at firms' productivity, communities' public good provisions and individuals' pro-social behavior for instance. Most of the economic literature on diversity deals with ethnic or ethno-linguistic diversity, which is usually measured using either a *polarization index* or a *fractionalization index*. *Polarization* measures the extent to which the population is divided into two large distinct homogeneous groups, e.g. when a large ethnic minority faces an ethnic majority.<sup>3</sup> This measure was initially developed by Esteban and Ray (1994) in order to characterize wealth distribution, and was adapted to the case of ethnic groups by Reynal-Querol (2002). Polarization has proved particularly relevant in explaining ethnic conflicts and civil wars (e.g. Montalvo and Reynal-Querol, 2005), which were practically uncorrelated with fractionalization (see Fearon and Laitin, 2003). The bulk of the literature on ethnic diversity - and this dissertation is no exception - rather rely on *fractionalization*, which indicates the probability that two randomly drawn individuals from the population belong to two different groups.

In the first paper on the topic, Easterly and Levine (1997) seek to explain Africa's cross-countries differences in growth rates by cross-countries differences in ethnic fragmentation. This seminal paper brings empirical evidence that high ethno-linguistic diversity is at least partially responsible for high black market premiums, poor financial development, low provision of infrastructure and low levels of education, which are key determinants of economic growth. The conclusion of this macroeconomics-oriented study is that Africa's high level of ethnic diversity helps understand its "tragic growth performance". In the same vein, Alesina et al. (2003) construct measures of ethnic, linguistic and religious heterogeneity for a very large set of countries. They find that fractionalization is on average more correlated to the politico-economic outcomes they consider than

---

<sup>3</sup>More formally, polarization captures how far the distribution of the ethnic groups is from the (1/2, 0, 0, ... 0, 1/2) distribution (bipolar), which represents the highest level of polarization.

polarization is. In addition, they confirm previous findings that ethnic and language diversity is negatively associated with measures of good governance, while the reverse holds for religious diversity.

A few recent papers are revisiting this trend of the literature, disregarding ethno-linguistic diversity in favor of different yet related forms of diversity. Using the fact that the genetic diversity within a country is inversely related to the migratory distance from the cradle of humankind (East Africa) [Ashraf and Galor \(2013\)](#) show that "the level of genetic diversity within a society has a hump-shaped effect on development outcomes in the precolonial as well as in the modern era, reflecting the trade-off between the beneficial and the detrimental effects of diversity on productivity." In another cross-country analysis, [Alesina et al. \(2013\)](#) measure diversity in terms of birth countries. Their findings uncover a positive effect of such diversity on countries' income per capita, contrasting with the results obtained with ethno-linguistic diversity.

Another trend of the literature departs from the cross-countries studies and focuses instead on within countries issues, adopting a more microeconomic perspective. Working on US cities, metropolitan areas and urban counties, [Alesina et al. \(1999\)](#) report that ethnic fragmentation is associated with a lower budget share spent on public goods such as schools, roads, and trash pickup. Both their empirical results and theoretical framework suggest that different ethnic groups have different preferences over the type of public good to be produced and have a disutility if the public good is used by members of another group, so that more heterogeneous communities contribute less to local taxes. [Miguel and Gugerty \(2005\)](#) also show that the provision of public goods is less efficient in more diverse Kenyan villages, but they put a different explanation forward. They explain that the threat of social sanctions is less credible across ethnic groups and hence it is more difficult to avoid the free-rider problem in more diverse communities.

A few papers alternatively focus on outcomes pertaining to social capital. Interestingly, these papers do not only look at ethnic or racial fragmentation, but also insist on income heterogeneity. Using data on US Metropolitan Statistical Areas, [Alesina and La Ferrara \(2000\)](#) find that income and ethnic heterogeneity are associated with less participation to social activities. The result is stronger for ethnic diversity, and is especially salient for activities that involve a high degree of interaction between members. Similarly, [Costa and Kahn \(2003\)](#) reveal that engagement in civic life is hindered in more unequal and ethnically diverse communities, while [Alesina and La Ferrara \(2002\)](#) bring empirical evidence that trust is lower in more racially heterogeneous communities and in those with higher income inequality.

Finally, a segment of this trend of the literature studies the effect of diversity on productivity. As we have seen above, diversity can be detrimental by imposing costs related to barriers to communication and reduced trust. On the other hand, diversity can be bene-

ficial by bringing together a variety in abilities and skills complementary. A few theoretical papers have shown that under certain conditions, the positive effects of diversity offset its drawbacks (see for instance Lazear, 1999b). The related empirical literature tends to disregard purely ethnic diversity and rather turns towards different yet related forms of diversity. For instance, using data on US metropolitan areas, Ottaviano and Peri (2006) find that birthplace diversity, which is understood as cultural diversity, has a net positive impact on US-born workers' productivity.

The findings of this large literature could be summarized as follows: economic performance (taken in a broad sense), is negatively related to diversity in terms of income, race and ethnicity but positively related to cultural diversity. However, dealing with such questions presents some challenges, and in particular raises the issue of endogeneity. In the case of ethnic diversity for instance, an important concern is that individuals may have a preference for living close to their co-ethnics and thus tend to gather along ethnic lines. Then, the risk is that individuals who are not constrained with respect to the location of their home, e.g the wealthiest, may actually be able to self-segregate, so that the most homogeneous areas systematically correspond to wealthy places, while diverse areas end up being the most deprived. Hence, any estimates on the social or economic implications of diversity will be biased. A similar problem may arise if, on the other hand, diverse areas attract unprejudiced or more trusting individuals. In this case, any estimated effect of diversity on outcomes related to social capital (for instance) might be biased. Another issue could arise if immigrants decide to settle in more economically dynamic areas on purpose, leading to reverse causality.

In order to estimate unbiased causal effects, it is therefore necessary to ensure that diversity is not driven by some factor that also affects the outcome considered, or else to find a way to disentangle the two effects. Only recently has the literature started to tackle this issue. A widely used approach, instrumental variable estimation, consists in finding variables explaining diversity but being unrelated to the outcome considered. Most of the papers on ethnic diversity seeking to establish causality rely for instance on past settlement of immigrants or analogous historical data as instruments (see for instance Miguel and Gugerty, 2005; Ottaviano and Peri, 2006; Glennerster et al., 2013). In the same vein, the results of Alesina et al. (2013) are obtained specifying a gravity model to predict the diversity of immigration based on exogenous bilateral variables. Alternatively, part of the literature relies on natural or randomized experiments, whereby diversity is necessarily exogenous. A famous example of such experiment is the Moving To Opportunity program that randomly allocated housing vouchers to deprived US households so that they could relocate into richer neighborhoods. Several papers take advantage of this setting to examine the effect of social diversity. For instance Ludwig et al. (2013) establish that

living in better neighborhood improved physical and mental health of initially unfavored households.

## Some conditions for a successful diversity

The results presented in the previous section draw a rather pessimistic picture of the effect of diversity, although some positive effects were highlighted. This section makes amends for the previous one by giving some hints about conditions for a positive diversity effect.

First of all, [Fearon and Laitin \(1996\)](#) show that in ethnically diverse communities, conflict is the exception rather than the norm. Using a social matching model, they uncover two channels sustaining inter-ethnic cooperation even if social sanctions are not credible across groups. The first, called *spiral* equilibrium is the mechanism by which "individual defections trigger an escalation and complete breakdown of intergroup relations", when each group may hold all members of the other group liable for the actions of its individual members. In this case, cooperation is sustained by the fear of losing all future payoffs from cross-group transaction as a result of this breakdown. The *in-group policing* equilibrium comes from the fact that each group has better information about the behavior of its own members than about the other group and so can target individuals rather than whole groups. In this case, defectors are identified and punished by their own group, hence containing inter-ethnic violence.

From a broader point of view, [Putnam \(2007\)](#) argues that the negative spillovers of ethnic diversity are only a short run matter. He states that in the long run, individuals from different background eventually get along together so that only the positive effect of diversity may persist. This process is closely related to the integration of immigrants, is achieved through the creation of "new, cross-cutting forms of social solidarity and more encompassing identities". Following this idea, it seems that cultural integration or assimilation of immigrants is a key condition to overcome social fragmentation. In a book devoted to this topic, [Algan et al. \(2012\)](#) review empirical evidence of integration of immigrants in European countries. An important pattern common to the various countries considered emerges: the knowledge of the host country language. Therefore, after some time, barriers to communication may be tremendously reduced. It is worth noting that such integration process also facilitates the adoption of common social norms, through secularization or education for instance.

Finally, several papers have highlighted possible moderating factors. For instance, [Miguel \(2004\)](#) finds no diversity impacts on local outcomes in Tanzania, a country in which the ruling authorities have sought to ameliorate ethnic cleavages by promoting a common language. This comes back to the previous idea that learning a common language facilitates communication and cooperation. Another example is provided by [Posner](#)

(2004), who show that changed electoral rules can create broader ethnic identities thereby reducing fragmentation. Finally, in a recent study, [Glennster et al. \(2013\)](#) argue that the presence of strong chiefs at the local level, although reinforcing the salience of ethnicity, translates into effective inter-ethnic cooperation.

## Contributions and road map of the dissertation

By assessing the social and economic effects of local diversity this dissertation follows the literature presented above. Yet, it enlarges the scope of this literature, complementing existing studies in several dimensions: it looks at unexplored outcomes, focuses on different geographical scopes, digs further into the mechanisms at play, relies on alternative measures of diversity, and proposes an innovative identification strategy. More specifically, the primary question of this thesis is the following: do individuals in more diverse neighborhoods fare better or worse than individuals in less diverse neighborhoods? Answering it leads me to deal with two secondary but not less important matters: understanding the mechanisms driving the relationship between diversity and the various outcomes considered, and addressing the endogeneity issue.

This thesis is divided into two main parts, dealing respectively with living conditions (chapters 1 and 2) and employment (chapters 3 and 4). The first chapter studies the impact of block level diversity of origins on housing conditions, and develops a new identification strategy to bypass the problem of endogenous diversity. It is directly related to the second chapter, which studies the local determinants of victimization. In particular, this chapter explores the impact of spatial heterogeneity in terms of unemployment. Chapter three studies the effect of local diversity on individuals' employment prospects, challenging alternative measures of diversity, both in terms of origins definition and of geographical scope. Finally, chapter four develops a model emphasizing the role of job information transmission in the choice of engaging into undeclared work, in a context where two ethnic groups co-exist. The content of each chapter is described below.

The first chapter, entitled *The Social Effects of Ethnic Diversity at the Local Level: A Natural Experiment with Exogenous Residential Allocation*, demonstrates the effects of diversity on housing quality at a very finite neighborhood level.<sup>4</sup> Its main contribution is to identify the social effects of diversity: working at a very low geographic level enables us to see how diversity shapes neighborhoods relationships, which in turn helps understanding the channels through which diversity impacts housing conditions. An additional contribution of this chapter is to provide a new identification strategy to overcome the endogeneity problem raised by residential sorting. We rely on a natural experiment of ex-

---

<sup>4</sup>This chapter is co-authored, hence the use of "we".

ogenous spatial allocation in the French public housing sector to identify the causal effect of diversity on those outcomes.

This study relies on the French housing survey, which reports specific information about neglect and degradation in the common areas of the apartment buildings, general housing quality and interpersonal conflicts. It also uses data from the 1999 population census, through which we measure diversity at the block level. These rich data enable us to identify various effects of local diversity on living conditions, and to explore the possible channels explaining these relationships. In order to make unbiased causal inferences, we provide a new strategy for identifying the causal effect of diversity on economic and social outcomes. It relies on a natural experiment in which households in France are allocated to public housing blocks without taking their ethnic origin or their preference for diversity into account. This implies that the neighborhood level of diversity faced by individuals living in the public housing sector can be considered as exogenous. We show that this identification assumption is in line with French law and with the official public housing allocation process. We also conduct a variety of formal statistical tests ruling out the possibility of self-sorting along ethnic lines in the French public housing sector. We therefore restrict our analysis to the public housing population to derive unbiased estimates of the effect of diversity.

We find that an increase in block diversity implies more voluntary degradation and vandalism in the common areas of the housing units (broken mailboxes, graffiti...), and a poorer quality of basic facilities and equipment that should be taken care of by the public housing directorate (heating system, elevator...). On the other hand, we do not find any significant impact of diversity on aggression, burglaries or robberies in the neighborhood. Our interpretation of these results is that diversity generates social anomie, i.e. the absence of common rules and social norms. As a consequence, there is a failure to impose social sanctions and punish defectors, hence more vandalism; a failure to generate collective action to pressure the public housing offices into improving housing quality; and fewer opportunities for violent confrontation at all levels of diversity. Although our data do not allow us to test directly this interpretation, we provide some support to this claim.

The title of the second chapter is self-explanatory: *The Local Determinants of Victimization*. This chapter relates to the previous one on two aspects. First, it deals with another dimension of living conditions, namely the probability of being victim of a criminal event. Although this question was considered in the first chapter, this one actually focuses on that outcome and makes use of more appropriate data. In addition, preceding results suggested that local diversity had no effect on crime related outcomes, which were more related to local unemployment. The second chapter hence emphasizes the role of unemployment rate, and in particular the spatial heterogeneity in unemployment rates, in

explaining victimization. Indeed, the starting point of this study is the thought that results obtained in the economics of crime literature at very aggregate levels may hide important spatial disparities. In particular, one finding of the existing literature is that unemployment increases crime rates. Yet, at the neighborhood level, it is not clear whether one is more exposed to risk when living in a high-unemployment neighborhood or when residing in a privileged area adjacent to a deprived one, specially if the criminals are mobile. Studying crime from a more microeconomic perspective can thus challenge some of the established results, and lead to a better understanding of the mechanisms behind criminal events.

Using the French victimization survey that provides information localized at a very low geographic level (a 2,000 inhabitants neighborhood), this chapter proves this idea well-founded, as will be explained below. As was the case in the previous chapter, this work is naturally confronted to the issue of residential sorting, that may bias any estimates of neighborhood characteristics, and in particular unemployment rate. To tackle this issue, I take advantage of the very precise localization of the data, and adopt the strategy developed by Bayer et al. (2008). The identifying assumption is that although households are able to select a given area in which they want to live, they are, however, unable to pick a precise neighborhood within this given area. This assumption means that even if households are able to choose a given residential area, there will not be any correlation in unobserved factors affecting risk of victimization among individuals living in the same neighborhood *within the larger selected area*. As a consequence, once we control for the characteristics of the larger area selected by the individual, the remaining spatial variance of unemployment across neighborhoods within the larger area is supposed to be exogenous. Although I do not formally test this assumption, I provide some arguments supporting the credibility of this assumption in the context of this study.

Three important findings emerge from this study. First, neighborhood characteristics explain victimization better than individual characteristics, except for assaults. Second, among the various neighborhood characteristics considered, unemployment rate appears as the most relevant factor having a positive effect on victimization, in line with the findings of the previous chapter. Third, adopting a spatial approach reveals that for crimes such as burglaries and thefts of objects from cars, the effect of unemployment rate in surrounding neighborhoods is stronger than the effect in the neighborhood where the crime took place, while the reverse is true for smaller crimes. This result gives some support to the idea that criminals are mobile across neighborhoods for more serious economic crimes, in line with the Beckerian theory of crime, but that petty crimes and vandalism do not involve any mobility, relating rather to the social disorganization theory.

In the third chapter, *Diversity and Employment Prospects: Do Neighbors Matter?*, I



assess the effect of local diversity of origins on individuals' employment prospects. As suggested in the previous chapters, diversity can have various effects at different geographic levels, and the case of unemployment is no exception. At a very local level, diversity may act as a barrier to job information transmission between neighbors hence lowering their chance to find a job. At a more aggregate level, on the other hand, it may enhance productivity through skills complementarity and increase employment opportunities. This brief discussion reveals that the relationship between diversity is not trivial. This chapter aims at understanding as precisely as possible the mechanisms lying behind the diversity-employment relationship, which remained unexplored so far.

I am able to deal with this question using the French employment survey. These data allow me to distinguish between a local and a more global effect of diversity by measuring diversity at two different geographic levels: the local neighborhood, composed of about 25 adjacent homes, and the employment zone, which corresponds to a local labor market. Naive regressions of individual employment status on both measures of diversity reveal that employment probability is negatively correlated with neighborhood diversity, but positively correlated with employment zone diversity. This suggests a negative effect through networks and a positive one through productivity. In addition, I rely on three alternative definitions of origins to measure diversity: nationality, birth country, and parents' origins. The results presented above are consistent across the three measures. Yet, diversity based on nationality has a larger impact than diversity based on birth country, which is itself more relevant than diversity based on parents' origins. I argue that this suggests a prominent role of cultural over ethnic diversity.

Once again, these results may be biased due to the likely endogeneity of diversity. I confront this problem in two different ways, according to the geographic level. To deal with the endogeneity of local neighborhoods diversity, I follow the same strategy as in the previous chapter. That is, I assume that households are unable to pinpoint the exact neighborhood in which they end up living within a larger selected area. Expectedly, as we control for the characteristics of this larger area, the estimated effect of neighborhood diversity becomes more negative. The same strategy cannot be applied to employment zones diversity, because these areas are too large for the identifying assumption to hold. Instead, the problem is handled through a more traditional instrumental variable approach, where two different instruments are proposed. The first one is the predicted level of diversity in each employment zone, based on the past distribution of each origin group across employment zones and on the current number of individuals from each origin in France, and inspired from [Card \(2001\)](#). The second one is the level of diversity within the public housing tenants of the employment zone, building on the methodology developed in the first chapter. Interestingly, once employment zone diversity is instrumented using any of these two variables, its positive relationship with employment is driven down to zero.

The last chapter, entitled *Ethnic Networks and the Informal Labor Market*, differs from the previous ones in that it is theoretical. Yet, because it deals with social relationships in a context where a majority group co-exist with a minority group, this chapter echoes the others. It is particularly related to the previous chapter, as it emphasizes the role of job information transmission within and across groups, on the decision to enter the formal or the informal labor market. The motivation for this model comes from several facts. First, ethnic minorities (e.g. immigrants) usually face adverse labor market situation (e.g. low wages and high unemployment rates). Second, the informal economy serves as an economic safety net for the most deprived individuals. Combined, these facts give the intuition for why minorities would engage into the informal sector. Third, information networks are an important channel for finding standard jobs, and *a fortiori* undeclared jobs. Therefore, if information does not circulate well across groups, individuals from the minority group are less likely to receive job offers through word of mouth communication. If they consequently face high unemployment rates, they might instead turn toward the informal economy, especially if they receive informal offers more easily through their minority contacts.

In this chapter, I provide a theoretical background for the fact that ethnic minorities circumvent adverse labor market outcomes by undertaking undeclared jobs. I develop a dynamic model in which agents belonging to two different groups (majority versus minority) form a network through which they can exchange information about declared or undeclared jobs. In the model, the network plays a central role: both formal and informal job opportunities can only be obtained through word of mouth communication. This network features homophily, so that information circulates imperfectly across groups. I analyse the flows of individuals between the various occupations and characterize all the steady-states of this dynamic economy. For this purpose, I solve for the endogenous individual decisions to accept or reject formal and informal job offers. I then derive the conditions under which these steady-states arise in equilibrium. Interestingly, when the net formal wage is larger than the net informal wage, all equilibria involving participation in the informal economy arise for low enough values of the unemployment benefit. This reveals that when the unemployment benefit is too low, individuals are more likely to immediately accept an informal job offer when they receive one rather than to decline and wait for a formal job opportunity.

In particular, there exists an equilibrium in which workers from both groups are employed in the formal sector but the minority workers also engage in the informal sector. This equilibrium is in line with the intuition that the informal sector may be used as an economic safety net when formal employment prospects are low. Indeed, the minority group faces such a low rate of formal job offers arrival (due to the size of the network)

that its members are better off immediately accepting an informal job offer than waiting longer for a formal one. In this setting, simple comparative statics allows me to consider and compare policies aiming at reducing informal employment. I show that improving the communication across groups implies an increase in the formal employment rate for both groups, together with a reduction of the informal employment rate. Similarly, intensifying the audit rates to destruct informal jobs has a positive impact on the formal employment rate for both groups and a negative effect on the informal employment rate, but its comes at the expense of increased unemployment rates.

## **Part I**

# **Diversity and Living Conditions**



# Chapter 1

## The Social Effects of Ethnic Diversity at the Local Level: A Natural Experiment with Exogenous Residential Allocation

### 1 Introduction

Recent research has drawn an ominous picture of the implications of cultural heterogeneity on social peace and economic growth. A large literature shows a negative relationship, though not always robust, between ethnic diversity and the quality of public goods (Alesina et al., 1999; Alesina and La Ferrara, 2000; Miguel, 2004; Miguel and Gugerty, 2005), welfare spending (Luttmer, 2001), civil conflict and trust (Fearon and Laitin, 2000; Putnam, 2007; Alesina and La Ferrara, 2002) and economic growth (Alesina et al., 1999). The leading explanations of why ethnic fragmentation affects those outcomes are the heterogeneity of preferences and the free-rider problem which undermines collective action. The literature thus views the problem of fractionalization in terms of voting behavior on aggregate outcomes such as public good provision at the country or county level. Yet, little is known on how diversity could affect directly social relationships and well-being at the neighborhood level. Our paper fills this gap by looking at the effect of ethnic diversity on social relations and the quality of common spaces within local communities at the housing block level. Besides we provide a new identification strategy to overcome the endogeneity problem raised by residential self-selection. We rely on a natural experiment of exogenous spatial allocation in the French public housing sector to identify the causal effect of diversity on those outcomes.

The main contribution of our paper is to identify the effect of ethnic diversity on social relationships and the quality of public goods at a very local block level. We use micro data on housing conditions where the units of observation are public housing blocks made

up of twenty adjacent households on average. This is a key improvement for the analysis of how diversity shapes social relationships compared to the previous literature which is based on aggregated data at the county, regional or country levels. Diversity might matter for various reasons at different levels and the channels through which diversity operates are likely to depend on the size of the unit of observation. By focusing on the provision of public goods at an aggregate level, the previous literature is mainly interested in the effect of diversity on collective action through lobbying or patronage (see [Alesina and La Ferrara \(2005\)](#) for a survey). Instead, we analyze in this paper how diversity within a small community affects individual well-being and satisfaction with housing conditions through relationships among neighbors. Indeed, the survey we use (French Housing Survey) reports specific information about the neglect and voluntary degradations of the public areas, the quality of the housing, and direct interpersonal conflicts. These data enable us to identify various effects of diversity on local social relationships and public good outcomes, and to explore the possible channels explaining this link.

When residents of more diverse blocks report that neglect and voluntary degradation is rife in their housing unit, we interpret this as a result of the residents' failure to develop social norms and other regarding preferences. When they report the breakdown and the poor quality of basic facilities (such as heating and soundproofing), we interpret this as a result of a diminished capacity for collective action for social improvement. Those goods are of course not directly degraded by diversity. But diversity might be associated with lower ability for collective action, explaining the irregularity of maintenance and the absence of repairs in more diverse blocks. In this case, the result could be supported in equilibrium if the housing directorate reckons that it can neglect facilities in ethnically heterogeneous housing projects, knowing that it will not face collective action from its residents demanding better services. Finally, when residents report incidents of direct interpersonal conflicts, we can interpret this as an effect of diversity on cultural enmity. We test these channels by using indirect objective measures of the quality of social relationships and common spaces, such as the number of repairs and the upkeep of the security equipment. We find that diversity decreases the quality of local common spaces, but has no effect on public safety. Instead, individuals are more likely to report the absence of any social relationship with their neighbors rather than interpersonal conflicts. Thus diversity leads to social anomie, preventing the emergence of social norms and collective action, rather than antipathy at the local neighborhood level.

In order to make unbiased causal inferences, we provide a new strategy for identifying the causal effect of diversity on economic and social outcomes. The general concern in this literature is that the endogenous residential sorting of individuals on ethnic grounds biases the estimate of the impact of diversity. We address this issue by using a natural

experiment in which households in France are allocated to public housing blocks without taking their ethnic origin or their preference for diversity into account. Due to a strongly republican ideology, the French public housing system allocates state planned moderate cost rental apartments (HLMs - *Habitations à Loyer Modéré*) to natives and immigrants without concern for their cultural and ethnic background, mixing people indiscriminately. Some HLM neighborhoods are consequently quite diverse, and others quite homogeneous. Furthermore, HLM inhabitants rarely move, as the rents are much lower than market rates, and moving between HLM blocks is quite difficult. Consequently, residents cannot choose whether to live near people like themselves. Rather, they accept their placement, whether next to co-ethnics or strangers. Methodologically, this means that we can take the degree of diversity in any one HLM block as exogenous, connect the level of diversity with the housing situation, and examine whether greater heterogeneity leads to poorer provision of public goods or more troubled social relationships in French communities. We extensively document the actual process of allocation of households within the public housing sector. We show that legal rules prohibit housing allocation based on ethnic backgrounds and that in practice, the characteristics of the public housing sector make it very complicated to bypass the law. Then, we conduct a variety of formal statistical tests to verify the absence of self-sorting on ethnic characteristics. In particular, we show that the observed spatial distribution of residents across public housing blocks is not statistically different from a random distribution generated by Monte Carlo simulations. We perform a variety of alternative tests. We show that while households moving into a new neighborhood tend to self-segregate in the unconstrained private housing market, there is no such evidence in the public housing market. We also examine potential self-selection prior to the move and show that households that have refused an offer end up living in public housing blocks that display the same ethnic diversity as those who accepted their first offer. Thus even if households try to be choosy with respect to the ethnic composition of their neighborhoods, they cannot self-segregate in the public housing sector due to the allocation process and the tight supply constraints of dwellings.

Naturally, this paper is not the first one to try to overcome this identification issue. But previous attempts to establish causality rely mainly on instrumental variables.<sup>1</sup> However convincing the instruments might be, this strategy cannot overcome the concern as to whether the instruments fulfill the exclusion restriction and do not have a direct effect on public goods. For instance, Miguel (2004) and Miguel and Gugerty (2005) use the pre-colonial patterns of settlement as instruments, assuming that these variables have no direct impact on present-day ethnic relations. More recently Glennerster et al. (2013) have also

---

<sup>1</sup>In their seminal contribution to the literature, Alesina et al. (1999) provide a first attempt to deal with this endogeneity issue by collecting data at different levels of aggregation (cities, metropolitan areas and counties). Their assumption is that different levels of aggregation allow for the correction of the potential biases introduced by Tiebout sorting.



relied on historical data of fractionalization as an instrument. But since past settlement patterns are likely to have at least some direct impact on present-day ethnic relations, the exclusion restriction might still be technically violated. Using a natural experiment with exogenous allocation of ethnic groups is thus an alternative strategy to deal with these traditional caveats. The paper which is the closest to ours is [Dahlberg et al. \(2012\)](#), which uses a nation-wide policy intervention program that exogenously placed refugees coming to Sweden across the Swedish municipalities. However, their paper examines in-group bias in preferences for redistribution rather than the effect of diversity on local public goods and social relationships.

Our paper is related to the large literature on the effects of ethnic diversity on economic and social outcomes. In US cities, higher ethnic diversity has been found to be associated with lower social capital ([Putnam, 2007](#); [Alesina and La Ferrara, 2000, 2002](#)), lower welfare spending ([Luttmer, 2001](#)), and poorer quality of public goods ([Alesina et al., 1999](#)). In Western Kenya, the greater the mixing of tribes, the less people have public spiritedness, and the lower the contributions to public goods ([Miguel, 2004](#); [Miguel and Gugerty, 2005](#)). In cross-national surveys, diversity correlates with low growth in GDP and low quality of institutions ([Easterly and Levine, 1997](#); [Alesina et al., 2003](#)). [Alesina and Zhuravskaya \(2011\)](#) show that islands of homogeneity amid a broadly diverse country do not decrease the negative effects of diversity on the quality of government. Theoretical contribution, in particular on ethnic conflicts, can be found in [Caselli and Coleman \(2013\)](#) and [Esteban and Ray \(2011\)](#).<sup>2</sup> These findings are depressing, in a normative sense, for those who herald gains from diversity ([Page, 2007](#)); and depressing, in an empirical sense, as in our globalized world, local cultural diversity is increasingly common ([Dancygier, 2010](#)). However, the robustness of the relationship and the channels at work remain to be determined. [Putnam \(2007\)](#) is careful to underline that his data allow him only to claim short run correlation between diversity and trust. [Miguel \(2004\)](#) finds no diversity impacts on local outcomes in Tanzania, a country in which the ruling authorities have sought to ameliorate ethnic cleavages by promoting a common language. [Posner \(2004\)](#) shows that changed electoral rules can create broader ethnic identities thereby reducing fragmentation. [Dunning and Harrison \(2010\)](#) show that inter-tribal polarization in Mali is reduced with cross-cutting cleavages. [Glennerster et al. \(2013\)](#) also argue that the presence of

---

<sup>2</sup>The magnitude of the relationship between those outcomes and ethnic diversity is substantial. [Putnam \(2007\)](#) finds that the difference between living in a highly homogeneous city (Bismarck, North Dakota) and the heterogeneous Los Angeles is as great as the difference between an area with a poverty rate of 7 percent and one with a poverty rate of 23 percent. [Alesina et al. \(1999\)](#) show that moving from complete homogeneity to complete heterogeneity is associated with a reduction in spending on roads by nine percentage points. [Luttmer \(2001\)](#) finds that interpersonal preferences based on negative exposure and racial group loyalty of recipients are associated with 33 percent of the cross-state variation in the support for welfare spending. [Alesina et al. \(2003\)](#) show that moving from perfect homogeneity to maximum heterogeneity would be associated with a reduction in a country's growth rate by two percentage points per year.

strong chiefs at the local level, although reinforcing the salience of ethnicity, translates into effective inter-ethnic cooperation. Finally, [Alesina and La Ferrara \(2005\)](#) document the potential positive effect of diversity on productivity through complementarity in skills.

Our paper is also incidentally related to empirical works examining neighborhood effects on social and economic outcomes. So far, the literature has mainly focused on the neighborhood effects on physical and mental health, economic self-sufficiency, risky and criminal behavior, or educational outcomes (see among many others [Katz et al., 2001](#); [Oreopoulos, 2003](#); [Goux and Maurin, 2007](#); [Kling et al., 2007](#)). In particular, [Katz et al. \(2001\)](#) and subsequent contributions use the Moving to Opportunity social experiment to estimate the externalities from neighbors. To avoid the problem of endogenous neighborhood selection, those authors use data from a randomized experiment in which some families living in high-poverty U.S. housing projects were offered housing vouchers to enable them to move to higher income areas. While our paper is not based on a randomized experiment, we also avoid the inferential issues of residential endogenous selection by using the exogenous spatial allocation of households with respect to ethnic characteristics. We enlarge the dimensions analyzed in this literature by looking at how immediate neighborhood diversity affects well-being and the quality of the local environment.

The remainder of this paper is organized as follows. Section 2 presents the data. Section 3 documents our identifying assumption on the absence of residential self-sorting in public housing. We conduct various tests to show that the spatial allocation of households across public housing blocks within localities is exogenous with respect to ethnic characteristics. Section 4 shows our main results. We document the effects of ethnic diversity on satisfaction with housing condition, local public goods quality and social relationships. We discuss the various dimensions and channels through which diversity might matter for households' well-being at the finite local level. Section 5 provides tests of the validity of self-reported outcomes, and we perform a series of robustness checks on our results in section 6. Section 7 documents additional tests on the exogeneity of the residential allocation in the public housing sector. Section 8 concludes.

## 2 Presentation of the data

### 2.1 Data sets

The analysis is based on two representative French national surveys. We use the French Labor Force Survey (Enquête Emploi, INSEE, hereafter the LFS) to test our identification assumption that spatial allocation in the public housing market can be considered as quasi-random and exogenous relative to ethnic characteristics. The survey covers the period

2003-2007 and provides all the relevant information about ethnic background, economic characteristics and geographic location of individuals to test our identification strategy. The sampling unit of this database is at the housing block level, which consists in twenty adjacent households on average.<sup>3</sup> Over the 2003 to 2007 period, more than 10,000 different housing blocks were sampled. All the households within a randomly selected housing block were surveyed and, within each household, all persons aged fifteen or over were interviewed. Using these data, we can work on real neighborhoods at a very small geographic level. Moreover, we have information on whether the respondent was living in a public housing unit, whether he or she has been living in his or her current public housing for at least one year or whether he or she has just moved into the neighborhood. These particular features enable us to compute the level of ethnic diversity prevailing within each public housing block and to test for the absence of self-sorting on ethnic background across public housing blocks.

We use the French Housing Survey 2002 (*Enquête Logement*, INSEE, hereafter the HS), to estimate the relationship between ethnic diversity and the quality of public space within the housing block. We identify the causal effect of diversity and control for self-sorting by focusing on the public housing sector. The HS provides detailed information on the intensity and quality of social interactions within housing block, including social relationships and the quality of local public spaces, ranging from vandalism in the common areas, to housing quality and conflicts in the neighborhood. The HS also reports detailed information about the ethnic, economic and social backgrounds of individuals within the neighborhood.<sup>4</sup> In the HS, all the individuals living in a given housing block are not systematically surveyed and are randomly drawn instead. We thus cannot compute ethnic fractionalization on a representative sample at the housing block from the HS. We overcome this concern by using the 1999 French Population Census. Each HS sample is drawn from the most recent Census and the geographical units of the HS are a subsample of those of the Census. As the Census provides variables such as birth country or nationality at birth, it allows us to compute representative fractionalization indices at the housing block level and then to match them with the corresponding housing block in the HS. Table 1.18 in Appendix A presents descriptive statistics for the housing blocks in the Census and the Housing Survey.

---

<sup>3</sup>INSEE has chosen this sampling strategy so as to reduce the traveling expenses of those who administer the survey.

<sup>4</sup>Some of the key variables for our study are not public. The French Statistical Institute (INSEE) made their access possible as part of a convention between the INSEE and Sciences Po. We were required to make use of the "sensitive" data within the confines of the INSEE.

## 2.2 Fractionalization indexes

We use the standard ethno-linguistic fractionalization (hereafter ELF) index used in the literature to construct our measure of ethnic diversity (e.g. [Alesina et al. \(2003\)](#) for a detailed description).<sup>5</sup> This index reflects the probability that two randomly drawn individuals from a given population belong to different groups (previous studies looked at ethno-linguistic or religious groups). More formally, the basic fractionalization index is computed as one minus the Herfindahl index of ethno-linguistic group shares:

$$ELF_j = \sum_{i=1}^{i=N} s_{ij} (1 - s_{ij}) = 1 - \sum_{i=1}^{i=N} s_{ij}^2 \quad (1.1)$$

where  $s_{ij}$  is the share of group  $i$  ( $i=1, \dots, N$ ) in area  $j$ . If the population living in area  $j$  is fully homogeneous,  $ELF_j$  equals 0 and it converges to 1 as the population heterogeneity increases. Note that  $ELF_j$  can increase for two reasons: it will increase with the number of ethno-linguistic groups, and it will increase the more equal the size of the groups. As mentioned above, the Census data and the LFS provide information about the country of birth and the nationality at birth of individuals, allowing us to construct two different measures of diversity. In the remainder of the paper, we focus on diversity as measured by nationality at birth, computed at the block level. The distribution of diversity in housing blocks is presented in Appendix A (Figure 1.1 and Table 1.19).<sup>6</sup> Unsurprisingly, given that immigrants and second generation French are more likely to be eligible for public housing dwellings on income criteria than native French, the public housing neighborhoods are characterized by higher levels of diversity than other neighborhoods.

## 2.3 Sample characteristics

We now document the characteristics of the 2002 Housing Survey. Most of the variables are given at the household level. The dataset contains 32,156 households, corresponding to 78,791 individuals. 39.6 percent of the French households are renters while 56 percent are owners. Overall, 15.77 percent of the households live in public housing units, representing 39.8 percent of the tenants.

Table 1.1 shows the main socio-demographic characteristics of the sample. We compare households living in the public and in the private housing sectors. Foreigners (or immigrants) are over-represented in the public housing population compared to the private housing population. Public housing neighborhoods are also characterized by a poorer

<sup>5</sup>We have also tried alternative indexes such as polarization indicators, yielding similar results.

<sup>6</sup>On the public housing graph, we see that 6 percent of public housing blocks are perfectly homogeneous. This high frequency is to a large extent explained by the fact that in many blocks we observe only very few inhabitants, thereby increasing the chance of getting a null ELF. We keep those blocks in our main analysis, but we checked that deleting them does not affect the results.

socio-economic environment: the unemployment rate is around twice as high as in private housing blocks. Individuals living in public housing dwellings are less educated and earn lower incomes. Around one third of the adults have no diploma at all, and the share of individuals having achieved graduate studies is less than half the corresponding share in the private housing sector. Column 3 shows that the two populations are statistically significantly different with respect to most of their characteristics. Column 4 shows the characteristics in the private housing sector when we restrict to tenants. Still, the two populations are statistically significantly different (Column 5).

Table 1.2 documents the perception of housing conditions by native French and immigrants in the public housing sector. On average, natives have a much better opinion about the quality of their housing than Maghrebians or other Africans.<sup>7</sup> Table 1.2 shows that 13.8 percent of the native French are very satisfied with their housing conditions while this is the case for only 8.42 percent of the Maghrebians. Conversely, only 9.9 percent of the native French complain about insufficient housing conditions versus 18.21 percent of the Maghrebians. The last line of Table 1.2 reveals that the poorer housing conditions are associated with lower levels of income, a situation more salient for Maghrebian and African families. In particular, the households that are very satisfied with their housing conditions earn on average 13,300 euros per year, while very unsatisfied households earn 10,127 euros a year on average. From the last column, we observe that the average Maghrebian family earns even less than that (8,603 euros).

### 3 The exogeneity of diversity in the public housing sector

This section addresses the main identification issue raised by the estimation of the effect of ethnic diversity on social interactions and the quality of public goods. The issue, common to all the literature on ethnic diversity, is that fractionalization presents a high risk of endogeneity. Individuals generally tend to self segregate: they prefer forming links with others like themselves, with whom they share common interests, and in particular people of the same ethnicity or the same social background.<sup>8</sup> If people can choose the area where they live, they would rather move into neighborhoods where people are similar to themselves. If individuals who are not constrained with respect to the location of their home choose to gather along ethnic lines, then the richest individuals will be able to move into the most homogeneous neighborhoods. Therefore, the level of diversity of the neighborhoods is probably endogenous and any estimates on the implications of diversity will be biased. In particular, if the wealthy families that live in diverse settings are those

---

<sup>7</sup>We observe the same pattern when we look at the various subjective and objective measures of the quality of public housing

<sup>8</sup>Race, or ethnicity, is the most salient characteristic along which homophilious relationship form.

that have a taste for diversity, the true effect of diversity on social outcomes should be smaller in absolute terms.<sup>9</sup>

To identify the effect of ethnic diversity, one must therefore study individuals who are assigned to their place of residence without consideration of ethnic characteristics. The purpose of this section is to bring forth evidence that spatial allocation of households across public housing blocks in France can be considered as exogenous with respect to ethnic characteristics due to French regulation. Naturally, the sample of households that apply to public housing dwellings is endogenous with respect to economic, social or cultural characteristics. But among the pool of selected households, we show that their spatial allocation across the public housing blocks of a given department is exogenous with respect to their ethnic characteristics, conditional on their other characteristics.

We also want to stress that the mere fact of working at the block level already decreases the extent of endogenous sorting. First, although households can generally choose the neighborhood in which they move, they may not be able to select a particular block in a given neighborhood. Second, while it is possible to have an idea of the socio-economic and demographic characteristics of a given neighborhood, it is much more difficult to observe such characteristics in a specific block prior to moving. Bayer et al. (2008) rely on this key assumption and provide empirical evidence that individuals characteristics are not correlated within blocks.

### 3.1 An ethnically-blind allocation process built into law

We first document the actual process of allocation of households across public housing dwellings. This gives a legal basis to our identifying assumption of the absence of self-sorting on ethnic characteristics in the public housing sector.<sup>10</sup> In France, the only eligibility requirements for admittance into the public housing sector are to be legally living in France (as a French citizen or migrant with a valid residence permit) and to be living under a certain threshold of income per unit of consumption. This income ceiling is usually rather high: in 2009, this threshold was between 36,748 and 50,999 euros per year for a four-person family, depending on the region of residence (the upper figure is nearly 3,000 euros higher than the average disposable income of four-person households in 2007). Using the 2002 Housing Survey data, Jacquot (2007) estimates that given their income, between two thirds and four fifths of households living in Metropolitan France

---

<sup>9</sup>Combes et al. (2012) use customer discrimination theory to show that owners will tend to discriminate against ethnic minorities when renting their apartment, bringing new evidence of why any causal claim of ethnic diversity on public goods in the private housing market would be biased.

<sup>10</sup>The process of allocation across public housing blocks in France is mainly inspired by theories from Le Corbusier (1887-1965). Le Corbusier insisted that France must avoid the homogeneous ghettos of the urban landscapes elsewhere, and should therefore allocate housing blind to ethnicity, not permitting family networks to grow within housing establishments. These ideas were translated into state regulation (Bernardot, 2008).

could apply for a public housing unit. As a consequence, the population eligible for public housing is about three times as large as the available space in vacant dwellings. This implies that other criteria must be taken into account in the distribution process. Hence, in addition to the income of the household, family situation and household size are taken into account to ensure a suitable match with the characteristics of vacant dwellings, as well as the emergency of the application. These are actually the main criteria used by the commission due to the boom in housing prices in the private sector during the mid-90s and the 2000s. In particular, five priority criteria are defined by law (Article L441-1 of law relative to construction and housing - *Code pour la Construction et l'Habitat*) at the national level to ensure that vacant housing will first be distributed to households with obvious social difficulties. Households satisfying these priority criteria are those in which there is a (mentally or physically) disabled person, those living in precarious or hazardous shelter due to financial constraints, those living in a temporary accommodation, individuals living in a precarious shelter who recently found a job after a long unemployment spell, and spouse-abused individuals.

To get on the queue for a housing unit, households submit a form revealing their identity and family situation, their employment status and the resources of the household, the reasons for applying to the public housing sector (currently or soon to be homeless, or reasons related to health situation, family situation, job situation, inappropriate current housing, unpleasant environment), the type of housing looked for, whether the applicant is disabled and whether this is the first application. It is important to stress the fact that the application form contains very limited information about the ethnicity of the applicant: he or she only needs to inform about his or her nationality, which is limited to three possible categories (French, European Union, or non European Union).

We now document the selection process of the applicants. The commissions of selection in charge of allocating households to vacant public housing dwellings are held at the department level (or at the city level in the case of Paris which is both a city and a department due to its size).<sup>11</sup> The composition of the commissions is regulated by law: it includes six members of the public housing offices board, a representative of associations for social and economic insertion (appointed by the head of the department -*préfet*), mayors of the cities (or districts) in which vacant housings are to be attributed, as well as a representative of any association defending tenant rights. In addition, another department representative may attend the commission. For each vacant housing unit, at least

---

<sup>11</sup>Metropolitan France is divided into 22 large administrative areas, called *régions* (regions henceforth), and into 96 smaller administrative areas, called *départements* (departments henceforth). Each department is hence a subdivision of a region, and several departments can belong to the same region. Each department is administered by an elected General Council (*Conseil Général*) and its President, whose main areas of responsibility include the management of a number of social and welfare programs, junior high schools (*collèges*), buildings and technical staff, local roads, schools, rural buses, and municipal infrastructure.

three households must be considered by the commissioners, who finally decide which household will be allocated to which housing unit, according to the eligibility and priority criteria detailed above. Other criteria such as the number of children in the household are also taken into account in order to allocate suitable dwellings.<sup>12</sup>

With the allocation process regulated by legal rules at the national level, it seems unlikely that households can be allocated according to their origin. The main concern of the commissions is to favor socially endangered households, as shown by the priority criteria. Finally and most importantly perhaps, any decision based on the origin of an applicant, i.e. discriminating on this basis, is prohibited in France. Public housing offices are also regularly audited. If evidence of discrimination is detected, they are judged and punished accordingly. This is why the lawyers [Rouquette and Lipietz \(1991\)](#) stress that the rules of allocation of public housing units that prohibit "localism", and the high administrative barriers that effectively prohibit exchanges of lodgings except for changing spatial needs of families, make the allocation of public housing units largely exogenous with respect to the ethnic origins of the applicants.

Despite this legal process of allocation, one might still be worried about the possibility of self-sorting of households that refuse the residential allocation proposed by the commission. In theory, households can refuse up to three offers. However, self-sorting, especially on ethnic characteristics, seems unlikely to be a common practice. Residential mobility within the public housing sector is very low, due to the current strong shortage of supply of public housing dwellings. This makes it unlikely that the selected households could be really picky about the diversity of their neighborhood (see the study by [Simon, 2003](#)). In addition, rents are considerably lower in public housing than in private

---

<sup>12</sup>Public housing allocation in Paris serves as a useful concrete example. We draw on the official audit of *Observatoire du Logement et de l'Habitat de Paris* ([Obsevatoire, 2011](#)). Paris is a special case as it is, due to its size, a department as well as a city. The application form, the commission, and the allocation process thus take place in Paris, at the city level. As of January 2010, there were 186,017 public housing dwellings in Paris. Public housing buildings are scattered across all Parisian areas, with a high concentration (69 percent) in six districts (the 13th, 14th, 15th, 18th, 19th and 20th *arrondissements*). Within Paris, 48.7 percent of households are under the income ceiling and could be theoretically eligible. In practice, only households with very modest incomes apply (71 percent have an income lower than the minimum ceiling for all France, equivalent to 2,345 euros per month for a household with two children). On the 31st of December 2010, there were 121,937 ongoing applications, to be compared to 12,500 public housing units allocated over the year 2010. The breakdown of the population that were granted a public housing unit in 2010 is the following. 67.7 percent came from precarious housing, 28.8 percent came from the private rental sector, and 2.3 percent came from the public housing sector. In the latter case, those are people who moved for larger space following an increase in their household size (only 12 percent of the public housing dwellings have more than three rooms). The mobility rate (defined as the ratio of new entrants over the total number of public housing dwellings) is particularly low: it reaches 5.5 percent in 2010. It is formally possible to indicate a precise neighborhood in the application form, but in practice, very few applicants (6.6 percent) do provide this information. More than half of the 121,937 applicants (52.9 percent) did not mention any particular area at all, probably due to the fear of being rejected on this ground. Among those who indicated an area of preference, 91.2 percent mentioned the area where they were already living.



housing, increasing the opportunity cost of moving, so that the turnover is very low. More specifically, the mobility rate in the public housing sector is even lower than for recent owners. Using data from the 2002 Housing Survey, [Debrand and Taffin \(2005\)](#) give precise measures of the mobility rate: it amounts to 10.3 percent for new owners, to 15.9 percent for tenants in the private housing sector, but only to 9.9 percent for tenants in the public housing sector. While even 9.9 percent may seem high, we show in section 7.2 that when households move, they almost never achieve a placement in a less diverse setting in the public sector. Besides, the mobility rates seem to have become even lower in recent years due to the boom of prices in the private sector, as shown in the Parisian case in footnote 11. The authors also document an increase in the gap in the mobility rates between the private and the public rental markets: there is was 6 point difference in 2002, to be compared to a 0.8 point difference in 1984. As a consequence of the size of the eligible population and of the low turnover, the waiting periods are rather long: the 2002 Housing Survey documents that over one third of the population applying for a public housing unit had been waiting for more than one year. A closer look at the distribution of waiting periods reveals a difference between natives and immigrants, but this difference is washed out once we control for household characteristics: the main determinant of a longer waiting period is household size. This is not surprising, as the public housing market in France is characterized by a shortage of large apartments. This is part of the explanation of the difference in waiting period between immigrants and native French, as the former tend to have more children than the latter, on average.

In a word, the public housing market is very tight, and highly regulated. This implies that households have very limited control over the time when they will be assigned a public housing dwelling and the precise place where it will be located. This is especially true at the block level, which is our level of analysis. This gives some initial support to our assumption that the distribution of households across public housing blocks is blind to ethnic characteristics and preferences of households.

### **3.2 Tests on quasi-random distribution of ethnic groups shares across public housing blocks**

In the remainder of this section, we provide a more formal statistical test to show that the spatial allocation of households across public housing blocks within a given locality is exogenous with respect to ethnic characteristics. We focus here on what we see as the most stringent test on the spatial distribution of residents, which consists in comparing the observed distribution with a random distribution. This allows us to test directly our assumption of quasi-randomness of the allocation of households across public housing

blocks within a given area. In Section 7, we perform a variety of alternative tests. First, we show that while households moving into a new neighborhood tend to self-segregate in the unconstrained private housing market, there is no such evidence in the public housing market. This test points out the absence of self-selection along ethnic lines among the movers. But self-selection could occur prior to the move. Thus we also look at households that have refused a public housing dwelling offer. We show that households that have refused an offer end up living in public housing blocks that display the same ELF as those who accepted their first offer. Thus even if households try to be choosy with respect to the ethnic composition of their neighborhoods, they cannot self-segregate in the public housing sector due to the allocation process and the tight supply constraints of dwellings.

We now document our test on the quasi-random residential allocation in the public housing sector. As mentioned in section 3.1, the allocation of households across public housing blocks takes place at the department level. If the members of the commission follow the legal criteria and do not take into account the ethnic characteristics in the allocation process, we should find an equal distribution of households of a given nationality across the various public housing blocks within each department. For the sake of illustration, let us assume that 10 percent of Maghrebians live in the public housing sector in Paris. We should find the same share of 10 percent of Maghrebians within each Parisian housing block if the allocation was truly exogenous with respect to ethnic characteristics. Naturally, this equality of distribution of ethnic groups shares across housing blocks can hold only if we have a sufficiently large number of individuals within each housing block. Instead, in the Labor Force Survey, on which we perform the test, we only observe an average of 40.55 different individuals corresponding to 15.53 households in each neighborhood. This is due to the sampling strategy of the French National Institute of Statistics and Economics (INSEE) that interviews all the individuals from a given neighborhood, but consequently limits the size of the neighborhood. With such a small sample size of observations at the neighborhood level, any analytical test of equality of distribution of ethnic groups shares across blocks would fail. We thus use Monte Carlo simulation to reproduce an artificially random distribution of the population. We randomly reallocate the public housing population across the different blocks within each department, and then compare this random distribution to the actual distribution.

Let us now describe more precisely this test. We perform Monte Carlo simulations generating artificial random allocations that we later compare to the observed allocation. As mentioned above, we use the Labor Force Survey to perform this test, because its unique design allow us to work on the entire population of the surveyed neighborhoods. For each *département*, we pool the public housing population and reallocate it

randomly, without replacement, across the different residential blocks of the corresponding *département*, maintaining unchanged the actual size of each block. We get a simulated random allocation of individuals with a given characteristic across blocks. We then run a Kolmogorov-Smirnov (KS) test of equality of distribution of this given characteristic across neighborhoods with its actual distribution. More precisely, we are interested here in the distribution of the share of Native French and of the share of Maghrebians across neighborhoods. We finally determine the percentage of *départements* for which the actual and simulated distributions across housing blocks are similar, i.e. those for which we cannot reject the null hypothesis of equality of the distribution at the 10% level. The Monte-Carlo simulation results presented here are based on one hundred replications of the process described above. For each draw we run the tests for the equality of distributions, and then average the results.

Table 1.3 shows the values of those tests averaging over 100 Monte Carlo draws. The labels in the first column indicate the ethnic characteristic under consideration by distinguishing Maghrebian origins and French nationality at birth. The second column reports the results of the Kolmogorov-Smirnov test within the public housing sector. For the sake of comparison, we run in Column (3) the same KS-test on the full sample, including both those who live in the public and private housing sectors. The equality of spatial distribution between the random simulated distribution and the real observed one is accepted in most *départements* in the public housing sector. In particular, the equality of distribution with respect to Arabic origin (respectively French origin) is not rejected in 80 percent (respectively 70 percent) of the departments in the public sector. In contrast, Column (3) shows that in the whole sample, the equality of distribution is not rejected in 54.3 percent of the departments for the share of Maghrebians and falls dramatically to 24.8 percent of the department for the share of Native French. This test shows that while French natives (and to a lesser extent Maghrebians) do self-segregate a lot when we consider the whole sample, this is no longer the case in the public housing sector. All in all, those tests are supportive of our identifying assumption that the allocation of households across the public housing blocks can be considered as exogenous with respect to their ethnic backgrounds.

## 4 Results

### 4.1 Specification

This section estimates the impact of diversity on social relationships and public goods at the local housing block level. We identify the effect of diversity by using data from the public housing sector where households are exogenously allocated with respect to ethnic

characteristics. Let  $j$ ,  $k$  and  $l$  indicate respectively households, buildings and blocks. For each outcome, we estimate the following equation:

$$Y_{jkl} = \alpha + \beta ELF_l + \gamma X_j + \delta Z_k + \mu W_l + \varepsilon_{jkl} \quad (1.2)$$

where  $Y_{jkl}$  denotes the housing outcome we are interested in, as stated by household  $j$  in building  $k$  and block  $l$ ,  $ELF_l$  is the level of ethnic diversity in the block,  $X_j$  is a vector of household characteristics,  $Z_k$  a vector of building characteristics and  $W_l$  a vector of socio-economic characteristics of the block. We also control for department fixed effects since the spatial allocation of households across public housings is decided at the department level.<sup>13</sup> All results derive from OLS estimates, with robust standard errors clustered at the block level.<sup>14</sup> We control for a large set of household characteristics: age, gender, level of education, labor market status and nationality of the household head,<sup>15</sup> as well as household size, and total household income per member. We also control for building characteristics, with the (log)-number of apartments in the housing project, and its date of construction. Indeed, the size and the number of occupants might affect the ability of the households to coordinate for improving the commons or to enforce norms, while the age of the building might explain part of the degradations observed and tenant satisfaction.

An important issue is whether the degree of fractionalization is picking up various dimensions of the environment where people are living, including the extent of inequality and the unemployment rate or the socio-economic background of the neighborhood (Alesina and La Ferrara, 2002). We therefore include a very detailed classification in 27 categories of the socio-economic environment of each neighborhood, constructed by Tabard (2002) from the INSEE. This classification characterizes each area according to the socio-professional category and the occupation of all men in the area. We use the classification that was built using the 1999 census data. This variable is the most detailed one available in French national surveys to capture the socio-economic background of an area. We also include the unemployment rate computed at the block level using the 1999 Population Census data. Finally, we include department fixed effects.

---

<sup>13</sup>Results were unaltered by the inclusion of city fixed effect to account for the fact that the mayors are members of the attribution committee.

<sup>14</sup>We have also run logistic regressions on dummy outcomes, with similar results. To ease the interpretation of the coefficients, we will report the OLS estimates henceforth.

<sup>15</sup>We distinguish between the following categories for nationalities: French at birth, naturalized French, from other European countries, Maghrebian, Sub-Saharan African, Asian and all others.

## 4.2 The effect of fractionalization on the opinion about housing conditions

We start by documenting the impact of diversity on the overall satisfaction about housing conditions. From the HS, we use the question: "In general how do you judge the quality of your housing conditions?". The variable takes on values from 1, for very good, to 5 for very bad. Over the public housing population, the average of this variable is of 2.5, with a 0.98 standard deviation. This question on well-being related to housing conditions is rather general. We will detail the different dimensions that could affect this well-being in the following subsection.

Table 1.4 shows that ethnic diversity has a negative effect on satisfaction with housing condition, statistically significant at the 5 percent level. A one standard deviation increase in ethnic diversity generates an increase in the dissatisfaction with housing conditions that amounts to 6.7 percent of its standard deviation. To get a better sense of the magnitude of this effect, we can say that the increase in the dissatisfaction with housing conditions generated by a one standard deviation increase in block unemployment rate corresponds to 13.1 percent of its standard deviation. Thus the effect of diversity on satisfaction is as sizeable as half the effect of the local unemployment rate. Two other variables seem to be related to household satisfaction with housing conditions: members of larger households tend to be less satisfied with their housing condition, while those living in newer buildings (constructed after 1990) have a significantly better opinion on the subject than others. Finally, older and more educated individuals also complain less than others, but to a lesser extent.<sup>16</sup>

## 4.3 The various effects of diversity

### 4.3.1 Vandalism, housing quality and public safety

This section looks further at the various dimensions of the dissatisfaction with housing conditions that could be affected by ethnic fractionalization. The HS covers a large variety of questions on social relationships and the quality of the housing environment. Table 1.5 reports descriptive statistics of the outcomes we look at. To organize the discussion about those questions, we distinguish three main dimensions: (a) Behaviors and Public goods that are directly under the control of the tenants. This category includes in particular the neglect or voluntary degradations that may be imputable to the tenants, (b) Behaviors and Public goods that are under the control of the landlords. This category include in particular the poor quality of housing due to a lack of maintenance and repairs by the landlords and

---

<sup>16</sup>The results reported in 1.4 also show that people of Asian nationalities tend to be more satisfied with their housing condition than native French. However, we do not give much credit to this figure given that we observe only 9 Asian individuals in the public housing sector in our dataset.

(c) Interpersonal conflict and hatred social relationship. This category includes personal aggression and robberies to which we refer as civil conflicts or public safety. We have also run an exploratory factor analysis that yields similar, if not identical, categories. Appendix C reports the results obtained with the three indices resulting from the factor analysis.

The first dimension of housing quality refers to actions or goods that are largely under the control of the tenants. In this category, we include all the variables reporting neglect or voluntary deterioration in the common areas of the building. First, households are asked a general question on degradations: *"Were the common areas of your building (lobby, staircase, floors) vandalized or neglected (destruction, deterioration) over the last twelve months?"*. The answers are 1 for "Never", 2 for "Minor degradations" and 3 for "Major or very frequent degradations". Households are then asked to mention which kind of degradations they observed over the previous year. They can choose several possible answers from the following list: *graffiti or degradations of the walls (or on the floor), trash and litter on the floor, broken windows, broken doors, broken light bulbs, degradation of mail boxes, degradation of the entry phone or entry code, deterioration of the elevator*. For each outcome, the variable is coded as 1 in case of a degradation, and 0 otherwise. All those items refer more or less directly to a willful degradation. We will thus refer to this set of questions as the category *Vandalism*. We also include in this category a question about noise pollution:<sup>17</sup> *"How frequently are you disturbed by the noise in your housing during the day?", "During the night?"*. The answers are 1 for "Infrequently or never", 2 for "Rather frequently", and 3 for "Very frequently".

The second category we consider refers to goods that are not directly produced or altered by residents. But they might be related to diversity by the lack of maintenance and repairs by the HLM office to improve the housing quality. We will henceforth label this category *Poor Quality of Housing*. We include in this category variables corresponding to housing problems that can neither be caused nor solved by the tenants, but for which HLM offices are responsible. The households are first asked: *"How would you qualify the way the common areas of your building are maintained and taken care of (cleaning, maintenance of collective facilities: lighting, trash cans,...)?"*. The answer ranges from 1 for good, to 2 for average, and 3 for bad. More specific questions are also asked: *"How does the façade of your building look?"*,<sup>18</sup> *"What is the quality of the soundproofing of your housing?"*,<sup>19</sup> *"Was the elevator out of order during more than 24 hours over the past three months?"*,<sup>20</sup> *"Did you experience toilet issues (leaks, flush breakdown, drainage*

<sup>17</sup>The underlying assumption is that the source of the noise in the hallways and apartments of the building is not due to poor soundproofing. As a matter of fact, we see in the following sections that more diversity increases the disturbance related to noise, but fails to explain the quality of soundproofing.

<sup>18</sup>There are five possible answers: 1=As new, 2=Good, 3=Average, 4=Dirty, 5= Bad, with cracks, 6=Very bad, the building threatens to collapse.

<sup>19</sup>The possible answers are: 1=Good, 2=Average, 3=Bad.

<sup>20</sup>In contrast, the question mentioned in the *Vandalism* section refers to the interior status of the elevator

problem) over the last three months?", or "Did you experience coldness in your apartment during more than 24 hours over the past twelve months?". We also include more detailed questions concerning the origin of coldness: "Did you experience coldness because of a bad insulation?", "Did you experience coldness because the heating equipment broke down?" and "Did you experience coldness because of a poor heating equipment?". For all the previous questions, the variable is 1 when the answer is "Yes" and 0 otherwise.

The last category of questions refers to personal aggressions and criminality. We will label this category *Public Safety*. Three questions correspond to this category: "Have you, or a member of your household, been a victim of or a witness to physical aggression in your neighborhood during the last twelve months?", "Have you, or a member of your household, been a victim of or a witness to a robbery in your neighborhood during the last twelve months?", and "Have you been victim of a burglary (or any attempt) over the past twelve months?". For these three questions, the variable equals 1 in case of the event, and zero otherwise.

#### 4.3.2 Results

Tables 1.6 to 1.8 show the effect of ethnic fractionalization on the various outcomes corresponding to the three different dimensions: "Vandalism", "Quality of housing" and "Public Safety". For each outcome, we run a separate regression according to equation 1.2, using the largest set of control variables we could think of. More precisely, we report the results when we control for households characteristics, building characteristics (which may explain a large part of the degradations observed in the housing projects), neighborhood characteristics and department fixed effects.

Table 1.6 reports the effect of ethnic diversity on outcomes related to voluntary degradations and vandalism. We report the results in the public housing environment, with the full set of controls as detailed above. For almost all the outcomes considered, the estimated effect of ethnic diversity is statistically significant at the 1 percent level,<sup>21</sup> and is sizeable. Let us for instance look at the results for graffiti in the full-specification: a one standard deviation increase in ethnic diversity is associated with a rise by 5.6 percentage points in the probability of observing graffiti, which represents 12.8 percent of the total standard deviation of this outcome. The effect of ethnic diversity is comparable to the effect of local unemployment: a one standard deviation increase in block unemployment rate is associated with a rise by 4.56 percentage in the probability of observing graffiti. Regarding the deterioration of elevators, a one standard deviation increase in ethnic diversity induces a 4.9 percentage points increase in the probability of observing degradation

---

rather than its mechanical breakdown.

<sup>21</sup>The exception is for the indicator for broken doors and noise during the day, for which the effect of diversity is only significant at the 5% level.

of the elevator, which represents 16.7 percent of the total standard deviation of this outcome. This effect is once again as sizeable as that of the local unemployment rate. Note also that the size of the building (i.e. the number of housings) has a strong positive impact on all the outcomes related to vandalism.

The second set of regressions, presented in Table 1.7 shows the effect of diversity on outcomes signaling poor quality of housing. The coefficient associated with ethnic diversity is generally significant, except for the outcome related to the quality of soundproofing.<sup>22</sup> More diverse neighborhoods are characterized by a lower care of the commons by the persons in charge, a poorer condition of the façade, more frequent concerns with heating, more frequent elevator breakdowns and toilet issues. It is worth noting that the estimated effects of diversity are much lower than those found for outcomes associated with vandalism. Consider the outcome associated with the probability that the elevator is out of order. We find that when the ELF increases by one standard deviation, the probability that the elevator was out of order during at least 24 hours over the last three months rises by 1.9 percentage points. This corresponds to only 5.24% of the standard deviation of this outcome. If we now turn to heating issues, our results indicate that a one standard deviation increase in ethnic diversity is associated with a rise by 2.43 percentage points in the probability to have experienced insufficient heat in the apartment during more than 24 hours over the past year, which represents 6.41 percent of the total standard deviation of this outcome. The date of construction of the building is also an important explanatory variable for most of the outcomes related to general housing quality, as it accounts for the general state of capital equipment under the responsibility of the public housing office (heating, façade, soundproofing,...). Living in a more recent building decreases particularly the probability to report poor condition of the outside walls, low quality of insulation or poor quality of soundproofing.

Finally, Table 1.8 reports the results for outcomes related to public safety, capturing direct aggression, robberies and burglaries. Remarkably, it shows that ethnic diversity does not have a significant impact on any of these outcome variables in our preferred specification. This finding is consistent with Fearon and Laitin (1996), who argue that despite inter-ethnic relations being generally more tense, in-group policing mechanisms typically keep violence off of the equilibrium path.

In sum, and taking advantage of data at a more micro level than has heretofore been available, we see that fractionalization operates with different degrees of impact for different sorts of public goods and social relationships.<sup>23</sup> To be sure, results are not significant for all of the outcomes that we examine. But overall, the results are clear that

---

<sup>22</sup>Two other variables are not explained by diversity, but these are not outcomes *per se*. These variables are reasons why the household experience coldness in the apartment.

<sup>23</sup>Our findings are unchanged with regressions on aggregated indices obtained with a principal component analysis (see Appendix C), and with a mean effect analysis (see Appendix C).



fractionalization at the local level increases vandalism by a great deal, decreases building maintenance by a moderate (but overall significant) degree, and has no effect on security. These findings allow us (as we do in the next section) to propose the various channels through which fractionalization works in the provision of public goods.

### 4.3.3 Rationalization of the channels

To rationalize our findings, we propose different interpretations of the channels through which fractionalization affects social relationships and local public goods. The category "Vandalism" refers to the neglect or voluntary degradations of the common areas of the building, such as damaging common property, graffiti, or depositing trash on the floor. These are outcomes over which public housing residents have control and for which they can be held responsible.<sup>24</sup> The category "Quality of housing" include variables such as the condition of the outside walls, quality of soundproofing or coldness in the apartment. Those variables are more the responsibility of the public housing managers. Finally, the "Public Safety" category represents outcomes that are less under the control of local authorities than of the state police. We find that both locally controlled outcomes are negatively affected by diversity, and we provide a different rationalization of the channels for each type.

Our interpretation for the results on "Vandalism" is that diversity prevents the creation of social norms to punish defectors, as the threat of social sanctions is lower across groups. The other-regarding preferences are less effective in more diverse areas. This has been a standard result in the literature since the seminal work of Coleman (1988), and it helps explain why we observe more voluntary degradations with diversity. Supporting our intuition, many households living in the public housing sector report having "no relationship at all" with their neighbors, rather than "bad" or "very bad relationships", which can be a barrier to the creation of other-regarding social norms. The increase in graffiti in more diverse areas might also illustrate the need to mark one's territory in a context where several groups co-exist.

We understand the result on "Quality of housing" as the inability of more heterogeneous communities to undertake collective action that would pressure the public housing office into improving housing quality. This could be sustained (though we have no direct evidence to support this) by beliefs in the housing directorate that it need not maintain public goods to high standards in heterogeneous housing projects because the likelihood of collective action against it is minimal. In this sense, the resulting poor housing quality associated with ethnic diversity can be seen as an equilibrium in which the lack of expectations of collective action would fail to incentivize the housing directorate to make

---

<sup>24</sup>Given that residents need to enter a code in order to gain entry into their building, it is unlikely these degradations are coming from outsiders.

costly improvements.<sup>25</sup> Coming back to the results concerning heating issues, we can find some support for this assumption: we find that households living in more diverse neighborhoods not only report more heating failures, but also report that this is due to the poor quality of the heating equipment, an appliance typically under the control of the HLM office.<sup>26</sup>

Finally, we can think of two possible interpretations of the absence of any diversity effect on aggressions and robberies. First, this could result from more physical security provided by the city and state police in more diverse neighborhoods. The second explanation would be that individuals living in the public housing sector in general experience social anomie. In fact, one third (32.7 percent) of the public housing population, irrespective of diversity, declares to have no relationship at all with individuals living in their same area. In addition, we find that individuals living in a more diverse neighborhood tend to even more social anomie and fewer relationships with their neighbors.

To summarize, our interpretation is that diversity generates social anomie, i.e. the absence of common rules and social norms. As a consequence of anomie, there is (a) less other-regarding preferences, hence more neglect and vandalism, (b) a failure to generate collective action to pressure the public housing offices into improving housing quality, and (c) fewer opportunities for violent confrontation at all levels of diversity. We also interpret the lack of an effect of diversity on violence by security provided at a higher level of administration, not subject to the constraints of local diversity.

#### 4.3.4 Interpretation of the channels based on repairs

We bring additional evidence on the interpretation of the channels by looking at maintenance and repairs performed in the building. Note first that these outcomes add an objective dimension to the previous subjective questions. The variation in the effects of diversity on the number of repairs depending on the type of public good also helps us to tease out the different channels through which diversity operates.

The Housing Survey asks whether elevators, staircase, windows, heating equipment, security equipment, and so on, have been repaired or installed during the previous year. We build three measures of repairs, corresponding to our three general outcomes. We define a first variable tracking repairs that can be fixed internally by the tenants. Those repairs concern staircase, windows, doors and lights of the commons, i.e repairs related to neglect or voluntary degradations. A second variable indicates repairs that can be fixed only by the external intervention of the public housing office. Those repairs include

---

<sup>25</sup>The collective action could also influence mayor's office. But the political logic of the public housing support is beyond the scope of the paper

<sup>26</sup>Another possible reason for having experienced coldness in the apartment that the household can mention is to have restricted heating in order to save money. The results are not reported here, but we find no significant effect of diversity on this outcome, in any specification.

revamping of the façade, or interventions to improve, among other things, the heating system or insulation quality, i.e. repairs related to the general quality of housing. Finally, we build a third variable accounting for the installation of security equipment in the building, which can be related to conflicts outcomes. We then regress each of these three variables (as well as less aggregated indicators of repairs) on the level of diversity of the block, controlling for factors that could explain the number of repairs: the number of dwellings in the building, and its date of construction. Table 1.9 reports these OLS estimates.

Column 1 of Table 1.9 reveals a positive and statistically significant correlation between the probability of repairs inside the building (windows, doors, lights... in the common areas) and local diversity: the more the diversity, the more the work for repairing the effects of vandalism. In the main regressions of the paper presented in section 4.3.2 (corresponding to the Table 1.6), we found that voluntary degradations increase with diversity. This implies that the larger number of repairs results from greater need due to a lack of publicly spirited social norms rather than from greater responsiveness by the housing authorities to regular maintenance.

Column 2 shows a negative and statistically significant correlation between the number of substantial works in the building (façade, heating, insulation...) and local diversity: the more the diversity, the less the work asked by the landlords for improving the general quality of housing. In the main regressions presented in section 4.3.2 (corresponding to the Table 1.7), we found that more diversity implies a lower quality of housing. Thus, it seems that more diverse neighborhoods are deprived of such substantial work, although the inhabitants actually complain (individually to survey enumerators) about the quality of housing. This supports our intuition that tenants in more diverse neighborhoods are unable to engage in collective action to pressure the public housing offices into undertaking important works.

Finally Column 3 shows a positive and statistically significant correlation between the existence of security equipment and local diversity. In the main regressions presented in section 4.3.2 (corresponding to the Table 1.8), we found no impact of diversity on aggression and robberies. The presence of security equipment in more diverse neighborhood might be part of the explanation of the absence of diversity effect on burglaries. In addition, it is supportive of our idea that vandalism in the common areas of the buildings is imputable to the tenants, who have access to the building.

## 5 Robustness checks on self-reported quality of public goods

One concern in the previous analysis is related to the subjective nature of the outcome variables used in our study. Self-reported perceptions might reflect personal bias rather than be correlated with objective measures of public good provision. Perhaps people are just happier when they are surrounded by people more like themselves, and this is reflected in their answer to the quality of housing. We conduct several tests challenging this alternative explanation of personal bias.

First, as mentioned in the previous section, the HS provides information about various types of repairs and work that have been done in the building or in the housing unit over the previous year. These variables present the advantage of being objective. The lower part of Table 1.9 reports simple correlations between the various outcomes and the associated repairs. We find that most of our subjective outcomes are strongly and positively correlated with the existence of repairs, i.e. objective outcomes, especially for the variables related to vandalism.<sup>27</sup> This is our initial evidence of the reliability of our subjective measures of housing quality and well-being.

We then conduct more formal tests. We replicate the regression on the dissatisfaction with housing conditions (section 4.2) including interaction terms between diversity and the various ethnic groups. Those estimates reveal whether different groups react in different ways to the level of diversity of their neighborhood. Column 2 of Table 1.10 shows that there does not seem to be a different effect for the various groups, and the coefficient for diversity remains unchanged (see Column 1 for the baseline specification). Then we concentrate on actual differences between "pure French" households<sup>28</sup> and fully Maghrebian households' dissatisfaction with housing conditions. In particular, we interact the dummies of being in a fully native French household or being in a fully Maghrebian household with the ELF: none of the coefficients is significant (see Table 1.10, Column 3). Thus for any given level of diversity, there is no significant difference in the answers given by pure French and fully Maghrebian households. In other words, the idea that bad opinions of housing conditions are driven by average bad feelings due to being surrounded by foreigners can be rejected. Moreover, including these additional controls only slightly affects the magnitude of the ethnic diversity coefficient, and does not affect its direction or its significance.

---

<sup>27</sup>An exception is the condition of the outside walls, which is negatively correlated with the probability that façade work was done. This is not surprising as the assessment of the façade's condition is done at the time of the survey, while repairs concern the previous year.

<sup>28</sup>Both children and parents were born French in France.

An alternative test to show that subjective perceptions have an objective foundation is to look at how much within-housing project variation there is in perceptions. We regress self-reported perception on housing project fixed effects and individual characteristics. Once we control for housing project fixed effects, assuming there is no within-project variation in public goods, the remainder of the variation tells us if certain ethnic or socio-demographic groups are more likely to be positively or negatively biased. If perceptions have a high signal-to-noise ratio, there should be less within-project variation because perceptions would be a good signal of project level public goods. Table 1.11 shows the regressions of our main indicator of satisfaction about housing conditions on individual characteristics. Column 1 shows the within-housing project estimates by including housing project fixed effects, exploiting variation within public housing. Column 2 shows the between-housing project estimates. Column 1 shows that the only individual characteristics statistically significantly correlated with within-project variation in the perception of the environment are age and household size. Income, education or the country of origin of the households are uncorrelated with perceptions of the environment. We also compute the standard deviation in the perception of the quality of housing between public housing projects and within public housing projects. The standard deviation is almost twice as high across blocks (.801) than within blocks (.435), and this difference is statistically significant. In sum, low levels of within block variation on perceptions adds confidence that there is an objective foundation for tenants' subjective reports.

Finally, we also estimate the effect of diversity on the perception of the quality of public goods that are financed by the city, the department or the state rather than locally financed by the HLM offices. If there is a reporting bias in general, then, the effect on all types of public goods should be the same. If it is related to localized collective action failures, then the impact should only be on locally provided/maintained public goods. Thus this test provides both an additional robustness check on the channels through which diversity affects public goods and on the absence of a reporting bias. The local public goods we have focused on so far (except for individuals' protection) are provided or maintained by the private company that owns and manages the public housing building. We now consider public goods that are managed at the city or department level. In the HS, three public goods enter this category. The first one is the perception of the quality of roads and streets with the following question: *"What is your opinion about the maintenance of the streets, roads and public spaces in the area?"*. The second question measures the access to public transportation: *"What is your opinion about the accessibility of your area by public transportation?"*. The third item measures the accessibility of the area by private transportation: *"What is your opinion about the accessibility of your area by private vehi-*

*cles (parking, congestion)?"*. The answer ranges for all three questions from 1 for good, 2 for neither good nor bad to 3 for bad. Table 1.12 reports the OLS estimates, controlling for all the previous household and local characteristics in addition to department fixed effects. We find that ethnic fractionalization is neither correlated with the quality of public spaces and roads in the areas (Column 1), nor with public transportation (Column 2), nor with car parking and general congestion (Column 3). Again, our confidence that the subjective reports to enumerators on housing quality have an objective foundation is increased.

## 6 Further tests: Fractionalization and ethnic shares

The basic regressions measure ethnic diversity using a standard ELF, controlling for household, building and neighborhood characteristics. Yet, as suggested by Vigdor (2002), it might be important to control for ethnic group shares to get a more comprehensive set of covariates for diversity. Column 1 of Table 1.13 reports the results once we control for ethnic group shares.<sup>29</sup> The estimated impact of ELF is now even stronger than in the previous specifications, confirming the robustness of our result along this dimension.

Moreover, we run regressions replacing the fractionalization index by ethnic groups shares (Column 2 of Table 1.13), and by ethnic group shares and their square (Column 3 of Table 1.13), controlling for the usual individual and local characteristics. Only one group (Maghrebian) seems to have a significantly negative effect on the dissatisfaction with housing conditions: the higher the share of Maghrebians in a block (relative to the share of French), the more likely individuals are to complain about their housing conditions. However, this negative effect decreases with the share of Maghrebians. From this result, we infer that our measure of diversity reflects not only the actual ethnic composition of the neighborhood, but also that some ethnic groups might have different effects on self-reported perceptions of the quality of public spaces as they become a majority of the neighborhood population. However, this result does not call into question the effect of diversity per se on which we have already reported (Column 1 of Table 1.13).

Finally, we re-run our main regressions using an alternative fractionalization index, trying to encompass another dimension of diversity. More precisely, in order to account for communication issues potentially related to the diversity of origins, we compute an alternative fractionalization index based on a proxy for French speaking. We use information provided by the International Organisation of La Francophonie to group countries

---

<sup>29</sup>In Table 1.13, we aggregate the various nationalities at birth into six different categories. The results are similar when we work with more detailed shares for all nationalities. The share of native French is the omitted category.

according to the share of the population that is French speaking (in 2010). Their classification allows us to distinguish among six groups of countries: countries which are not members of the organisation,<sup>30</sup> countries in which French speakers represent less than 5 percent of the population, countries in which French speakers represent between 5 and 15 percent of the population, countries in which French speakers represent between 16 and 35 percent of the population, countries in which French speakers represent between 36 and 60 percent of the population, and countries in which French speakers represent more than 60 percent of the population. We then assign one of the six French-speaking levels to each individual (from the 1999 census dataset), according to his or her nationality at birth. This sorting of individuals captures the probability that they actually speak French. Finally, we compute a standard ELF for each block (using the same methodology as for our main index of diversity) relying on the shares of the block population belonging to one of the six groups. We replicate the regressions of the paper<sup>31</sup> using this alternative index instead of the one based on nationality at birth. The results are reported in Table 1.14. Each coefficient comes from a separate regression. The columns indicate the four dependent variables under study. Panel A and panel B respectively correspond to the measure of diversity used in each regression. Our results are unaffected when we use the new index based on French speaking origin. This is not very surprising given that the correlation between the two indices is very large (98.45 %).

## 7 Additional tests on the exogeneity of residential allocation in the public housing sector

### 7.1 Absence of self-sorting on ethnic backgrounds

Our first set of alternative tests consists in showing that while households tend to self-segregate in the unconstrained private housing market, there is no such evidence in the public housing market. We test this using the LFS and focusing on individuals who recently moved into an area (within the previous year).

We first estimate the correlation between the hourly wages of the movers and the level of diversity of the area into which they just moved. Without prior beliefs over agents' preferences, if individuals have a taste for or against homogeneity, there should be a significant relationship between the level of ethnic diversity prevailing in their neighborhood

---

<sup>30</sup>For these countries, the organisation does not provide any data, but we can reasonably assume that the share of French speaking population in non-member countries is close to zero.

<sup>31</sup>We replicate the regression of the dissatisfaction with housing conditions (Columns 1) as well as of the three aggregate indices we obtained with the principal component analysis presented in Appendix B.

and their wages when their choice is not constrained by legal rules. Indeed, in an unconstrained market (e.g. the private housing market), the richer the individual, the easier it should be for him or her to choose his or her neighborhood. Therefore, if the level of diversity of the area enters one's preferences, there should be a correlation between individual wealth and the level of diversity in the area. In the public housing market as well, some public housing units are more expensive than others, depending on the location and the date of construction. The wealthiest inhabitants could thus have some control over the diversity of their neighborhood, in theory.

To test this, we compute the fractionalization index of the area to which a household recently moved taking into account only the neighbors who had been living there for more than one year. We thus calculate the fractionalization indices at stake prior to the move. For the private housing market, we compute the fractionalization index of the whole area, including both the population living in private and public housing dwellings within this area. We follow this strategy since there are a few areas with both public and private housing units in the LFS. It is reasonable to think that it is the level of diversity of the whole neighborhood that will matter in the mobility decision in the private market.<sup>32</sup> Regarding the public housing sector, we compute the level of diversity including residents of the public housing only, our identification assumption being that households do not have control over the level of diversity of their neighborhood *within* the public housing sector. We run OLS estimates of the hourly wage of newly arrived individuals on the level of diversity of the area in which they just moved, controlling for the department of residence.<sup>33</sup> First, we focus on individuals having just moved into a private dwelling. We find a very strong negative relationship between income and diversity (the estimated coefficient is -0.14 and is significant at the 1 percent level).<sup>34</sup> Then we look at the sample of individuals having moved into a public housing dwelling within the past year. In this case, the simple OLS regression reveals that there is no significant correlation between the income of individuals moving into a public housing dwelling and pre-existing diversity of nationalities within the neighborhood.<sup>35</sup> These correlations show that while the wealthiest households tend to self-segregate in less diverse areas in the (unregulated) private housing market, it does not seem to be the case in the (regulated) public housing market. In other words, although diversity enters households' preferences as revealed by the private housing market result, the location in the public housing market seems to be unaffected by such preferences.

---

<sup>32</sup>The results are unchanged if we consider only the population living in the private housing sector: the magnitude of the correlation decreases marginally, but remains statistically significant.

<sup>33</sup>If we reverse the dependent and the explanatory variables, the sign and significance level of the coefficient remains the same.

<sup>34</sup>This is powerful evidence of the bias introduced with endogenous sorting

<sup>35</sup>The results are not displayed but are available upon request.



Our second test uses the same methodology and estimates the link between the origin (nationality) of individuals moving into a new area and the share of the area's "long term" population of the same origin.<sup>36</sup> We expect a significant relationship in the private housing market where location choice is relatively unconstrained but not in the public housing sector. Table 1.15 reports the results from an OLS regression of the share of neighbors from the same origin as new movers on new movers' characteristics: nationality group, public housing dummy, quadratic function of age, hourly wage (log) education, socio-economic category, department fixed effects, and interaction of individual characteristics with the public housing dummy. We consider seven different nationality groups: native French, naturalized French, Europeans, Maghrebians, other Africans, Asians, and other nationalities, which is taken as the reference group.

Three facts are worth noting here. First, there is indeed evidence that on average native French are significantly more likely to move in neighborhoods where the share of natives is higher, compared to households from other nationalities. This is not surprising given the fact that natives make up a large majority in the French population. The second interesting point is that the coefficient for living in the public housing sector is negative and significant at the 5% level. More precisely, it reveals that HLM households move in areas where the share of individuals from the same origin is on average 15.5% lower than for households in the private housing sector. This result strengthens the idea that the extent to which households in the public housing sector live close to their co-ethnics is lower than in the private sector. Finally, when we interact nationalities with the public housing dummy, none of the coefficients but one are significant. This comforts us with the idea that there is no particular self-segregation along ethnic lines in the public housing sector. The only coefficient that is significant is for naturalized French individuals, which is not a clear ethnic group. However, when we control for the share of one's own ethnic group in one's department in column (2)<sup>37</sup>, even this interaction term turns insignificant.<sup>38</sup>

---

<sup>36</sup>A similar test was proposed by Goux and Maurin (2007) to show that the educational achievement of the children of newcomers in public housing is uncorrelated with that of the current residents. Individuals do not self-select in public housing neighborhoods according to the educational achievement of the neighbors' children. By contrast, the authors find a strong self-selection on the educational characteristics in the private housing sector.

<sup>37</sup>The purpose of including this variables is to account for the fact that part of the ethnic sorting is likely to be due to an over-representation of some groups in given departments and in particular in the public housing sector.

<sup>38</sup>In a previous version of the paper, we regressed individuals' origin on the share of the population of his or her new neighborhood from each nationality in each housing sector, rather than pooling individuals from all origins. The results were the following. In the private housing sector, a significant relationship between one's nationality and the share of same-origin neighbors showed up for most of the nationality groups. By contrast, in the public housing sector, there was no statistically significant relationship between the nationality of the individual and the share of the "long term" population in the area having the same nationality. The correlation was close to zero for households with African origins, and was around three times as low as in the private sector for households with Maghrebian origins. The only significant relationship showed up for immigrants from Europe, who represent a marginal share of the whole immigrant population

We conduct the same kind of test on other individual characteristics, and reach similar conclusions. We find that in the private sector, highly educated (respectively low skilled) individuals are very likely to move into neighborhoods with higher levels of highly educated (respectively low skilled) people. This is not surprising and illustrates self segregation along education level in the private sector. On the contrary, such segregation does not appear in the public housing sector. The only characteristic for which we find a positive correlation between the new and the old inhabitants in public housing blocks is the fact of being a factory worker. This is perhaps not too surprising either given that factory workers represent more than 30 percent of the public housing population, and due to the history of public housing, which was initially (and over several decades) dedicated to factory workers.

We also regress the probability of having moved in a new HLM dwelling in the past year (dummy equal to one in this case and to zero if the individual was already living in the same HLM apartment one year before) on individual characteristics (nationality, age, gender, wage, education, socio professional group), and the interaction of these characteristics with the ethnic diversity among the public housing population of the block.<sup>39</sup> As would be expected in the absence of sorting, the coefficients on the interaction terms are not significantly different from zero. The only exception is for the interaction of ELF with the dummy for African nationality, for which the coefficient is negative and significant at the 10 percent level.

Finally, we build on the information about relationships provided by the Housing Survey to bring additional evidence on the absence of sorting. More specifically, respondents to the HS are asked the following questions. "*How would you qualify the relationships with your close neighbors?*" and "*How would you qualify the relationships in your neighborhood?*", where the first question refers to the direct neighborhood, while the second one refers to a larger neighborhood. The possible answers are 1 for "Good", 2 for "Average", 3 for "Bad" and 4 for "No relationship". From this, we build two dummy variables indicating whether the individual reports any relationship (good, average or bad) in the direct or in the larger neighborhood, respectively. Simple descriptive statistics and correlations reveal that living in the public housing sector is associated with a significantly higher probability to have no relationship at all with your neighbors, the more so the larger the neighborhood considered. However, when we control for neighborhood, households and building characteristics, there is no significant difference between the private and the

---

compared to immigrants from the Maghreb and Africa. However, because the standard errors were quite large in the public housing sector due to a low number of observations, we decided to run this alternative test with pooled data.

<sup>39</sup>The results are not displayed in the paper but are available from authors upon request.

public housing sectors in terms of existence of relationships. In addition, for individuals reporting having relationships, we built two variables characterizing the quality of the relationship in the direct or in the larger neighborhood, respectively. In this case, simple correlations reveal that individuals living in the public housing sector tend to have worse relationship with their neighbors, if any. This relationship remains significant when we control for neighborhood, households and building characteristics.<sup>40</sup> To summarize, public housing neighborhoods are characterized by an absence of any relationship, and by a decrease in the quality of relationship when they exist. This brings additional support to our assumption that public housing tenants do not choose their neighbors, otherwise we would expect them to report more frequent relationship, of better quality.

## 7.2 Tests on the refusal rate of public housing offers

The previous tests point out the absence of self-selection along ethnic lines among the movers. But self-selection could occur prior to the move. In this case the sample of movers that we observe in the database would be biased. We address this issue by looking at households that have refused a public housing dwelling offer. Actually, a disturbing fact for our assumption is that a non-negligible share of households waiting to be allocated into a public housing unit report to have declined at least one offer. In the Housing Survey, 24.2 percent of households currently living in a public housing dwelling report to have rejected at least one proposal before finding their current place. Besides, 16.5 percent of the households that are still waiting for an offer at the time of the survey - whether they are already living in a public housing dwelling or not- have previously turned down at least one offer. An additional concern is that 47.9 percent of the households in public housing that had declined an offer at the time they were on the waiting list declared that one of the major reasons for this decision was they found the local environment unpleasant.<sup>41</sup> The corresponding figure for the households still waiting for an answer amounts to 57 percent. The answer "unpleasant local environment" is hard to interpret at this stage, since it could refer to diversity as well as the proximity to public transport and infrastructure, lack of green spaces and so on.

Yet, we show that even if households declined at least one offer, possibly due to the ethnic diversity of the neighborhood, they were still unable to choose the level of diversity of the area in which they end up living, and would not be able to do so for any neighborhood to which they would receive an allocation in the future. To put it another way, although households may have a distaste for diversity, we find evidence that this is

---

<sup>40</sup>The precise figures corresponding to the previous results are not reported in the paper but are available from the authors upon request.

<sup>41</sup>The other possible answers were: inconvenient place, rent too expensive, low quality building, and apartment not corresponding to household needs.

not taken into account in their allocation process by the attribution commissions. In principle, households can decline up to three offers. But due to the strong shortage of public housing dwellings, we find that households that have declined an offer in the past cannot self-select into less diverse neighborhoods in the future. We provide evidence of that fact in what follows.

First, if there were self-selection upon diversity, we should expect households that turned down proposals before being allocated to their current public housing dwelling to end up living in less diverse neighborhoods. To test this conjecture, we run OLS regressions of a variable indicating whether the household declined at least one offer (during the latest application process) on the level of diversity of the neighborhood in which it now lives.<sup>42</sup> Panel A-I of Table 1.16 shows various estimates of the effect of ethnic diversity on the probability of having turned down offers. Column 1 shows the correlation without any additional control variables. In Column 2, we control for household characteristics. We add up the characteristics of the housing project in Column 3. Column 4 finally includes neighborhood characteristics and department fixed effects since the allocation of a public housing dwelling takes place at the department level. In each specification, the coefficient on ELF is not significantly different from zero, showing that households having declined offers during their past allocation process do not end up living in neighborhoods with significantly different levels of diversity.

We explore further the validity of this conjecture by focusing on the reasons adumbrated by households for refusing an offer. If public housing residents were to sort themselves on the basis of their (dis)taste for diversity, those who declined "because of the local environment" should now live in significantly less diverse neighborhoods. We thus regress a dummy variable indicating whether an "unpleasant environment" was the reason why the household declined at least one offer (during the past application process) on the level of diversity of its current neighborhood. Panel A-II of Table 1.16 reports the estimates on the level of diversity, using the same specifications as above. Here again, none of the coefficients is significant. Instead, household characteristics such as the labor market status of the head of household and the size of the household are the only ones that matter in these regressions.

Alternatively, we perform these tests on the subsample of individuals currently waiting for an HLM offer. Panel B of Table 1.16 shows the regressions of the refusal dummy (B-I) and the "refusal due to unpleasant environment" dummy (B-II), for individuals who are currently applying for public housing on the diversity in their neighborhood. We still control for household, building and neighborhood characteristics. Once again, we

---

<sup>42</sup>In this paper, we always rely on OLS estimations, even when the dependent variables are dummies. Using probit estimates does significantly affect our results.

find that the ethnic diversity of a block is uncorrelated with households wait-listed for an HLM assignment having turned down offers since the beginning of their request (B-I). This suggests that the current level of diversity in the block does not rush households out of the area, as their propensity to decline an offer is independent of the ELF in the current neighborhood. The high refusal rates of HLM offers do not therefore seem driven by a hope to reduce diversity by waiting.

Let us now focus on individuals who left their previous housing unit because they did not like the environment. In the Housing Survey, 5 percent of households that moved over the past four years mention an unpleasant environment as one of the main reason they moved. In this question, the phrase "unpleasant environment" explicitly refers to troubles such as "noise, lifestyle or insecurity". Again, this could be related to high levels of diversity. If this is true, and if households can actually select the block to which they move, then we expect that those households having moved because they disliked their environment ended up living in less diverse neighborhoods than the households that moved for a different reason.

We perform OLS regressions of a variable indicating whether the household left its previous housing due to an unpleasant environment, on the level of diversity of its current neighborhood. Table 1.17 shows the coefficients on diversity in the specification controlling for household, building and neighborhood characteristics, and including department fixed effects. Column 1 shows the results for households that moved within the private housing market. As expected, households that left their previous housing to escape from an unpleasant environment now live in blocks where the diversity is significantly lower. Column 2 shows that this result does not hold for households that moved within the public housing market. This result suggests once again that in the public housing sector, households do not have control over the diversity of the block to which they are allocated.

A potential concern with the previous result is due to the small sample of observations (only 627 in the public housing case), generating large standard errors. Therefore, we replicate this test on a larger subsample. Instead of focusing on households that have moved within a housing sector, we now concentrate on households having moved into each sector, no matter the sector in which they were living prior to their move.<sup>43</sup> As previously, we see that for households living in the private housing sector, the probability that they left their previous housing due to an unpleasant environment is negatively correlated with the diversity in the current neighborhood (Column 3). However, no such significant relationship shows up for households living in a public housing dwelling (Column 4), and the estimates are now more precise than in Column 2. We can infer from those tests that

---

<sup>43</sup>To summarize, Columns 1 and 2 report the results for households moving from a housing dwelling in the private (1) and public (2) sectors into a housing dwelling in the same sectors. Columns 3 and 4 report the results for households moving from any housing sector into the private (3) and public (4) sectors.

households tend to self-select in low-diversity neighborhoods in the private housing sector, but are unable to do so in the public housing sector.

## 8 Conclusion

This paper exploits French public housing policy as a natural experiment to identify the causal effect of diversity on well-being, social relationships and the quality of local public goods. The French Housing Survey provides a unique micro level of analysis of social interaction between adjacent neighbors within housing blocks. We provide a detailed analysis of the channels through which diversity operates at the local level while the previous literature focused so far on aggregate outcomes and channels. We use the exogenous allocation of households within public housing with respect to ethnic characteristics in France to address the bias from endogenous residential sorting that reduces the confidence in previous empirical findings on fractionalization. We find that fractionalization has a negative impact on other-regarding preferences, leading to higher neglect and vandalism in the housing commons. Fractionalization also undermines collective action for the improvement of the quality of housing. But in our context, fractionalization has no effect on public safety, diversity being associated with social anomie within the housing blocks rather than violent confrontations among neighbors.

This natural experiment calls for future research on the specific role of national, local and informal institutions in mitigating or magnifying the effect of ethnic diversity on the provision of public goods. France is a country with a republican tradition that resolutely refuses to reify ethnic identification as a strategy to prevent the ethnification of everyday life. Yet we find a significant negative effect of diversity on local public goods in its public housing sector, comparable to the association found in the US localities where multiculturalist institutions regulate ethnic relations (Putnam, 2007) and in cases where public institutions are weak (Alesina and La Ferrara, 2005). However, on issues of physical security in French public housing, the costs to ethnic diversity disappear. This may be due to the emergence of informal institutions (such as in-group policing as in Fearon and Laitin, 1996) or the supremacy of state-level institutions in which local diversity plays no role in the supply of order. In any event, the results raise a puzzle, to be addressed in future research, on the general power of institutional arguments in overcoming the negative implications of ethnic heterogeneity on the provision of public goods.

# Tables

**Table 1.1:** Public Housing and Private Housing population characteristics (households heads, Housing Survey 2002)

	Public Housing (HLM) (1)	Private Housing (2)	p-val (1)/(2) (3)	Private Rental Housing market (4)	p-val (1)/(4) (5)
<b>Birth Country</b>					
France	78.63	88.34	0.000	86.16	0.000
Portugal	1.72	1.23	0.028	1.05	0.004
Spain	1.12	0.78	0.089	0.68	0.033
Italy	0.72	1.10	0.007	0.51	0.210
Other E.U. country	0.71	1.05	0.091	1.11	0.086
Turkey	1.24	0.30	0.000	0.51	0.000
Other European country	0.74	0.75	0.767	0.89	0.457
Maghreb	11.06	4.14	0.000	5.00	0.000
Other African country	2.66	1.06	0.000	2.17	0.327
Cambodia, Vietnam, Laos	0.69	0.41	0.016	0.51	0.227
Other countries	0.72	0.84	0.381	1.40	0.001
<b>Nationality</b>					
French at birth	82.07	91.53	0.000	88.98	0.000
French by acquisition	5.81	3.72	0.000	3.47	0.000
Portuguese	1.33	1.00	0.090	1.02	0.202
Spanish	0.62	0.31	0.004	0.44	0.265
Italian	0.47	0.43	0.865	0.20	0.012
Other E.U. nationality	0.21	0.63	0.002	0.68	0.000
Turkish	1.01	0.19	0.000	0.41	0.000
Other European nationality	0.33	0.32	0.959	0.53	0.157
Maghrebian	6.34	1.02	0.000	2.10	0.000
Other African nationality	1.50	0.38	0.000	1.03	0.096
Cambodian, Vietnamese, Laotian	0.17	0.07	0.004	0.14	0.354
Other nationalities	0.14	0.40	0.016	0.99	0.000
<b>Employment status</b>					
Employed	58.19	56.11	0.090	63.74	0.000
Unemployed	10.82	4.08	0.000	7.75	0.000
Inactive	30.99	39.81	0.000	28.51	0.004
<b>Level of education</b> (highest diploma obtained, individuals above 25 years old)					
No diploma	28.26	14.85	0.000	14.40	0.000
Lower education	50.62	48.33	0.009	37.38	0.000
Baccalaureate	9.37	12.44	0.000	16.26	0.000
Higher education	11.74	24.38	0.000	31.96	0.000
<b>Socio professional group</b>					
Farmer	0.18	1.96	0.000	0.67	0.000
Craftsman, Shopkeeper	1.50	5.03	0.000	3.99	0.000
Executive or other high position	3.64	13.03	0.000	14.14	0.000
Intermediate occupation	12.01	14.29	0.000	16.50	0.000
Employee	20.18	9.93	0.000	15.44	0.000
(Factory) Worker	31.10	16.02	0.000	20.35	0.000
<b>Age (mean)</b>	47.09	51.71	0.000	41.55	0.000
<b>Annual income (mean)</b>	12,226	18,041	0.000	15,902	0.000

Column 3 reports the p-value from a t-test for the null hypothesis that the mean of a given variable is the same for the public housing (Column 1) and private housing (Column 2) populations. Column 5 reports the p-value from a t-test for the null hypothesis that the mean of a given variable is the same for the public housing population (Column 1) and for the population of tenants in the private housing market (Column4).

**Table 1.2:** Dissatisfaction with housing conditions by income level and ethnic origin in the Public Housing sector

	<b>Dissatisfaction with Housing Conditions</b>					<b>Mean Annual Income</b> (in euros)
	Very satisfying	Satisfying	Average	Insufficient	Very insufficient	
<b>Ethnic origin</b>						
French born	13.8	44.01	28.54	9.9	3.75	12,758
Naturalized French	10.68	43.51	29.01	13.8	2.99	10,459
Other European	10.71	39.58	31.21	13.14	5.35	12,292
Maghrebian	8.42	33.27	34.83	18.21	5.26	8,603
African	7.82	20.77	41.29	25.14	4.99	7,865
Asian	0.00	60.64	11.25	28.11	0.00	12,892
<b>Mean Annual Income</b>	13,300	12,856	11,842	10,288	10,127	

**Table 1.3:** Monte-Carlo Test of Random Allocation

% departments without residential sorting relative to households' characteristics		
Household's characteristics	Public Housing	Total Sample
Nationality from Maghreb/Middle East	80.08	54.36
French Nationality at Birth	70.23	24.89

Comparison between the actual and simulated distributions by ethnic groups shares across public housing blocks (Col. 1) and across the whole sample of housing blocks (Col. 2). Percentage of *départements* where equality is not rejected at the 10 percent level using a Kolmogorov-Smirnov test.



**Table 1.4:** Ethnic diversity and dissatisfaction with housing condition

Dependent Variable:	<b>Dissatisfaction with housing condition</b>	
	Coefficient	(Std error)
<b>Ethnic Diversity</b>	0.368**	(0.129)
<b>Household characteristics:</b>		
Gender	-0.018	(0.032)
Age	-0.005***	(0.001)
Level of Education	-0.013*	(0.007)
Income (log)	-0.043	(0.031)
Household size	0.105***	(0.013)
<i>Employment status (ref: Employed)</i>		
Unemployed	0.056	(0.055)
Inactive	-0.049	(0.047)
<i>Nationality (ref: French at birth)</i>		
Naturalized French	-0.047	(0.066)
European	0.040	(0.089)
Maghrebian	-0.097	(0.067)
African	0.109	(0.143)
Asian	-0.623**	(0.311)
Other nationality	0.557	(0.633)
<b>Building characteristics:</b>		
Nb of dwellings (log)	0.020	(0.013)
<i>Date of construction of the building (ref: before 1948)</i>		
1949 ≤ t < 1974	0.017	(0.077)
1975 ≤ t < 1981	-0.094	(0.081)
1982 ≤ t < 1989	-0.109	(0.082)
1990 ≤ t < 1998	-0.423***	(0.085)
1999 ≤ t	-0.751***	(0.166)
<b>Neighborhood characteristics:</b>		
Block unemployment rate	1.003***	(0.176)
Socio-economic background (Tabard)	Yes	
Department Fixed Effects	Yes	
Intercept	4.377***	(0.394)
R-squared	0.128	
Observations	4379	

Robust standard errors adjusted for block clustering are in parentheses. \*\*\* p<0.01, \*\* p<0.05, \* p<0.1

**Table 1.5:** Descriptive statistics for each outcome in the public housing sector

	Mean	std dev	Values
<b>1. Degradation of the common areas due to vandalism</b>			
Damaging the premises	1.637	0.778	1 - 3
Graffiti	0.257	0.437	0 - 1
Garbage on the floor	0.188	0.391	0 - 1
Broken windows	0.136	0.343	0 - 1
Broken doors	0.127	0.333	0 - 1
Broken light bulbs	0.094	0.291	0 - 1
Broken mailboxes	0.154	0.361	0 - 1
Vandalism on the elevator	0.085	0.279	0 - 1
Noise in daytime	1.595	0.748	1 - 3
Noise in night time	1.374	0.627	1 - 3
<b>2. Poor quality of housing due to low maintenance</b>			
Care of the common areas	1.593	0.752	1 - 3
Condition of the outside walls	2.433	0.962	1 - 5
Cold in the apartment	0.175	0.380	0 - 1
Cold due to bad insulation	0.065	0.246	0 - 1
Cold due to breakdown in heating equipment	0.045	0.207	0 - 1
Cold due to poor equipment	0.059	0.236	0 - 1
Quality of soundproofing	1.981	0.823	1 - 3
Breakdown of the elevator	0.155	0.362	0 - 1
Toilet malfunction	0.153	0.360	0 - 1
<b>3. Public Safety</b>			
Robberies	0.095	0.293	0 - 1
Aggressions	0.081	0.273	0 - 1
Burglary (or attempt)	0.041	0.198	0 - 1

Depending on the questions, we have between 4,310 and 5,189 observations

**Table 1.6:** Diversity and Public goods: Degradation of the common areas due to vandalism (To be continued)

Outcomes:	Damaging the premises	Graffiti	Garbage on the floor	Broken windows	Broken doors	Broken light bulbs	Broken mailboxes	Vandalism on the elevator	Noise in daytime	Noise in Nighttime
<b>Ethnic Diversity</b>	0.630*** (0.127)	0.313*** (0.063)	0.298*** (0.060)	0.200*** (0.053)	0.151** (0.051)	0.271*** (0.048)	0.330*** (0.058)	0.168*** (0.047)	0.288** (0.106)	0.313*** (0.091)
<b>Household characteristics:</b>										
Gender	0.006 (0.028)	0.006 (0.015)	0.003 (0.013)	0.004 (0.011)	0.007 (0.012)	0.012 (0.010)	0.004 (0.012)	0.016* (0.010)	-0.057** (0.026)	-0.018 (0.022)
Age	-0.000 (0.001)	-0.000 (0.001)	-0.000 (0.001)	-0.000 (0.000)	-0.000 (0.000)	-0.000 (0.000)	-0.001** (0.000)	0.000 (0.000)	-0.005*** (0.001)	-0.004*** (0.001)
Level of Education	-0.001 (0.006)	-0.004 (0.003)	0.004 (0.003)	-0.001 (0.003)	-0.003 (0.003)	-0.000 (0.002)	-0.009*** (0.003)	-0.002 (0.002)	-0.008 (0.005)	-0.014** (0.005)
Income (log)	0.002 (0.028)	-0.014 (0.015)	0.003 (0.014)	0.000 (0.012)	-0.018 (0.012)	-0.006 (0.010)	-0.005 (0.012)	-0.008 (0.009)	-0.011 (0.026)	-0.034 (0.021)
Household size	0.044*** (0.010)	0.015** (0.005)	0.006 (0.005)	0.013** (0.004)	0.009** (0.004)	0.018*** (0.004)	0.014** (0.005)	0.005 (0.004)	0.031*** (0.009)	0.016** (0.008)
<i>Employment status (ref: Employed)</i>										
Unemployed	0.047 (0.045)	0.017 (0.023)	0.012 (0.021)	0.021 (0.020)	-0.009 (0.019)	-0.013 (0.016)	0.001 (0.020)	-0.011 (0.015)	-0.043 (0.039)	-0.023 (0.034)
Inactive	-0.044 (0.039)	-0.031 (0.020)	-0.020 (0.018)	-0.011 (0.017)	-0.028* (0.017)	0.012 (0.015)	0.013 (0.018)	-0.022* (0.013)	0.043 (0.037)	0.047 (0.032)
<i>Nationality (ref: French at birth)</i>										
Naturalized French	-0.088 (0.057)	-0.026 (0.030)	-0.037 (0.026)	0.007 (0.027)	0.001 (0.024)	-0.001 (0.022)	0.009 (0.027)	0.012 (0.024)	-0.019 (0.055)	-0.051 (0.044)
European	0.003 (0.078)	-0.049 (0.039)	-0.027 (0.035)	0.060* (0.035)	0.032 (0.037)	-0.009 (0.029)	0.041 (0.037)	-0.037 (0.024)	-0.024 (0.065)	0.078 (0.063)
Maghrebian	-0.179*** (0.054)	-0.054* (0.030)	-0.072** (0.026)	-0.048** (0.024)	-0.058** (0.025)	-0.062** (0.022)	-0.057** (0.027)	-0.063** (0.019)	-0.073 (0.049)	-0.076* (0.043)
African	0.019 (0.112)	-0.063 (0.064)	-0.041 (0.059)	-0.027 (0.054)	0.037 (0.059)	-0.027 (0.052)	-0.029 (0.057)	0.006 (0.051)	0.102 (0.100)	0.056 (0.090)
Asian	-0.111 (0.231)	-0.009 (0.156)	-0.258** (0.121)	0.228 (0.156)	0.002 (0.148)	0.025 (0.158)	0.134 (0.151)	0.305* (0.167)	0.070 (0.231)	0.107 (0.194)
Other nationality	-0.311 (0.306)	-0.338*** (0.073)	-0.235** (0.084)	0.023 (0.152)	0.027 (0.161)	0.093 (0.158)	0.039 (0.155)	-0.086** (0.039)	0.233 (0.306)	0.552 (0.391)

**Table 1.6:** Diversity and Public goods: Degradation of the common areas due to vandalism (Continued)

Outcomes:	Damaging the premises	Graffiti	Garbage on the floor	Broken windows	Broken doors	Broken light bulbs	Broken mailboxes	Vandalism on the elevator	Noise in daytime	Noise in Nighttime
<b>Building characteristics:</b>										
Nb of dwellings (log)	0.105*** (0.018)	0.068*** (0.006)	0.055*** (0.005)	0.036*** (0.005)	0.037*** (0.004)	0.029*** (0.004)	0.047*** (0.004)	0.040*** (0.004)	.027** (0.010)	0.025** (0.008)
<i>Date of construction of the building (ref: before 1948)</i>										
1949 ≤ t < 1974	0.253*** (0.061)	0.115*** (0.025)	0.081*** (0.022)	0.054** (0.020)	0.058** (0.021)	0.017 (0.015)	0.047** (0.020)	0.023 (0.017)	0.041 (0.059)	-0.024 (0.051)
1975 ≤ t < 1981	0.277*** (0.070)	0.118*** (0.028)	0.088*** (0.026)	0.053*** (0.023)	0.049** (0.024)	0.034* (0.018)	0.043** (0.022)	0.038** (0.019)	-0.079 (0.062)	-0.059 (0.053)
1982 ≤ t < 1989	0.357*** (0.073)	0.134*** (0.029)	0.081** (0.026)	0.060** (0.023)	0.061** (0.025)	0.051** (0.020)	0.087*** (0.024)	0.053** (0.021)	-0.106* (0.063)	-0.048 (0.055)
1990 ≤ t < 1998	0.261*** (0.071)	0.093*** (0.027)	0.066** (0.025)	0.038* (0.021)	0.047** (0.023)	0.009 (0.017)	0.049** (0.021)	0.025 (0.018)	-0.171** (0.063)	-0.105* (0.055)
1999 ≤ t	0.210 (0.154)	0.073 (0.067)	0.039 (0.075)	0.113 (0.082)	0.071 (0.067)	-0.037 (0.028)	-0.039 (0.034)	-0.034 (0.030)	-0.321** (0.140)	-0.341*** (0.066)
<b>Neighborhood characteristics:</b>										
Block unempl. rate	1.063*** (0.166)	0.351*** (0.068)	0.415*** (0.072)	0.207*** (0.058)	0.230*** (0.055)	0.263*** (0.045)	0.294*** (0.063)	0.100* (0.056)	0.478*** (0.129)	0.414*** (0.112)
Socio-economic background (Tabard)	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Department	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Fixed Effects										
Intercept	0.588** (0.298)	-0.034 (0.155)	-0.186 (0.140)	-0.132 (0.126)	0.085 (0.128)	-0.095 (0.105)	-0.096 (0.128)	-0.100 (0.093)	1.899*** (0.269)	1.852*** (0.222)
R-squared	0.136	0.157	0.134	0.091	0.072	0.086	0.129	0.093	0.091	0.081
Observations	3661	4379	4379	4379	4379	4379	4379	4379	4379	4379

Each column corresponds to a different regression, for each dependent variable, as reported in the first line. Robust standard errors adjusted for block clustering are in parentheses. \*\*\* p<0.01, \*\* p<0.05, \* p<0.1

**Table 1.7:** Diversity and Public goods: Poor Quality of Housing due to low Maintenance (To be continued)

Outcomes:	Care of the common areas	Condition of the outside walls	Cold in the apartment	Cold due to bad insulation	Cold due to breakdown of heating equipment	Cold due to poor heating equipment	Quality of soundproofing	Breakdown of the elevator	Toilet malfunction
<b>Ethnic Diversity</b>	0.384** (0.121)	0.260* (0.145)	0.136** (0.059)	0.021 (0.040)	0.007 (0.037)	0.084** (0.036)	0.021 (0.110)	0.106** (0.051)	0.133** (0.051)
<b>Household characteristics:</b>									
Gender	-0.027 (0.028)	-0.030 (0.032)	-0.029** (0.014)	-0.013 (0.009)	-0.019** (0.008)	-0.022** (0.009)	-0.009 (0.028)	-0.002 (0.012)	-0.017 (0.013)
Age	-0.003** (0.001)	-0.003** (0.001)	-0.000 (0.001)	-0.000 (0.000)	-0.000 (0.000)	-0.000 (0.000)	-0.006*** (0.001)	0.001* (0.000)	-0.001** (0.000)
Level of Education	-0.000 (0.006)	-0.010 (0.007)	-0.000 (0.003)	-0.000 (0.002)	0.001 (0.002)	-0.001 (0.002)	-0.007 (0.006)	-0.000 (0.003)	0.001 (0.003)
Income (log)	0.051* (0.029)	0.024 (0.032)	-0.009 (0.014)	-0.004 (0.008)	0.009 (0.008)	-0.010 (0.008)	0.004 (0.029)	0.011 (0.012)	-0.022 (0.013)
Household size	0.051*** (0.011)	0.041*** (0.011)	0.013** (0.005)	0.009** (0.003)	0.004 (0.003)	0.006* (0.003)	0.027** (0.010)	0.013** (0.004)	0.020*** (0.005)
<i>Employment status (ref: Employed)</i>									
Unemployed	0.041 (0.044)	-0.030 (0.049)	0.008 (0.022)	0.006 (0.014)	-0.003 (0.011)	0.012 (0.015)	-0.019 (0.042)	0.000 (0.019)	0.021 (0.021)
Inactive	0.026 (0.039)	0.026 (0.043)	-0.015 (0.019)	0.005 (0.013)	-0.001 (0.010)	-0.004 (0.012)	-0.011 (0.038)	-0.014 (0.016)	0.004 (0.019)
<i>Nationality (ref: French at birth)</i>									
Naturalized French	-0.060 (0.056)	-0.075 (0.059)	0.046 (0.029)	0.010 (0.018)	0.016 (0.017)	0.020 (0.020)	-0.027 (0.060)	-0.047** (0.023)	-0.033 (0.025)
European	0.155** (0.079)	-0.094 (0.087)	-0.022 (0.036)	0.007 (0.025)	-0.002 (0.020)	0.006 (0.024)	0.037 (0.072)	-0.026 (0.034)	-0.021 (0.033)
Maghrebian	-0.080 (0.054)	-0.247*** (0.061)	0.068** (0.029)	0.002 (0.020)	0.016 (0.016)	0.051** (0.021)	-0.170** (0.052)	0.002 (0.025)	-0.061** (0.025)
African	0.122 (0.101)	-0.033 (0.116)	0.171** (0.063)	-0.001 (0.040)	0.055 (0.038)	0.035 (0.043)	0.068 (0.094)	0.016 (0.056)	0.071 (0.059)
Asian	0.162 (0.343)	0.033 (0.301)	-0.241*** (0.033)	-0.104*** (0.021)	-0.072*** (0.021)	-0.094*** (0.020)	0.172 (0.223)	0.088 (0.147)	0.057 (0.143)
Other nationality	0.177 (0.329)	0.487 (0.393)	-0.164*** (0.041)	-0.082*** (0.022)	-0.030 (0.027)	-0.050** (0.018)	0.038 (0.440)	-0.143** (0.060)	0.104 (0.185)

Robust standard errors adjusted for block clustering are in parentheses. \*\*\* p<0.01, \*\* p<0.05, \* p<0.1

**Table 1.7: Diversity and Public goods: Poor Quality of Housing due to low Maintenance (Continued)**

Outcomes:	Care of the common areas	Condition of the outside walls	Cold in the apartment	Cold due to bad insulation	Cold due to breakdown of heating equipment	Cold due to poor heating equipment	Quality of soundproofing	Breakdown of the elevator	Toilet malfunction
<b>Building characteristics:</b>									
Nb of dwellings (log)	-0.026 (0.016)	-0.025* (0.015)	0.003 (0.005)	0.001 (0.003)	-0.000 (0.003)	-0.001 (0.003)	0.042*** (0.011)	0.069*** (0.005)	-0.007 (0.005)
<i>Date of construction of the building (ref: before 1948)</i>									
1949 ≤ t < 1974	0.049 (0.070)	0.041 (0.077)	0.017 (0.027)	-0.003 (0.019)	0.004 (0.016)	0.029** (0.014)	-0.004 (0.063)	0.028 (0.022)	-0.004 (0.027)
1975 ≤ t < 1981	-0.005 (0.075)	0.138* (0.081)	0.046 (0.029)	0.013 (0.020)	0.006 (0.017)	0.036*** (0.016)	-0.161** (0.068)	0.097*** (0.025)	0.007 (0.029)
1982 ≤ t < 1989	0.030 (0.079)	0.048 (0.084)	0.004 (0.030)	-0.011 (0.021)	0.002 (0.018)	0.022 (0.016)	-0.388*** (0.067)	0.061** (0.026)	-0.050* (0.030)
1990 ≤ t < 1998	-0.055 (0.080)	-0.439*** (0.082)	-0.050* (0.029)	-0.041** (0.020)	-0.001 (0.018)	-0.003 (0.014)	-0.574*** (0.068)	0.046* (0.026)	-0.089** (0.030)
1999 ≤ t	0.039 (0.162)	-1.055*** (0.152)	-0.078 (0.057)	-0.071** (0.024)	0.026 (0.049)	-0.009 (0.020)	-0.849*** (0.150)	-0.081** (0.039)	-0.009 (0.094)
<b>Neighborhood characteristics:</b>									
Block unempl. rate	1.044*** (0.158)	0.827*** (0.166)	0.216*** (0.065)	0.104** (0.042)	0.054 (0.046)	0.063* (0.038)	0.641*** (0.126)	0.074 (0.057)	0.084 (0.060)
Socio-economic background (Tabard)	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Department	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Fixed Effects									
Intercept	0.812** (0.296)	2.321*** (0.330)	0.226 (0.142)	0.047 (0.084)	0.033 (0.083)	0.074 (0.077)	1.939*** (0.296)	-0.312** (0.127)	0.332** (0.138)
R-squared	0.085	0.097	0.040	0.018	0.008	0.023	0.132	0.151	0.040
Observations	3661	4379	4379	4379	4379	4379	4379	4379	4379

Each column corresponds to a different regression, for each dependent variable, as reported in the first line. Robust standard errors adjusted for block clustering are in parentheses.  
 \*\*\* p<0.01, \*\* p<0.05, \* p<0.1

**Table 1.8:** Diversity and Public goods: Public safety in the neighborhood (To be continued)

Outcomes:	Robberies	Aggressions	Burglary (or attempt)
<b>Ethnic Diversity</b>	0.043 (0.039)	-0.024 (0.038)	-0.001 (0.027)
<b>Household characteristics:</b>			
Gender	-0.002 (0.010)	-0.012 (0.011)	-0.010 (0.007)
Age	-0.000 (0.000)	-0.001* (0.000)	-0.000 (0.000)
Level of Education	0.005** (0.002)	0.006** (0.002)	-0.001 (0.002)
Income (log)	-0.004 (0.010)	-0.005 (0.010)	-0.004 (0.006)
Household size	0.010** (0.004)	0.012** (0.004)	-0.001 (0.003)
<i>Employment status (ref: Employed)</i>			
Unemployed	0.008 (0.016)	0.037** (0.017)	0.009 (0.011)
Inactive	-0.002 (0.014)	0.008 (0.015)	0.008 (0.010)
<i>Nationality (ref: French at birth)</i>			
Naturalized French	-0.007 (0.019)	-0.024 (0.018)	0.024 (0.017)
European	-0.025 (0.026)	-0.018 (0.024)	0.025 (0.021)
Maghrebian	-0.043** (0.019)	-0.034* (0.020)	0.008 (0.015)
African	-0.029 (0.041)	-0.056 (0.037)	0.037 (0.040)
Asian	0.005 (0.129)	0.016 (0.113)	-0.036** (0.014)
Other nationality	-0.075** (0.027)	-0.106*** (0.031)	-0.051** (0.018)

Robust standard errors adjusted for block clustering are in parentheses.

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

**Table 1.8:** Diversity and Public goods: Public safety in the neighborhood (Continued)

Outcomes:	Robberies	Aggressions	Burglary (or attempt)
<b>Building characteristics:</b>			
Nb of dwellings (log)	-0.000 (0.004)	0.004 (0.004)	-0.002 (0.003)
<i>Date of construction of the building (ref: before 1948)</i>			
1949 ≤ t < 1974	0.050** (0.019)	0.034* (0.020)	0.012 (0.015)
1975 ≤ t < 1981	0.071** (0.022)	0.020 (0.021)	0.019 (0.015)
1982 ≤ t < 1989	0.071*** (0.021)	0.047** (0.022)	0.011 (0.015)
1990 ≤ t < 1998	0.071** (0.022)	0.022 (0.022)	0.011 (0.015)
1999 ≤ t	0.027 (0.052)	-0.033 (0.025)	0.076 (0.064)
<b>Neighborhood characteristics:</b>			
Block unempl. rate	0.019 (0.046)	0.200*** (0.051)	0.042 (0.031)
Socio-economic background (Tabard)	Yes	Yes	Yes
Department	Yes	Yes	Yes
Fixed Effects			
Intercept	0.007 (0.097)	0.006 (0.098)	0.092 (0.069)
R-squared	0.020	0.024	0.002
Observations	4379	4379	4379

Each column corresponds to a different regression, for each dependent variable, as reported in the first line. Robust standard errors adjusted for block clustering are in parentheses. \*\*\* p<0.01, \*\* p<0.05, \* p<0.1



Table 1.9: Type of repairs done

	<i>Vandalism</i> Work in the commons: staircase, doors, lights, glass (1)	<i>Poor Housing Quality</i> Major works: façade, heating, elevator, toilets (2)	<i>Public Safety</i> Security Equipment: entry code, locks (3)
Ethnic Diversity	0.134** (0.054)	-0.213*** (0.069)	0.141*** (0.041)
R-squared	0.024	0.012	0.010
N	2220	2220	2220

## Correlation (in %) with perception of degradations

**1. Vandalism**

Damaging the premises	5.79***	
Graffiti	16.94***	
Garbage on the floor	14.35***	
Broken windows	11.86***	
Broken doors	13.74***	
Broken light bulbs	12.24***	
Broken mailboxes	13.10***	
Vandalism on the elevator	13.54***	12.72***

**2. Poor Housing Quality**

Condition of the outside walls		-3.46***
Cold in the apartment		3.89***
Cold due to bad insulation		1.29
Cold due to breakdown in heating equipment		4.70***
Cold due to poor equipment		0.012
Breakdown of the elevator		-0.001
Toilet malfunction		4.31***

**3. Public Safety**

Robberies		2.52***
Aggressions		4.15***
Burglary (or attempt)		2.35***

Robust standard errors adjusted for block clustering are in parentheses \*\*\* p<0.01, \*\* p<0.05, \* p<0.1

**Table 1.10:** Are results driven by some major ethnic groups disliking being around foreigners? (to be continued)

	Dissatisfaction with Housing Conditions		
	(1)	(2)	(3)
<b>Ethnic Diversity</b>	0.368** (0.129)	0.359** (0.141)	0.313** (0.154)
<b>Household characteristics:</b>			
Gender	-0.018 (0.032)	-0.018 (0.033)	-0.018 (0.032)
Age	-0.005*** (0.001)	-0.005*** (0.001)	-0.005*** (0.001)
Education	-0.013* (0.007)	-0.013* (0.007)	-0.012* (0.007)
Income (log)	-0.043 (0.031)	-0.045 (0.031)	-0.038 (0.031)
Household size	0.105*** (0.013)	0.105*** (0.013)	0.094*** (0.015)
<i>Employment status(ref: Employed)</i>			
Unemployed	0.056 (0.055)	0.057 (0.055)	0.055 (0.055)
Inactive	-0.049 (0.047)	-0.049 (0.047)	-0.051 (0.047)
<i>Nationality (ref: French at birth)</i>			
Naturalized French	-0.047 (0.066)	-0.030 (0.144)	
European	0.040 (0.089)	0.043 (0.197)	
Maghrebian	-0.097 (0.067)	-0.095 (0.186)	
African	0.109 (0.143)	-0.239 (0.324)	
Asian	-0.623** (0.311)	-0.150 (0.814)	
Other nationality	0.557 (0.633)	0.788 (1.203)	
<i>Interaction terms: ELF * origin</i>			
ELF * naturalized French		-0.052 (0.373)	
ELF * European		-0.007 (0.542)	
ELF * Maghrebian		-0.001 (0.401)	
ELF * African		0.894 (0.751)	
ELF * Asian		-1.123 (1.340)	
ELF * Other nationality		-1.136 (4.552)	

Each coefficient is estimated from a separate regression, according to equation 1.2. Robust standard errors adjusted for block clustering are in parentheses. \*\*\* p<0.01, \*\* p<0.05, \* p<0.1

Table 1.10: Continued

Dependent Variable:	Dissatisfaction with housing condition		
	(1)	(2)	(3)
<b>Major groups in HLM: Native French and Maghrebians</b>			
Native French household			-0.043 (0.056)
ELF * Native French household			0.027 (0.169)
Maghrebian household			0.138 (0.259)
ELF * Maghrebian household			-0.300 (0.559)
<b>Building characteristics:</b>			
Nb of dwellings (log)	0.020 (0.013)	0.020 (0.013)	0.019 (0.013)
<i>Date of construction of the building (ref: before 1948)</i>			
1949 ≤ t < 1974	0.017 (0.077)	0.017 (0.077)	0.022 (0.077)
1975 ≤ t < 1981	-0.094 (0.081)	-0.094 (0.082)	-0.087 (0.082)
1982 ≤ t < 1989	-0.109 (0.082)	-0.109 (0.082)	-0.104 (0.082)
1990 ≤ t < 1998	-0.423*** (0.085)	-0.424*** (0.086)	-0.421*** (0.086)
1999 ≤ t	-0.751*** (0.166)	-0.753*** (0.166)	-0.748*** (0.165)
<b>Neighborhood characteristics:</b>			
Block unemployment rate	1.003*** (0.176)	0.999*** (0.176)	0.995*** (0.175)
Socio-economic background (Tabard)	Yes	Yes	Yes
Department	Yes	Yes	Yes
Fixed Effects			
Intercept	3.991*** (0.360)	4.010*** (0.363)	4.004*** (0.353)
R-squared	0.128	0.127	0.127
Observations	4379	4379	4379

Each coefficient is estimated from a separate regression, according to equation 1.2. Robust standard errors adjusted for block clustering are in parentheses \*\*\* p<0.01, \*\* p<0.05, \*p<0.1

**Table 1.11:** Variation in Perception of Housing quality: Within and Between Public Housing Blocks

	Dissatisfaction with Housing conditions	
	Within correlation	Between correlation
Gender	0.01 (0.04)	-0.03 (0.04)
Age	-0.00* (0.00)	-0.01*** (0.00)
Education	0.01 (0.01)	-0.03*** (0.01)
Income (log)	-0.07* (0.04)	-0.09** (0.04)
Unemployed	0.05 (0.07)	0.08 (0.06)
Inactive	-0.08 (0.06)	0.02 (0.06)
Household size	0.09*** (0.02)	0.13*** (0.01)
Naturalized French	-0.11 (0.08)	-0.00 (0.09)
European	-0.00 (0.11)	0.19* (0.11)
Maghrebian	-0.09 (0.09)	-0.01 (0.08)
African	-0.05 (0.15)	0.44** (0.20)
Asian	-0.34 (0.52)	-0.56 (0.53)
Other nationality	-0.11 (0.79)	0.14 (0.53)
Building size (log)		0.05*** (0.01)
Housing Project FE	Yes	No
R-squared	0.056	0.172
Observations	5105	5105

Robust standard errors adjusted for block clustering are in parentheses  
\*\*\*p<0.01, \*\* p<0.05, \* p<0.1

Table 1.12: Ethnic Diversity and Distant public goods

	Maintenance of streets	Accessibility to public transports	Accessibility to private transports
Ethnic Diversity	0.158 (0.096)	-0.050 (0.134)	0.142 (0.102)
Gender	0.001 (0.022)	-0.036 (0.028)	0.015 (0.024)
Age	-0.000 (0.001)	-0.002** (0.001)	-0.003** (0.001)
Education	-0.004 (0.005)	0.001 (0.006)	0.006 (0.005)
Income (log)	0.013 (0.024)	0.010 (0.030)	0.038* (0.023)
Unemployed	0.012 (0.036)	0.059 (0.041)	-0.027 (0.037)
Inactive	0.039 (0.032)	0.104** (0.039)	0.008 (0.032)
Household size	0.018** (0.008)	0.002 (0.010)	0.011 (0.009)
Naturalized French	-0.096** (0.045)	-0.009 (0.055)	0.030 (0.048)
European	0.040 (0.070)	0.023 (0.067)	-0.099* (0.060)
Maghrebien	-0.057 (0.045)	-0.005 (0.051)	-0.069 (0.046)
African	0.037 (0.095)	0.119 (0.097)	-0.032 (0.101)
Asian	-0.105 (0.183)	-0.128 (0.282)	0.176 (0.246)
Other nationality	-0.330*** (0.073)	0.112 (0.343)	-0.376*** (0.083)
Block unemployment rate	0.386*** (0.114)	-0.102 (0.144)	-0.041 (0.112)
Intercept	0.814** (0.276)	4.365*** (0.471)	0.783*** (0.233)
Department Fixed Effects	Yes	Yes	Yes
Socio economic backgrounds	Yes	Yes	Yes
R-squared	0.053	0.283	0.101
N	4451	4451	4451

Robust standard errors adjusted for block clustering are in parentheses. \*\*\* p<0.01, \*\* p<0.05, \* p<0.1

**Table 1.13:** Ethnic Diversity and Dissatisfaction with Housing conditions – Robustness Checks (to be continued)

	Dissatisfaction with Housing Conditions		
	(1)	(2)	(3)
Ethnic Diversity	1.392** (0.530)		
<b>Household characteristics:</b>			
Gender	-0.018 (0.032)	-0.019 (0.032)	-0.021 (0.032)
Age	-0.005*** (0.001)	-0.005*** (0.001)	-0.005*** (0.001)
Education	-0.013* (0.007)	-0.014* (0.007)	-0.014* (0.007)
Income (log)	-0.041 (0.031)	-0.041 (0.031)	-0.040 (0.031)
Household size	0.105*** (0.013)	0.104*** (0.013)	0.105*** (0.013)
<i>Employment status (ref: Employed)</i>			
Unemployed	0.050 (0.055)	0.059 (0.055)	0.051 (0.055)
Inactive	-0.050 (0.047)	-0.049 (0.047)	-0.051 (0.047)
<i>Nationality (ref: French at birth)</i>			
Naturalized French	-0.044 (0.066)	-0.043 (0.066)	-0.036 (0.066)
European	0.059 (0.088)	0.058 (0.089)	0.062 (0.089)
Maghrebien	-0.107 (0.068)	-0.106 (0.068)	-0.108 (0.068)
African	0.092 (0.143)	0.098 (0.143)	0.097 (0.143)
Asian	-0.615* (0.316)	-0.607* (0.318)	-0.605* (0.315)
Other nationality	0.577 (0.630)	0.565 (0.632)	0.569 (0.634)
<b>Building characteristics:</b>			
Nb of dwellings (log)	0.017 (0.013)	0.020 (0.013)	0.019 (0.013)
<i>Date of construction of the building (ref: before 1948)</i>			
1949 ≤ t < 1974	0.007 (0.077)	0.022 (0.077)	0.012 (0.076)
1975 ≤ t < 1981	-0.115 (0.082)	-0.101 (0.081)	-0.110 (0.081)
1982 ≤ t < 1989	-0.134 (0.082)	-0.122 (0.082)	-0.129 (0.082)
1990 ≤ t < 1998	-0.444*** (0.085)	-0.438*** (0.086)	-0.439*** (0.085)
1999 ≤ t	-0.769*** (0.168)	-0.770*** (0.168)	-0.751*** (0.170)

Robust standard errors adjusted for block clustering are in parentheses.  
 \*\*\*p<0.01, \*\* p<0.05, \* p<0.1

Table 1.13: Continued

Dependent Variable:	Dissatisfaction with housing condition		
	(1)	(2)	(3)
<b>Neighborhood characteristics:</b>			
Block unemployment rate	0.909*** (0.175)	0.954*** (0.175)	0.911*** (0.174)
Socio-economic background (Tabard)	Yes	Yes	Yes
Department	Yes	Yes	Yes
Fixed Effects			
<b><i>Ethnic group shares:</i></b>			
% European	-3.249** (1.007)	-0.928* (0.493)	-2.273** (0.961)
% Maghrebian	-1.092 (0.707)	0.633** (0.277)	2.014*** (0.521)
% African	-1.203 (1.101)	0.941 (0.823)	1.218 (1.428)
% Asian	-2.186* (1.323)	0.167 (0.974)	0.315 (1.838)
% Other nationality	-1.488* (0.855)	0.260 (0.501)	-0.263 (0.864)
<b><i>Squared ethnic group shares:</i></b>			
(% European) <sup>2</sup>			7.587* (4.332)
(% Maghrebian) <sup>2</sup>			-3.767** (1.220)
(% African) <sup>2</sup>			-4.225 (7.839)
(% Asian) <sup>2</sup>			-4.792 (14.523)
(% Other nationality) <sup>2</sup>			1.351 (2.895)
Intercept	4.009*** (0.361)	4.071*** (0.362)	4.023*** (0.363)
R-squared	0.131	0.129	0.131
Observations	4379	4379	4379

Robust standard errors adjusted for block clustering are in parentheses \*\*\* p<0.01, \*\* p<0.05, \* p<0.1

**Table 1.14:** A proxy for language fractionalization

Dependent Variable:	Dissatisfaction with housing conditions (1)	Neglect of the commons (2)	Quality of housing (3)	Insecurity (4)
<i>Panel A:</i>				
Diversity based on nationality at birth	0.368*** (0.129)	1.532*** (0.422)	0.727*** (0.263)	0.0252 (0.183)
Observations	4,379	1,693	3,797	4,379
R-squared	0.156	0.201	0.192	0.063
<i>Panel B:</i>				
Diversity based on share of the population speaking French in country of origin	0.366*** (0.141)	1.560*** (0.472)	0.741** (0.290)	-0.0377 (0.199)
Observations	4,365	1,689	3,788	4,365
R-squared	0.157	0.202	0.193	0.064

Standard errors in parentheses \*\*\* p<0.01, \*\* p<0.05, \* p<0.1. Controls: head of household characteristics (gender, age, education, activity status, aggregated nationality), household characteristics (log of income, household size), building characteristics (date of construction, log number of housing units), socio-economic background of the neighborhood (unemployment rate, Nicole Tabbarð classification), department fixed-effects. Each coefficient comes from a separate regression. The columns indicate the four dependent variables under study. Panel A and panel B respectively correspond to the measure of diversity used in each regression.



**Table 1.15:** Correlation between new inhabitants' nationality and share of the area population of the same nationality

Dep Var: Share of population of the same ethnic group as new movers in a given area		
	(1)	(2)
Share of the department population of the the same ethnic group		0.890*** (0.034)
<b>Nationality</b> (reference group: Other nationalities)		
Native	0.816*** (0.010)	0.067** (0.030)
Naturalized French	-0.044*** (0.012)	-0.037** (0.012)
European	-0.022* (0.012)	-0.007 (0.011)
Maghrebian	0.001 (0.014)	0.007 (0.013)
African	-0.025 (0.015)	-0.016 (0.015)
Asian	-0.085* (0.049)	-0.019 (0.047)
<b>Public Housing (HLM)</b>	-0.158** (0.068)	-0.184** (0.066)
<b>Nationality * HLM</b>		
HLM * Native	-0.038 (0.034)	0.040 (0.033)
HLM * Naturalized	0.095** (0.037)	0.051 (0.036)
HLM * European	-0.018 (0.038)	0.010 (0.037)
HLM * Maghrebian	0.055 (0.037)	0.024 (0.036)
HLM * African	0.046 (0.040)	0.017 (0.039)
HLM * Asian	0.083 (0.080)	0.027 (0.078)
<b>Household characteristics</b>		
Age	0.002** (0.001)	0.002** (0.001)
Age squared	-0.000** (0.000)	-0.000** (0.000)
Hourly wage (log)	-0.001 (0.002)	-0.001 (0.002)
Education	Yes	Yes
HLM * Education	Yes	Yes
Socio-Econ Category	Yes	Yes
HLM * SEC	Yes	Yes
Department Fixed Effect	Yes	Yes
Intercept	0.085** (0.036)	0.070** (0.035)
R-squared	0.856	0.864
N	11519	11519

\* p<0.10, \*\* p<0.05, \*\*\* p<0.001

**Table 1.16:** Rejection of HLM offers and Ethnic diversity

Rows: Dependent Variables	Coefficient associated with Ethnic Diversity			
	(1)	(2)	(3)	(4)
<b>Panel A:</b> Sample of households who currently live in public housing:				
I. Probability of having declined at least one HLM offer during the previous application process	0.058 (0.058)	0.069 (0.063)	0.017 (0.067)	0.123 (0.0886)
N	1,779	1,779	1,748	1,744
R <sup>2</sup>	0.001	0.021	0.023	0.089
II. Probability that the reason for having declined an HLM offer during the previous application was "unpleasant environment"	0.162 (0.144)	0.061 (0.158)	0.017 (0.171)	-0.0310 (0.258)
N	417	417	415	414
R <sup>2</sup>	0.003	0.035	0.050	0.308
<b>Panel B:</b> Sample of households who are currently applying to public housing:				
I. Probability of having declined at least one HLM offer during the current application process	-0.063 (0.057)	-0.043 (0.064)	-0.088 (0.071)	-0.116 (0.103)
N	1,192	1,192	1,173	1,171
R <sup>2</sup>	0.001	0.014	0.024	0.121
II. Probability that the reason for having declined an HLM offer during the current application was "unpleasant environment"	0.004 (0.194)	-0.007 (0.237)	-0.104 (0.250)	-0.122 (0.506)
N	198	198	195	194
R <sup>2</sup>	0.000	0.083	0.115	0.590

Each of the coefficients is estimated from a separate regression of each of the four dependent variables described in the first column on ethnic diversity. Column 1 does not include any control. Column 2 includes households characteristics (gender, age, education, employment status and nationality of the head of household, total income (in log) of the household per unit of consumption, and household size). Column 3 adds up the characteristics of the building (number of apartments (in log) and construction date). On top of that, column 4 includes neighborhood characteristics (socio-economic background (Tabard index), and local unemployment rate), as well as department fixed effects. In addition, a dummy variable indicating whether the household already lives in the public housing sector is included in specifications 2 to 4 of Panel B. The coefficients for all the controls are available from authors upon request. Robust standard errors adjusted for block clustering are in parentheses. \*\*\* p<0.01, \*\* p<0.05, \* p<0.1

**Table 1.17:** Do households having left their previous housing due to an “unpleasant environment” now live in less diverse neighborhoods?

Dependent Variable:	Main reason for leaving previous housing: unpleasant environment (noise, lifestyle or insecurity)			
	Households who moved <b>within</b> the		Households who moved <b>toward</b> the	
	Private Housing sector	Public Housing sector	Private Housing sector	Public Housing sector
	(1)	(2)	(3)	(4)
Ethnic Diversity	-0.073** (0.030)	0.083 (0.140)	-0.061* (0.032)	0.016 (0.052)
Observations	5,955	627	6,560	1,793
R-squared	0.030	0.207	0.031	0.079

In each regression, we control for household characteristics (gender, age, education, income (in log), employment status, nationality, household size), building characteristics (number of apartments and construction date), neighborhood characteristics (socio-economic background (Tabard index), and local unemployment rate) and department fixed effects. The coefficients for all the controls are available from authors upon request. Robust standard errors adjusted for block clustering are in parentheses. \*\*\*  $p < 0.01$ , \*\*  $p < 0.05$ , \*  $p < 0.1$

## 1.A Fractionalization index and French blocks

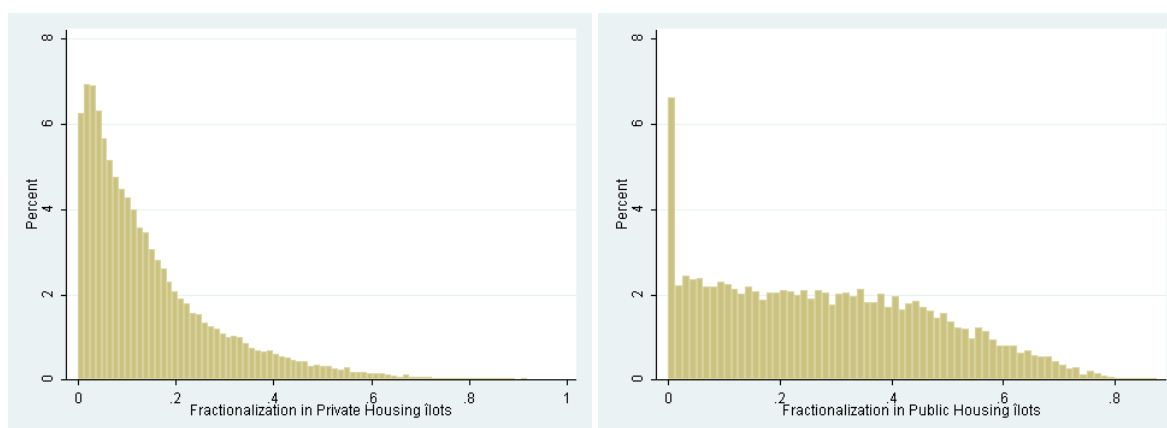
**Table 1.18:** Housing blocks in the Census 1999

	Number of blocks sampled per department		Number of individuals sampled per block	
	All France	HLM Population	All France	HLM Population
Mean	2,894.5	932.9	24.6	18.4
Median	2,236	740.5	15	8

**Table 1.19:** Fractionalization by nationality at birth in housing blocks

	1999 Census			2002 Housing Survey	
	Whole France	Private Housing	HLM Population	Whole France	HLM Population
Mean	16.65	14.29	27.68	16.23	25.33
Median	11.82	10.29	25.18	11.98	23.37
Std Dev	15.33	13.36	18.75	14.2	17.94
Minimum	0	0	0	0	0
Maximum	91.83	91.83	91.3	84.94	80.26
N	6,643,287	5,027,235	1,616,052	28,744	4,465

**Figure 1.1:** Fractionalization by nationality at birth within private and public housing blocks, Census 1999



## 1.B Principal component analysis

To decompose the various effects of fractionalization, we alternatively run an exploratory analysis to extract the main dimensions with which the various questions reported in the HS correlate the most. We then interpret those factors as different dimensions of the quality of public spaces that could be affected by ethnic diversity. The principal component analysis lets the correlation patterns among the various questions emerge endogenously from the data, rather than grouping them in an arbitrary way. We select (following the Kaiser criterion) three main factors with eigenvalues higher than one that emerged from the principal component analysis of the relevant survey questions. Table 1.20 reports those three factors and the rotated matrix of correlations between those factors and each question. Three main patterns of correlation emerge that refer broadly speaking to three dimensions of the quality of the public space. Table 1.5 reports descriptive statistics of the various questions. For each variable, a lower value represents a better outcome (e.g. greater care of the commons, less graffiti, better soundproofing...).

**Table 1.20:** Principal component analysis

	Factor 1	Factor 2	Factor 3
Care of the commons	0.142	0.225	0.053
Voluntary degradations of the commons	0.675	0.172	0.091
Graffiti on the walls	0.209	-0.126	0.023
Trash in the commons	0.247	0.047	0.085
Broken doors in the commons	0.591	0.166	0.123
Broken lights in the commons	0.564	0.072	0.031
Degradation of mail boxes	0.528	0.130	0.034
Broken elevators	0.528	-0.031	-0.041
Quality of the building's façade	-0.038	0.239	-0.030
Problem with heating in the building	-0.007	0.336	0.029
Quality of soundproofing	0.042	0.703	0.004
Noise disturbance during the day in the housing	0.060	0.831	0.052
Noise disturbance at night in the housing	0.113	0.807	0.105
Victim or witness of aggression in the neighborhood	0.098	0.136	0.746
Victim or witness of robbery in the neighborhood	-0.006	0.028	0.810

We then create summary indices from the three groups of questions identified in the previous section. We run a principal component analysis on each group of questions, and take the first principal component of each. We refer to those indices as "Neglect of the public areas", "Quality of housing" and "Public Safety". The higher the indices, the more unfavorable are the outcomes. We also check the robustness of the results by looking at alternative summary indices, taking the sum of the questions belonging to each group, or performing a mean effect analysis for each group. The estimates for these alternative

indices are reported in Appendix C, yielding similar results.

Table 1.21 shows the effect of ethnic fractionalization on those three different dimensions: "Neglect of the public areas", "Quality of housing" and "Public Safety".<sup>44</sup> For each index, we run separate regressions on ethnic diversity controlling for the usual household, building and neighborhood characteristics as specified at the bottom of each column in Table 1.21. As can be surmised from an examination of three sets of regressions on Table 1.21, the results relying on categories derived from the principal component analysis rather than the categories that followed from the theoretical literature on public goods, and relying on identical model specifications, are basically similar. For the effect of ethnic diversity on the synthetic index Neglect of Public Areas (see the first panel of Table 1.21), the effect of ethnic diversity is always statistically significant at the 1 percent level, and is substantively sizeable. For the index of Poor Housing Quality, the coefficient associated with ethnic diversity is statistically significant at the 1 percent level for the first two specifications but only at the 5 percent level when all the controls are included. However, as with the results using the theoretically inferred categorization in the main body of the paper, its effect is much lower than for the index for voluntary degradations. For the index of Public Safety, the data here show that ethnic diversity does not have a statistically significant effect on public safety in the public housing sector in our full specification, as is the case in our main results. In sum, categorization by principal components analysis yields quite similar results as to those reported with the theoretically derived categorization.

---

<sup>44</sup>The coefficient estimates for the control variables are not reported here but are very similar to those reported in Table 1.4. The full regression results are available upon request.

**Table 1.21:** Ethnic Diversity and Public Goods (public housing)

	Ethnic Diversity		
	(1)	(2)	(3)
<b>Index for Neglect of Public Areas</b>			
	1.791*** (0.322)	1.514*** (0.330)	1.532*** (0.422)
Observations	1,700	1,693	1,693
R-squared	0.060	0.084	0.134
<b>Index for Poor Quality of Housing</b>			
	2.132*** (0.187)	1.382*** (0.200)	0.727** (0.263)
Observations	3,869	3,805	3,797
R-squared	0.092	0.128	0.161
<b>Index for Public Safety</b>			
	0.330** (0.129)	0.273* (0.139)	0.025 (0.183)
Observations	4,464	4,388	4,379
R-squared	0.014	0.014	0.033
Household characteristics	Yes	Yes	Yes
Building characteristics	No	Yes	Yes
Neighborhood characteristics	No	No	Yes
Department fixed effects	No	No	Yes

Each coefficient is estimated from a separate regression, according to equation 1.2. The four dependent variables considered include the answer to the general opinion / dissatisfaction question and the three indices that were derived from principal component analysis as described in section 4.2. Each index is regressed on either ethnic diversity, controlling for the usual household and neighborhood characteristics unless otherwise indicated. Robust standard errors adjusted for block clustering are in parentheses. \*\*\*  $p < 0.01$ , \*\*  $p < 0.05$ , \*  $p < 0.1$

## 1.C Alternative summary indices - Mean effect analysis

As a robustness check, we have also experimented with alternative indexes for measuring these three dimensions of housing conditions. We have first looked at basic indices defined as the sum of the outcome variables related to each dimension. For each of the three dimensions considered, we thus obtain a variable which increases with the number of adverse outcomes one faces. Table 1.22 reports the results of the regression of these three indices on ethnic diversity for our favorite specification. Our results are robust to these alternative indices: the effect of ethnic diversity is still strongly significant for the index of Housing Quality, and is even stronger for the index for Neglect of Public Areas. As noted previously, there is no effect of diversity on public safety.

**Table 1.22:** Diversity and Public goods: sum of the various outcomes, Public Housing

	Ethnic Diversity (1)
<b>1. Neglect of the Public Areas</b>	
Sum of the outcome variables	2.321*** (0.575)
<b>2. Quality of Housing</b>	
Sum of the outcome variables	1.511*** (0.360)
<b>3. Public Safety</b>	
Sum of the outcome variables	0.029 (0.059)
Socio-economic	
Background of area	Yes***
Department fixed effects	Yes***

Each coefficient is estimated from a separate regression, according to equation 1.2. The three dependent variables considered are the three indices reported in bold. Each index is regressed on either ethnic diversity, controlling for the usual household and neighborhood characteristics unless otherwise indicated. Robust standard errors adjusted for block clustering are in parentheses. \*\*\*  $p < 0.01$ , \*\*  $p < 0.05$ , \*  $p < 0.1$

To be more thorough, we next perform a mean effect analysis. Following Kling et al. (2007), we construct summary indices aggregating information across the various related outcomes for each of the three dimensions studied above. To build the three summary indices, we first normalize each outcome using a pseudo-control group defined by individuals living in blocks characterized by a below-median fractionalization index, as in Glennerster et al. (2013). Let  $Y_k$  be the  $k^{th}$  of  $K$  related outcomes. Each standardized outcome  $Y_k^*$  is obtained by subtracting the mean  $\mu_k$  and dividing by the standard



deviation  $\sigma_k$  of the outcome variable among the low diversity pseudo-control group:  $Y_k^* = (Y_k - \mu_k) / \sigma_k$ . We then average the related standardized outcomes to form the summary index:  $Y^* = \sum_k Y_k^* / K$ . Accordingly, our summary index for neglect of the commons averages nine standardized measures including graffiti, broken mailboxes, broken elevator, low care of the commons, voluntary degradations and garbage on the floor; the index for poor housing conditions averages five standardized measures of quality of apartment's soundproofing, of efficiency of the heating system, and of the general state of the outside walls; and finally the public safety indicator averages standardized measures of robbery and personal aggression.

**Table 1.23:** Components of Summary Indices, Public Housing

	Low ELF		High ELF - low ELF	
	Raw (1)	Norm (2)	Raw (3)	Norm (4)
<b>1. Neglect of the Public Areas</b>				
Care of the commons	1.49	0	0.14	0.19
Damaging in the premises	1.45	0	0.25	0.36
Graffiti	0.63	0	0.03	0.07
Garbage on the floor	0.42	0	0.08	0.16
Broken glass	0.31	0	0.06	0.13
Broken doors	0.32	0	0.01	0.01
Broken light bulbs	0.18	0	0.07	0.18
Broken mailboxes	0.33	0	0.09	0.19
Broken elevators	0.14	0	0.1	0.28
<b>2. Quality of Housing</b>				
Condition of the outside walls	2.42	0	0.01	0.01
Quality of soundproofing	1.83	0	0.23	0.28
Noisy in daytime	1.48	0	0.16	0.22
Noisy in night time	1.27	0	0.15	0.27
Cold in the apartment	0.14	0	0.08	0.23
<b>3. Public Safety</b>				
Robberies	0.08	0	0.01	0.05
Aggressions	0.06	0	0.02	0.08

Table 1.23 presents the raw and normalized components of the three broad summary indices. The first column displays the mean of each outcome among the low-diversity group. The normalized outcomes for this pseudo-control group are displayed in column 2, with mean equal to zero by construction. Column 3 reports the difference between

the mean of each outcome among the households living in high diversity neighborhoods (a treated group of sorts) and that of the low diversity population. All but one of the differences are positive indicating that the average outcome is generally worse in more heterogeneous areas. Column 4 shows the difference between the normalized outcomes for treatment and control group, and allows for a more comprehensive reading. For instance, we know from column 3 that the raw difference between care of the commons in low and high diversity areas is of 0.14. Column 4 now tells us that this difference is of 0.19 standard deviations, relative to the control group standard deviation.

**Table 1.24:** Diversity and Public goods: mean effect estimates, Public Housing

	Ethnic Diversity (1)
<b>1. Neglect of the Public Areas</b>	
Mean effect estimate	0.545*** (0.130)
<b>2. Quality of Housing</b>	
Mean effect estimate	0.467*** (0.107)
<b>3. Public Safety</b>	
Mean effect estimate	0.050 (0.112)
Socio-economic	
Background of area	Yes***
Department fixed effects	Yes***

Each coefficient is estimated from a separate regression, according to equation 1.2. The three dependent variables considered are the three summary indices reported in bold. Each index is regressed on either ethnic diversity, controlling for the usual household and neighborhood characteristics unless otherwise indicated. Robust standard errors adjusted for block clustering are in parentheses. \*\*\* p<0.01, \*\* p<0.05, \* p<0.1

Table 1.24 reports mean effect estimates from regressing the summary indices for negligence, quality of housing and public safety on ethnic diversity and other variables, as in specification 1.2. The coefficient on ethnic diversity is the mean effect size. As expected, for the negligence index and the quality of housing index, mean effect estimates of ethnic diversity are strongly positive (column 1). Using summary indices also allows us to compare the mean effect of diversity on those two normalized outcomes: lines 1 and 2 of Table 1.24 tell us that ethnic diversity has a more adverse impact on the neglect of common areas than on the average quality of housing. This gives us an insight concerning the mechanisms at play: high levels of ethnic diversity are more likely to generate uncivic behaviors that could be avoided by higher quality social norms. By contrast, the mean ef-

fect estimate in the third line indicates that ethnic diversity plays no role on public safety.

The broad picture drawn in the three previous sets of regressions is largely confirmed by the regressions of each separate normalized outcome. The corresponding mean effect estimates of ethnic diversity are presented in table 1.25. Although diversity has no significant impact on a few outcomes, such as broken glass or broken light bulbs in the commons, we still have a very strong negative effect of diversity on every other negligence or housing quality outcome in the public sector. As noted earlier, the mean effects estimates for negligence outcomes are on average larger than those for housing quality. The effect measured on broken light bulbs is the strongest, with a more than one standard deviation difference between low and high diversity neighborhoods, while the lowest is obtained for the quality of sound proofing, with a difference of about one third in terms of its standard deviation. Turning to public safety, the mean effect estimates on robberies and direct aggressions are both insignificant in the public housing sector in our favorite specification.

**Table 1.25:** Ethnic diversity and disaggregated housing outcomes : mean effects analysis,  
Public Housing

	Ethnic diversity	
	(1)	(2)
<b>1. Neglect of the Public Areas</b>		
Care of the commons	0.752*** (0.124)	0.549*** (0.164)
Damaging the premises	1.271*** (0.140)	1.019*** (0.183)
Graffiti	0.387** (0.151)	0.488** (0.216)
Garbage on the floor	0.668*** (0.162)	0.625*** (0.227)
Broken glass	0.475*** (0.182)	0.368 (0.238)
Broken doors	0.228 (0.176)	0.110 (0.241)
Broken light bulbs	0.878*** (0.187)	1.049*** (0.248)
Broken mailboxes	0.652*** (0.176)	0.927*** (0.239)
Broken elevators	0.694*** (0.227)	0.656** (0.288)
<b>2. Quality of Housing</b>		
Condition of the outside walls	0.629*** (0.109)	0.414*** (0.151)
Quality of soundproofing	0.963*** (0.099)	0.393*** (0.138)
Noisy in daytime	0.935*** (0.110)	0.613*** (0.148)
Noisy in night time	1.096*** (0.121)	0.676*** (0.159)
Cold in the apartment	0.634*** (0.128)	0.418** (0.184)
<b>3. Public Safety</b>		
Robberies	0.207** (0.103)	0.149 (0.140)
Aggressions	0.231** (0.104)	-0.0489 (0.149)
Socio eco. background and department. fixed effects	No	Yes

Each entry is the coefficient estimate on ethnic diversity from a separate regression.

All the regressions include controls for household characteristics.

Robust standard errors adjusted for block clustering are in parentheses. The components of the three summary indices are the variables listed below each of them. Descriptive statistics for these outcomes are in Table 1.5.

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



## Chapter 2

# The Local Determinants of Victimization

### 1 Introduction

Some factors such as a high population density or a large unemployment rate, are known for rising crime rates. Living in a deprived US county, Italian province or French department hence puts one at a higher risk of being victim of a criminal event than living in a prosperous region. Yet, such a statement may hide important spatial disparities: a region characterized by well-defined social and economic attributes generally encompasses very heterogeneous areas. As we zoom in and focus on smaller and smaller areas, the relationship between social, economic or demographic characteristics and crime rates established at more aggregate levels may be altered. Consider for instance two adjacent neighborhoods, a prosperous one and a depressed one. Admittedly, unemployment breeds crime, so that the depressed neighborhood will be a nest for criminals. Those criminals may act in their own neighborhood, but they may as well travel to the more attractive adjacent neighborhood. At some point, there could even be a negative relationship between a neighborhood unemployment and crime rates. Alternatively, a dual crime market could exist, with different types of crimes committed in different neighborhoods (e.g. vandalism in poor neighborhoods and burglaries in wealthy areas). This simple example illustrates the idea that studying crime from a more microeconomic perspective can challenge some of the established results, and lead to a better understanding of the mechanisms behind criminal events. In particular, adopting a local approach allows the researcher to ask or revisit the following questions. What are the local determinants of crime? To what extent is the probability to be victim of a criminal event in a given neighborhood affected by the characteristics of surrounding areas? Can we observe a duality in crimes, with some crimes explained by intrinsic neighborhoods characteristics and others explained by the characteristics of more distant areas?

This paper answers these questions taking advantage of the French victimization survey that provides detailed information localized at a very low geographic level (a 2,000 inhabitants neighborhood). Note upfront that the survey asks whether individuals have been victims of any criminal event, that is *victimized*, but does not inform about people committing crimes.<sup>1</sup> Hence, I am able to characterize the circumstances of a crime and the victim, but not the criminal. Three important findings emerge from this study. First, neighborhood characteristics explain victimization better than individual characteristics, except for assaults. Second, among the various neighborhood characteristics considered, unemployment rate appears as the most relevant factor having a positive effect on victimization. Third, adopting a spatial approach reveals that for crimes such as burglaries and thefts of objects from cars, the effect of unemployment rate in surrounding neighborhoods is stronger than the effect in the neighborhood where the crime took place, while the reverse is true for smaller crimes.

The present work differs from the previous literature by exploring the determinants of victimization at a very low geographic level. The geographic unit considered, called IRIS, is a 2,000 individuals neighborhood and is the smallest census tract unit for which representative indicators can be constructed in France. Instead, existing results are generally obtained using more aggregate data: Gould et al. (2002) and Kelly (2000) rely on US counties, which add up to 3,140 units for the whole country; Buonanno et al. (2009) and Bianchi et al. (2012) are based on 95 Italian provinces; Machin and Meghir (2004) rely on 43 police force areas for England and Wales; and Fougère et al. (2009) use data from the 96 French *départements*. An exception is Bell et al. (2010) who study the impact of immigration on crime using data from 371 local authorities across England and Wales. Although I am not questioning the validity of the results based on aggregate data, I think that they present an important drawback. These studies fail to account for criminals' mobility, implicitly assuming that the offenders commit crimes in the area where they live (e.g. provinces), ignoring the heterogeneity across neighborhoods within this broad area (in terms of economic conditions for instance). By contrast, I argue that according to the type of crime and the expected pay-off, criminals might either operate in their own neighborhood or travel to a remote area. It is for instance reasonable to think that thieves are more likely to live in deprived neighborhoods and to steal from wealthier (possibly neighboring) areas, while young delinquents will not have any incentives to bear transportation costs in order to vandalize cars in a distant neighborhood. Zenou (2005) provides some theoretical background for this idea in an urban economics model explaining the link between crime and location by highlighting the role of the housing market. In particular,

---

<sup>1</sup>I will henceforth use the term *victimized* to refer to someone having been victim of any criminal event (from property to violent crimes and vandalism). Similarly, a *victimization* will refer to the event making one a victim.

distance to the city center (where jobs and crime opportunities are located) affects the decision to commit crime instead of working by increasing commuting costs and reducing housing rents. The idea that distance and mobility matter in criminal decision also finds some empirical support in the criminology literature. It documents that the places where perpetrators commit crimes often differ from their area of residence, and that the distance between the two locations varies with the accessibility of the target area, the type of crime and the offender's characteristics (see [Bruinsma \(2007\)](#) for a detailed survey on the Netherlands, and [Bernasco and Luykx \(2003\)](#) for an analysis of criminals' target location choice). Working at a very local level enables me to add a spatial dimension to the study of crime, which is a key input to the literature. I am indeed able to compare the effects of the characteristics of adjacent neighborhoods on crime and therefore to capture more precisely the mechanisms behind the relationships obtained with aggregate data.

In addition to allowing for the localization of the surveyed individuals at the neighborhood level, the victimization survey data used in this paper present several valuable features. First, in some cases, it is possible to know where the victimization took place, and hence to control for the characteristics of this location. By contrast, studies relying on police data usually consider the location of the police station where the crime was reported rather than the location of the event itself. Second, these data provide detailed information on individuals, so that relevant individual characteristics pertaining to potential victims' attractiveness can be taken into account, while they are ignored in most of the existing literature. Finally, victimization surveys are known for avoiding the under-reporting issue from which reported police data suffer. Not only are individuals less likely to report personal offenses or small property crimes to the police, but criminal attempts or threats are also not always taken into account by police forces. Relying on victimization survey data is thus particularly insightful regarding petty crimes and assaults. Distinguishing between petty crimes such as vandalism and more important economic crimes shows quite relevant, as these different types of crimes turn out to be driven by different channels.

The nature and the quality of these data enable me to answer the questions asked above. Regarding the local determinants of crime, I show that social, economic and demographic neighborhood characteristics are more important than individual characteristics in explaining victimization. This result is particularly strong for petty crimes such as motorbike theft or car vandalism. It holds for all types of crime considered except for assault for which individual characteristics dominate. It therefore looks as if offenders target neighborhoods rather than precise households or individuals. In particular, among the various neighborhood characteristics considered, unemployment rate appears as the most relevant factor, while factors such as the share of immigrants in the neighborhood are not important in explaining victimization. The subsequent results of this paper hence focus



on the role of unemployment on victimization. The coefficient for local unemployment rate is positive and its magnitude is particularly strong for small crimes such as motorbike theft or vandalism. Therefore, it seems that crimes committed in more deprived areas relate more to social disorganization theory (e.g. [Shaw and McKay, 1942](#)) than to rational economic crime theory *à la* Becker. Note that these results are obtained after correcting for the biases related to endogenous sorting, as will be explained below.

Finally, in order to test the idea that perpetrators may actually move across neighborhoods to commit economic crimes, I adopt a spatial approach that consists in controlling for both reference neighborhood and adjacent neighborhoods characteristics. The results show that for crimes such as burglaries and thefts of objects from cars, the effect of unemployment rate in distant neighborhoods is stronger than the effect in the reference neighborhood, while the reverse still holds for smaller crimes. Otherwise stated, for a given local unemployment rate, being surrounded by higher unemployment areas increases the probability of being burgled, but does not affect vandalism. Rather, vandalism is boosted by larger local unemployment rates for a given level of unemployment in the surrounding neighborhoods. This tends to support the idea of criminals mobility for some types of crime, e.g. economic crimes, in line with Becker's theory, but not for other types of crimes (petty crimes and vandalism), relating instead to the social disorganization theory. This result is, to my opinion, the most important finding of this work. It helps understand the mechanisms behind the finding that, at larger geographic level, unemployment increases crime: unemployment would have a direct local effect on small crimes versus a remote effect on more serious economic crimes. Not only does it mean that the relationship between unemployment and crime is not trivial as we focus on smaller areas, but that this relationship also depends on the type of offense. This result shows the importance of taking criminals' mobility into account, and implies that distance, geography and transport infrastructure might be worth getting more attention in future research on crime.

This paper obviously relates to the large economics of crime literature, initiated by [Becker \(1968\)](#) and [Ehrlich \(1973\)](#). The hypothesis developed in these seminal papers is that the decision to engage into criminal activities is the result of a rational cost-benefit analysis. Most empirical research on the economics of crime aims at testing this hypothesis, which implies that economically weaker individuals (e.g. unemployed workers) have a higher propensity to commit crime because they face lower opportunity costs. Part of the literature hence focuses on economic factors, revealing that lower wages ([Gould et al., 2002](#)), larger unemployment rates ([Fougère et al., 2009](#)) or more inequality ([Kelly, 2000](#)) generate higher crime rates. Alternatively, several studies concentrate on demographic factors such as population density: [Glaeser et al. \(1996\)](#) show that crime is rife in denser and more populated areas due to extended social interactions. A similar idea is devel-

oped by [Calvó-Armengol et al. \(2007\)](#) and [Patacchini and Zenou \(2008\)](#) who show the importance of social relationships, in particular of weak ties,<sup>2</sup> in criminal behavior. A growing literature also focuses on the role played by immigration, and provides evidence that its causal impact on crime is not significant or only very moderate. For instance, [Bianchi et al. \(2012\)](#) demonstrate that the share of immigrants in Italian provinces has only a marginal effect on crime rates through robberies. Other studies, such as [Spenkuch \(2010\)](#) or [Bell et al. \(2010\)](#) show that the effect is driven by the most economically deprived immigrants. On another aspect, [Buonanno et al. \(2009\)](#) insist on the role of social norms and show that they tend to reduce property crimes.

Incidentally, the low geographic focus of this study binds it to the literature on neighborhood effects. A major concern in this literature is that households usually sort across neighborhoods in a non-random fashion. It is then possible that some unobserved household or individual characteristics influence both the propensity to be victim of a criminal event and neighborhood characteristics, therefore biasing the results. Several methods have been used in the literature to overcome this endogenous sorting issue, such as randomized experiments or instrumental variables, that will be detailed more carefully in the paper. The approach adopted in this study follows [Bayer et al. \(2008\)](#) and builds on the very local nature of the data. The idea is that although households are able to select a given area in which they want to live, they are, however, unable to select a precise neighborhood within this given area. Therefore, once the characteristics of the larger selected area are controlled for, the remaining variation of unemployment rates across the smaller neighborhoods can be considered as exogenous.

The rest of the paper is organized as follows. Section 2 presents the data: I describe the victimization survey, and explain the particular geographic structure of the data. The empirical results are given in three distinct sections. Section 3 explores the determinants of the various types of victimization and compares the role of contextual versus individual characteristics. Section 4 deals with the issue of endogenous location selection, following the approach developed by [Bayer et al. \(2008\)](#). Section 5 is devoted to the new spatial approach, where I focus on the role of unemployment in the reference neighborhood versus adjacent neighborhoods, in an attempt to account for distance. Section 6 concludes.

## 2 Data overview

The French victimization survey (*Cadre de Vie et Sécurité* - Living Environment and Security, INSEE, CVS henceforth) is a repeated cross section, representative of mainland

---

<sup>2</sup>Weak ties are simple acquaintances, doing contrast to strong ties which are usually close friends and close relatives

France households. It has been conducted annually since 2007 and each wave contains approximately 16,000 observations (one per household). The latest wave I use is the 2011. For each type of victimization considered in the CVS survey, the respondent is asked whether it occurred at least once over the two years preceding the survey. Various types of victimization affecting households in general are considered. These are mostly property thefts (or attempts) and acts of vandalism: burglary, attempt of burglary or theft without breaking in the main home (*burglary*), car theft or attempt, (*car theft*), motorbike (or scooter) theft or attempt (*motorbike theft*), *bike theft*, act of vandalism on the main home (*home vandalism*), act of vandalism on the car (*car vandalism*), and theft of objects from the car (*car objects theft*). I will henceforth refer to these types of victimization as household victimization. In addition, one randomly selected individual in each household is asked about his/her personal experience of victimization over the past two years.<sup>3</sup> In this paper, I will consider three types of individual victimization: *robbery*, *theft* and *assault*.<sup>4</sup> The shares of households and individuals victims of a given type of victimization at least once over the previous two years are displayed in the first column of Table 2.1. These figures are obtained pooling the 2007 to 2011 waves of the victimization survey. The other columns report the figures for various types of urban units, according to their population size and their degree of urbanization. Expectedly, the probability of victimization is higher in larger urban units (more than 50,000 inhabitants) and in the Paris urban unit than in less populated and rural areas. It is also clear from this table that occurrences of victimization are very rare events, which does not ease their study. The survey reveals that very few households or individuals report repeated occurrence of a given type of victimization, so that considering the occurrence of an event or its number does not make a large difference (this is not in the table).

When a victimization is reported, the respondent gives details about the circumstances, declaration to the police, consequences (physical injuries, protection behavior), and offender (e.g. when s/he was seen or arrested). The data also contain detailed information on households such as income, home ownership status or number of children, as well as individual characteristics such as age, gender, socio-economic category, education, income and national origins. Descriptive statistics of household and individual characteristics are reported in Table 2.2. In addition, the survey describes the neighborhood: the pollster characterizes the type of housings in the neighborhood and indicates whether s/he observes evidence of vandalism (burnt cars for instance). The respondent also reports whether s/he was aware of any crime or alcohol or drug related incident in the neighborhood and characterizes the general quality of the neighborhood (street light, green spaces, buildings aspect, bunch of people hanging around). Finally, the respondent rates her/his

<sup>3</sup>The member of the household selected to answer to the individual part of the survey is the person above 14 years old whose birthday is the closest to the 1st of January.

<sup>4</sup>The survey also informs about threats or insults, but I decide to let these types of victimization aside.

feeling of insecurity.

All this information is available in the public version of the survey. I also have access to more sensitive information, through a Secure Remote Center of Access to the Data (*Centre d'Accès Sécurisé Distant, CASD*). In particular, I am able to localize the precise neighborhood where the surveyed households live. This local area, called IRIS (*Ilots Regroupés pour l'Information Statistique*) is the smallest census tract unit for which representative indicators can be constructed in France. All French municipalities with more than 10,000 inhabitants and most of the municipalities with more than 5,000 inhabitants are divided into several IRISes. Each IRIS is defined so as to be an homogeneous area in terms of living environment, and its borders follow the main topographical and landscape frontiers (e.g. roads, railways and rivers). The target size of an IRIS is 2,000 inhabitants, so that IRISes actually include between 1,800 and 5,000 inhabitants.<sup>5</sup> To give an idea of the level of aggregation, there are about 50,000 IRISes in France (for around 36,000 municipalities). By comparison, there are 96 *départements* in France, the geographical unit used by Fougère et al. (2009).<sup>6</sup> For the sake of illustration, Figure 2.1 shows a map of Paris divided into IRISes. This is of course an extreme example with very small IRISes due to the high population density in Paris. In the remainder of the paper, I will interchangeably refer to IRIS or neighborhood.

**Figure 2.1:** Paris map of IRISes



Because each wave of the CVS survey comprises about 16,000 observations, there are very few observations in each IRIS (2.3 observations per IRIS per year on average). Working at such a small scale thus prevents me from computing representative victimiza-

---

<sup>5</sup>The IRISes are thus comparable in terms of population size, but not necessarily in terms of geographical space. Typically, a small village in the countryside is not divided into IRISes and is actually an IRIS of its own, while cities with more than 5,000 inhabitants are divided into several IRISes. The denser the city considered, the smaller the size of its IRISes.

<sup>6</sup>Bianchi et al. (2012) rely on Italian Provinces, that adds up to a total of 95 units, and most studies on the US are done at the county level, that adds up to 3,140 units.

tion rates at the IRIS level. Instead, I use variables indicating whether each individual or household has ever experienced victimization over the past two years. On the bright side, working at the IRIS level presents a major advantage: it enables me to supplement the victimization data with social, economic and demographic characteristics representative of the IRIS. Indeed, the INSEE designed the IRIS to be the primary statistical unit of the census. Most of the French statistical data sources are therefore based on this geographical unit, so that it is easy to match information at the IRIS level. Using the French population censuses from 2006 to 2009, I can enrich the CVS survey data with socio-economic and demographic characteristics of the IRIS, at the time of the survey. Since 2004, the population census has been conducted annually, in a continuous way, and each wave contains information collected over five consecutive years.<sup>7</sup> For instance, the 2006 census was conducted over the 2004 to 2008 period. Individuals living in municipalities of less than 10,000 inhabitants are all surveyed once over the period. For municipalities of more than 10,000 inhabitants, 8 % of the population is surveyed each year, so that 40 % of the population is included in the final census data. Because the CVS survey data of a given year concern events that happened over the previous two years, I match them with the census data of the previous year, to be as close as possible in terms of dates: the 2007 wave of the CVS survey is hence matched with the 2006 census data and so on. To be more precise, I enrich the CVS survey data with the following characteristics, representative at the IRIS level: unemployment rate, share of single-parent households, share of immigrants, share of public housing units, share of households arrived less than two years ago and share of 14-18 years old. Furthermore, I can retrieve the IRIS median household income (per consumption unit) from tax surveys. The median household income of a given IRIS is averaged over two consecutive years (weighted by the number of consumption units) and then matched to the following wave of the CVS survey. For instance, the observations from the 2007 wave of the CVS survey are matched with the average median income of 2005 and 2006. Table 2.3 describes the most relevant contextual variables accounting for households' living environment.

### 3 Preliminary results

#### 3.1 Empirical methodology

When a victimization is reported in the survey, information is gathered about the circumstances in which it happened. In particular, the respondent has to indicate whether it took place in his/her own neighborhood or in some non-specified other place. As I am inter-

---

<sup>7</sup>Prior to 2004, the population census was conducted every decade on average, the latest one dating back from 1999.

ested in the local determinants of victimization, I restrict the victimization occurrence to the events that happened in one's own neighborhood, which I am actually able to identify and to characterize. I can then control for the socio-economic environment of the IRIS where the event occurred. Hence, for all types of victimization considered, the dependent variable takes on value 1 if the household or the individual was offended in his own neighborhood and 0 otherwise, i.e. if the offence happened outside the neighborhood or if no offence happened at all. Table 2.4 documents the extent to which victimization happens in the neighborhood. It shows that most of the household victimization happens in the neighborhood, while the reverse is true for individual victimization. Limiting the study to victimization that happened in the victim's neighborhood can hence be an issue for individual victimization, but it is the only way to control for contextual characteristics.<sup>8</sup>

Let  $i$ ,  $j$  and  $k$  indicate respectively individual, household and IRIS. For each outcome considered, I estimate the following equation.

$$VICT_i = \alpha + \beta X_i + \gamma Y_{j(i)} + \delta Z_k + \varepsilon_j \quad (2.1)$$

where  $VICT_i$  is a dummy variable indicating the occurrence of a given type of victimization at least once over the preceding two years. In the case of a household victimization  $i$  stands for the household head,  $j$  for the household and  $k$  for the IRIS, while in the case of an individual victimization,  $i$  stands for the surveyed individual.  $X_i$  is a vector of characteristics of the household head or of the interviewed individual according to the type of victimization considered (household or individual respectively). Then,  $Y_{j(i)}$  is a vector of household characteristics and  $Z_k$  a vector of social, economic and demographic characteristics of the IRIS, along with other contextual variables that are detailed below. All results presented below derive from the estimation of a linear probability model, using OLS estimates, with robust standard errors clustered at the IRIS level.<sup>9</sup>

Two broad sets of variables are used in the regressions: one to control for the living environment in a general sense ( $Z_k$ ) and another to control for individual and household characteristics ( $X_i$  and  $Y_j$ ). Regarding contextual variables, I control for social, economic and demographic neighborhood (IRIS) characteristics: median annual household income (in log), unemployment rate, share of immigrants, share of households living in the public housing sector, share of 14-18-year-old individuals, share of single-parent families (*monoparental*) and share of households that have been living in the IRIS for less than two years (*recent movers*). As population density is known to be an important factor of

<sup>8</sup>Excluding the observations for which a victimization happened outside of the neighborhood does not significantly affect the results.

<sup>9</sup>I have also run probit regressions, obtaining qualitatively similar results, which are available upon request.

crime, I control for the population density of the municipality (I do not know the density of the IRIS), along with an indicator of the size of the urban unit in which the IRIS is located. I also include a variable from the CVS survey describing the type of buildings in the neighborhood (dispersed houses out of the city, houses in a lot or in the city, apartment blocks in the city or in the suburbs). *Département* fixed effects are also included as contextual variables, with the intent of capturing more aggregate characteristics. In particular, the police force is organized at the *département* level (*préfecture*). Regarding household characteristics, the following controls are used: household monthly income (in three categories), ownership status (owner, tenant in the private housing sector or tenant in the public housing sector) and number of children in the household. As far as individual (respectively household head) characteristics are concerned, age, gender, nationality, occupation status and socio-economic category of the surveyed individual (respectively household head) are included in regressions of individual (respectively household) victimization.

### 3.2 Contextual versus individual determinants

As a first step of the analysis, I compare the role played by contextual variables to that played by individual variables, with two purposes in mind. One is to give a broad idea of the type of characteristics determining victimization, especially as victims' characteristics were never accounted for in the preceding literature. The other purpose is to give a first insight about the way offenders behave: do they target a specific house, car, or individual, or do they rather primarily target a neighborhood? To do so, I run regressions of the various types of victimization on different sets of controls, alternatively controlling for contextual and individual characteristics. Table 2.5 displays the adjusted R-squared of the various regressions, where each row stands for a given dependent variables (victimization), while each column corresponds to a different specification. Columns (1) to (4) include various sets of contextual variables, with column (4) including them all. Similarly, columns (5) to (7) include various sets of individual and household characteristics, with column (7) including them all. Finally, specifications including contextual, individual and household sets of variables are reported in column (8).<sup>10</sup> Of course, none of the R-squared is very large, mostly due to the nature of the dependent variables: not only are they binary variables, but also with a very small number of "ones". In addition victimization is very likely explained to a large extent by unobserved factors such as individuals' behavior or the way people or goods look.<sup>11</sup> Still, we can note from this table and in par-

<sup>10</sup>The results of the corresponding regressions are not shown here but are available upon request.

<sup>11</sup>For instance, an individual is less likely to get his/her phone stolen if it is in a bag than if it is on a table at a cafe terrace. Similarly, a very strong and fit man risks less of being assaulted than a very thin one.

ticular from the comparison of columns (4) and (7), that contextual variables play a more important role than household and individual characteristics for both household and individual victimization. Assaults are the exception for which the reverse is true. Therefore, it looks as if the decision to commit a crime was determined by neighborhood rather than potential victims' characteristics. It is also interesting to note that among the various sets of contextual variables used, *département* fixed effects seem to matter the least, meaning that local environment characteristics explain victimization better than those measured on a larger geographic scale.

Let me now turn to a more detailed description of the results. Table 2.6 reports the estimates from the regressions of the various types of household victimization on all contextual and individual characteristics and including year fixed effects to ensure that we capture any time trend in victimization, related to changes in laws, or economic situation for instance. The corresponding results for individual victimization are reported in Table 2.7. Regarding neighborhood characteristics, the coefficients for the unemployment rate and the share of households recently arrived are positive and significant for most of the victimization types. A possible interpretation for the latter is that the larger the share of recent movers, the less neighbors know each other, and hence the less likely they are to organize some kind of collective neighborhood watching. Another interpretation pertaining to the social disorganization theory is that weaker social ties undermine the ability of a community to exercise informal control over its members. The effect of unemployment rate is particularly large for motorbike theft, car vandalism and assault. Because these are non-lucrative and violent offenses, this effect also seems in line with social disorganization theory, reflecting social rather than financial deprivation. The strong relationship of unemployment with burglaries is less intuitive, as it suggests that burglars live and burgle in the same neighborhood. One interpretation could be that the burglaries happening in high unemployment neighborhoods are more about stealing goods for their personal use (e.g. TV sets, video game consoles, food) than for reselling them (e.g. jewelry, works of art). A possible alternative explanation is that it is easier to observe one's own neighbors and to know when they are away from home.

Although it is one of the main drivers of people's feeling of insecurity,<sup>12</sup> the share of immigrants is almost never significant, in line with the recent paper by Bianchi et al. (2012).<sup>13</sup> The type of neighborhood is also particularly relevant for most types of victimization. As expected, households living in residential areas made of groups of houses are

---

It is important to bear in mind that what we observe from the survey might be intrinsically biased precisely because of behavior. Indeed, some people are more cautious, do not walk alone at night, protect their home and their car, and are therefore less likely to record a victimization.

<sup>12</sup>These results are not shown in the paper but are available upon request.

<sup>13</sup>Exceptions are for thefts of objects from cars, and for non-violent individual theft, but these effects are limited.



more likely to be offended than those living in isolated houses in the countryside. Households living in apartments buildings are less likely to be burgled, but more likely to have their car vandalized, especially if they live in the suburbs.<sup>14</sup> Now taking a quick look at household and individual characteristics, we can see that wealthier households suffer more of home vandalism, but less of car vandalism, while their members are less likely assaulted. The result for cars is probably explained by the fact that wealthier household park their cars in a closed or secured space. Households with an unemployed head are more likely to be victims of burglary, home vandalism, and car theft. Similarly, unemployed individuals are more likely to suffer from violent crimes (robberies and assaults). Older individuals tend to be less victimized. Gender does not affect the probability of theft or robbery, but males are more victims of assaults than females.

#### 4 The issue of location selection

A major concern with this low geographic setting, that is common in the literature on neighborhood effects, is that households usually sort across neighborhoods in a non-random fashion. It is then possible that some unobserved household or individual characteristics influence both the propensity to be victim of a criminal event and neighborhood characteristics, therefore biasing the results. Several methods have been used to overcome this endogenous sorting issue, that I briefly summarize here. A first approach consists in using a measure of the variable of interest aggregated to a higher geographic level as an instrument for this variable. For instance, [Evans et al. \(1992\)](#) instrument neighborhood poverty with metropolitan area poverty. An alternative method is to rely on randomized experiments designed such that the choice of neighborhood is actually exogenous. One of the most famous examples is the Moving To Opportunity program in the US, through which randomly selected households are given housing vouchers, enabling them to relocate in richer neighborhoods. In particular, and related to the topic of the present study, [Ludwig et al. \(2001\)](#) and [Kling et al. \(2005\)](#) use this experiment to examine the role of neighborhood characteristics on juvenile crime. These are, to the best of my knowledge, the only two existing studies researching the impact of neighborhood effects on crime. However, the very particular setting in which these results are derived brings some concern regarding their validity in a more general context. [Bayer et al. \(2008\)](#) review more extensively these alternative methods and discuss their limitations.

The approach adopted in this paper builds on the very local nature of the data. It follows [Bayer et al. \(2008\)](#) who study the role of neighbors on work location. The idea is

---

<sup>14</sup>They are also less likely to have their home vandalized, but this is mechanically due to their home type.

that although households are able to select a given area in which they want to live, they are, however, unable to pinpoint a precise neighborhood within this given area. This assumption means that even if households are able to choose a given residential area, there will not be any correlation in unobserved factors affecting risk of victimization among individuals living in the same neighborhood *within the larger selected area*. It is now in order to present a few arguments supporting this assumption. First, because the housing market is very tight, it is reasonable to think that a household targeting a given area very unlikely has a choice over the precise neighborhood where it will end up in this area. This would indeed require that at least one housing unit satisfying the other decision criteria of the household (e.g. size) be vacant in each neighborhood within the larger area at the time when the household is looking for a new place. A second consideration is that it may be difficult for prospecting households to identify neighborhood-by-neighborhood variation in neighbors and contextual characteristics, prior to moving into the neighborhood. To put it differently, although the household may have a realistic *ex-ante* view on the characteristics of the targeted area, it is less likely to be actually able to identify differences in these characteristics across the various neighborhoods of the area. This makes even more sense in the context of victimization, for which *ex-ante* information is particularly difficult to gather. Finally, an interesting feature of the French neighborhoods studied here (the IRISes) is that they do not follow any kind of administrative frontier. For instance, they are distinct from police districts, and from school zones determining to which school children must go. Rather, the neighborhoods considered here were designed to encompass 2,000 inhabitants on average, and to be homogeneous in terms of living environment, with borders following the main topographical and landscape frontiers (e.g. roads, railways and rivers). People ignore where these borders are, and more generally do not even know what an IRIS is, as it is only used for statistical purpose. For those reasons, it is practically impossible that households purposely decide to live in a given IRIS rather than in a contiguous one.

All these arguments support the validity of the assumption that there should be no correlation in unobserved factors affecting victimization among neighbors living in the same neighborhood (IRIS) *within the larger selected area*. As a consequence, once we control for the characteristics of the larger area selected by the individual, the remaining spatial variation of characteristics across neighborhoods within the larger area is supposed to be exogenous. This is done through the inclusion of fixed effects of larger areas than the IRISes, literally called a "large neighborhood" in the French statistical jargon (*grand quartier*). Large enough municipalities are divided into several large neighborhoods, which themselves encompass several contiguous IRISes. If the municipality is too small, then all the IRISes of the municipality belong to the same large neighborhood, so

that the large neighborhood actually corresponds to the municipality. Although there is no formal evidence that large neighborhood is the geographical unit targeted by households when looking for a housing, this area makes a reasonable reference neighborhood compared to the IRIS. Figure 2.2 depicts the four large neighborhoods of Paris 7th *arrondissement*: each area of specific color is a large neighborhood, and each subdivision of a large neighborhood is an IRIS.

**Figure 2.2:** Paris 7th *arrondissement* map of Large Neighborhoods

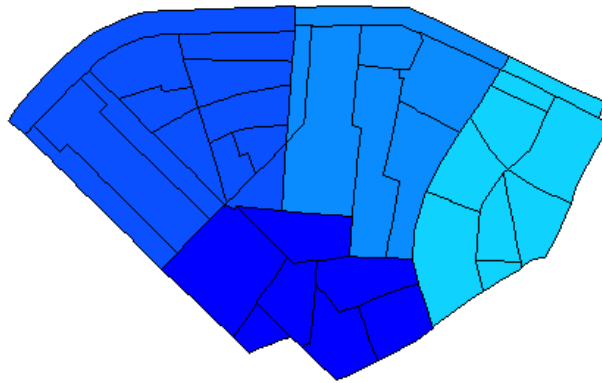


Table 2.8 summarizes the results of the regressions including large neighborhood fixed effects instead of *département* fixed effects. The specification is otherwise similar to the full specification presented in the previous section.<sup>15</sup> According to these estimates, a larger share of immigrants in the IRIS would imply a lower probability of being burgled and robbed. On average, the coefficients on the share of immigrants are lower than in the previous specification (Tables 2.6 and 2.7): the coefficients for *theft of objects from cars* and *theft* were significantly positive and are now driven down to zero (in part due to an increase in the standard errors), while some of the coefficients that were not significantly different from zero are now significantly negative (*burglary* and *robbery*). This suggests that controlling for large neighborhood fixed effects actually corrects a bias induced by the fact that immigrants tend to settle in more criminogenic areas because of lower rents for instance. The estimates for unemployment are rather stable for burglary, car vandalism and assault. Yet, they are driven down to zero for car theft, motorbike theft and bike theft, while the coefficient for robbery becomes slightly positive. Note also that unemployment seems to be the most relevant characteristic explaining crime, both in terms of the number of victimization types for which is involved and in terms of magnitude of the effect.

<sup>15</sup>The coefficients for the variables other than IRIS characteristics are available upon request. They are not significantly affected by the inclusion of large neighborhood fixed effects.

## 5 A spatial approach

An important dimension to take into account in the study of crime is criminals' mobility. To put it simply, if criminals are not mobile, the larger the number of criminals living in a neighborhood, the more likely the other inhabitants of this neighborhood experience victimization. On the other hand, if criminals are mobile, then even individuals living in a criminal-free neighborhood may face a risk of victimization if they are located at some reasonable distance of a neighborhood populated with criminals. Consider for instance a high unemployment neighborhood, more likely to breed criminals according to Becker's theory. The potential offenders could commit crime in the neighborhood where they reside, if, for instance, they cannot afford the cost of commuting to a more distant neighborhood, or if they benefit from observing their neighbors' habits and routine activities. Alternatively, they could decide to commit a crime in a more distant neighborhood if they fear to be more easily identified in their own neighborhood, or if their neighborhood is too deprived to be attractive. Whether an offender decides to act in his own neighborhood or in a remote one thus reflects a weighting of the expected gains, the direct costs and the opportunity costs of committing crime in another neighborhood. Therefore, even if we find that unemployment increases victimization on average, the effect might actually depend on where one lives relative to high unemployment neighborhoods.

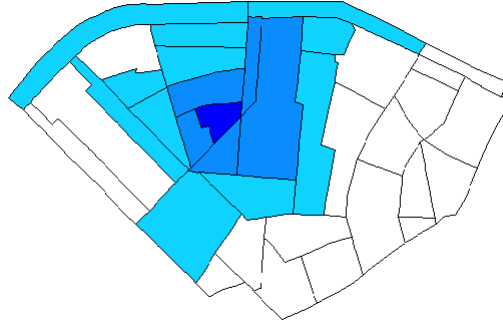
Although the question of criminals' mobility seems highly relevant when studying determinants of victimization, it is not addressed in the economics of crime literature, mainly due to the fact that most studies rely on aggregate data. By contrast, because I work with data localized at a low geographic level, I am able to connect individuals not only to the characteristics of the neighborhood where they live, but also to the characteristics of the neighborhoods that are further away. This new spatial approach enables me to indirectly account for criminals mobility.<sup>16</sup> To this aim, I consider the IRIS where the surveyed individual lives as the reference neighborhood, and I construct two successive circles of adjacent IRISes to represent more distant neighborhoods. More precisely, all the IRISes contiguous to a given IRIS constitute the first ring of adjacent neighborhoods (denoted as *IRIS1*), while all the IRISes contiguous to those in the first ring, excluding the reference IRIS and the first ring IRISes themselves, constitute the second ring of adjacent neighborhoods (denoted as *IRIS2*). The map on Figure 2.3 illustrates the three geographic levels on which I rely. The total area depicted here represents the seventh *arrondissement* of Paris. Each subdivision of this district is an IRIS. Consider for instance the IRIS colored in the darkest shade as the IRIS of reference. Then, the set of IRISes colored in a slightly lighter shade constitute the first ring of IRISes, i.e. the area made of all the

---

<sup>16</sup>It is not direct as I have no information about the offenders, so I am not actually able to locate them.

adjacent IRISes. Finally, the lightest IRISes constitute the second ring of IRISes, with respect to the reference IRIS.

**Figure 2.3:** Paris 7th *arrondissement* map of IRIS



Using this setting, it is possible to relate any individual or household in the survey to the characteristics of the neighborhood where it lives, as well as to those of the first and second rings of adjacent neighborhoods. Thus, I can explore whether a given factor matters more within the neighborhood or from a remote one. In the following empirical analysis, I will focus on one particular factor: the unemployment rate. First, as noted at the end of the previous section, this factor shows the most relevant in explaining victimization. Second, I make this choice in order to avoid a likely collinearity issue with other IRIS characteristics. The unemployment rate of a given neighborhood is indeed highly correlated with most of the other IRIS characteristics (share of immigrants, median income, share of single parents families and share of public housing), as can be seen from Table 2.9. Concretely, I compute the average unemployment rate over all the first and second rings of IRISes respectively, weighted by the size of the active population in each IRIS. To summarize, and using the same notation as in section 3, I estimate the following equation:

$$VICT_i = \alpha + \beta X_i + \gamma Y_{j(i)} + \delta U_k + \eta U_{k+1} + \nu U_{k+2} + \varepsilon_j \quad (2.2)$$

Where  $U_k$  is IRIS unemployment rate,  $U_{k+1}$  the average unemployment rate of the first ring of adjacent IRISes, and  $U_{k+2}$  the average unemployment rate of the second ring of adjacent IRISes. Note however that this geographical approach has some drawbacks. First, as I do not have any information for the road or transportation networks, I am not effectively capturing transportation time or cost, which are key determinants of mobility. This could be addressed using the information about road networks provided by the French Institute of Geography (IGN), and performing a Geography Information System analysis. However, because I do not have access to these data nor to the technology necessary to deal with it, I keep this step for future research. Second, this approach with adjacent IRISes does not enable me to directly capture distance, as IRISes are heterogeneous in terms of

size. As mentioned in Section 2, only municipalities with more than 5,000 inhabitants are divided into IRISes, and the target size of an IRIS is 2,000 inhabitants so that denser cities tend to have smaller IRISes. To deal with this issue, I restrict the sample to municipalities that are actually divided into IRISes, hence reducing the variation in the size of the IRISes. Such a restriction typically excludes rural villages, which are quite large IRISes (in terms of surface).

The estimated effects of the unemployment rate in the three successive neighborhood rings (IRIS, IRIS 1 and IRIS 2) on the various types of victimization are reported in Table 2.10. The first set of results displayed are the estimates obtained when no other control is included, while the second set of results is obtained including the full set of controls. Note that for each specification, I control for large neighborhood fixed effects to avoid the endogeneity issue, as discussed in the previous section. Let us first look at the burglaries in column (1). In the no other control specification, only the unemployment rate in the first ring of adjacent IRISes is positive and significant at the 10% level. In the full specification, both the IRIS and the IRIS 1 unemployment rates are significant at the 5% level, with a larger coefficient for the latter. The results are similar for thefts of objects from cars. The unemployment rate in the first contiguous neighborhoods is the only significant one that in the full specification. This suggests that economic types of crimes such as burglary and theft of objects from cars are better explained by the unemployment rate from more distant places than from the immediate neighborhood. This is in line with the idea that when it turns to economic crimes, offenders are more likely to travel to some remote area. Several considerations can help rationalize this: stealing from one's direct neighbors is not financially attractive when one lives in a more economically deprived neighborhood; the expected financial gain from an economic crime allows the offender to afford the cost of travelling to a more distant neighborhood; and the criminal limits the chance to be identified by witnesses when committing an offence in a different place than the one where he lives. It is then a bit puzzling that the unemployment rate in the reference IRIS still matters for burglaries. A possible explanation is that there are two types of burglars: those who travel to a remote place to steal expensive goods they can resell such as jewelry and works of art, and those who steal very basic goods such as food or TV sets from their own neighbors for their personal consumption. An alternative explanation could be that the habits and general behavior of one's direct neighbor are more easily observed, so that it simplifies the planning of the crime. It is also surprising not to find any significant effect on car and motorbike thefts. A possible explanation could be that stealing this type of goods requires an even longer distance, so that it is more easy to stock or use the car.<sup>17</sup> On the other hand, Table 2.10 also shows that the unemployment

---

<sup>17</sup>Note that when *département* fixed effects are included instead of large neighborhood fixed effects,

rate in the immediate neighborhood is particularly relevant in explaining non-economic and violent crimes such as acts of vandalism, whether on the home or on the car, and assaults. In this case, the social disorganization theory is more appropriate to understand the mechanisms than the Beckerian approach.

The concern linked to the IRIS size may persist, with the existence of very small IRISes in densely populated cities such as Paris. In this case, the distance between two IRISes may not be relevant, with a null transportation cost across the three contiguous rings of IRISes. In what follows, I therefore exclude the observations of the three largest cities (Paris, Lyon and Marseille), hence getting rid of the smallest IRISes. The regressions presented above are then replicated on this sub-sample. The estimates for unemployment rates in the successive rings of IRISes in the full specification are reported in Table 2.11. The previous results are robust to this sample restriction. The results are stable for burglaries, with a positive effect of direct (IRIS) unemployment rate and a larger positive effect of more distant (IRIS 1) unemployment. Note that the gap between the two coefficients is even slightly larger than in the previous table. The estimates for both acts of vandalism, theft of objects from car and assaults are also similar to those presented above. There are however two differences compared to the regressions including large cities: the coefficient for IRIS unemployment rate is now significantly positive (at the 10% level) in the motorbike theft and in the robbery equation. To the extent that robberies are violent crimes by opposition to thefts, this new result tends to comfort the idea that the direct exposure to unemployment affects violent rather than economic crimes. The result for motorbike theft could be at odds with this intuition, unless most of the thefts observed apply to motorcycles rather than to more powerful motorbikes.

## 6 Conclusion

This paper is, to the best of my knowledge, the first study on victimization at the neighborhood level. This local approach brings new insights to the economics of crime literature as it enables me to characterize precisely the context (both location and victim) in which criminal events occur. By contrast with previous papers based on aggregate police data, I am therefore able to distinguish between factors related to the opportunity cost of committing crime (e.g. unemployment, wages) and factors pertaining to the attractiveness of the victims (e.g. wealth). I find that household and individual characteristics are minor determinants of household and individual victimization respectively, while the economic situation of the neighborhood actually matters. In other words, victims (individuals or

---

the coefficient of  $u_{iris1}$  is positive and significant at the 5% level in the full specification for the car theft regression, supporting this intuition.

households) are not directly targeted (except in the case of assaults): rather, it is the neighborhood where the mischief occurs that is coveted. In particular, local unemployment rate is found to be strongly related to household victimization. In order to address the endogenous neighborhood selection issue, I included "large neighborhood" fixed effects in order to control for the characteristics of the larger area that the households are likely to have actually selected. Most of the estimates of neighborhood characteristics are attenuated once selection is corrected for. Yet, the local unemployment rate remains a strong predictor of several types of victimization.

This paper also sheds new light on the mechanisms behind this relationship, through the adoption of an original spatial approach. I take advantage of the precise location of the data to control for the characteristics of both the reference neighborhood and the first and second rings of adjacent neighborhoods. That way, I can account for heterogeneity across neighborhoods and hence indirectly for criminals mobility. This is an improvement over the existing literature which ignores this dimension. The results reveal that for burglaries and thefts of objects from cars, unemployment rate in the adjacent neighborhoods have a stronger explanatory power than unemployment in the precise neighborhood where the misdeed occurred. On the contrary, local unemployment rate dominates over more distant unemployment rates in explaining vandalism and assaults in particular. A natural interpretation of these findings is that criminals are mobile for economic crimes but not for violent crimes. In other words, they can afford some transportation cost when they expect a financial reward from their mischief, in line with the Beckerian theory of crime. On the other hand, violent crimes and vandalism escape from this logic and relate more to the social disorganization theory. This new method helps understanding more precisely criminal behavior according to the different types of crimes, and is therefore a key contribution to the literature.

Naturally, the empirical design endorsed in this paper presents some drawbacks and will be subject to future improvements. For instance, considering only two rings of adjacent neighborhoods is somehow arbitrary and is an important limitation as criminals may travel from more remote areas. In particular, car thefts may involve longer distances so as to reduce the risk of apprehension. This could explain why none of the unemployment rate estimates (IRIS, IRIS1 and IRIS2) is significant for this type of crime. One of the next developments of this work will therefore be to take into account all neighborhoods in an exhaustive fashion. The idea would be to express crime as a function of the sum of unemployment rates in all surrounding neighborhoods, weighted by distance or transportation costs. In other words, this would consist in adapting the market-potential function developed in the new economic geography literature (e.g. [Harris, 1954](#); [Hanson, 2005](#)) to the economics of crime literature. Because it reveals the relevance of a spatial approach and stresses its necessity, the present paper is a first step in this direction.



## Tables

**Table 2.1:** Share of households or individuals victimized at least once over the past two years.

		Full Sample (1)	Rural Areas (2)	Less than 50,000 inhab. (3)	More than 50,000 inhab. (4)	Paris Urban Unit (5)
<b>Household victimization</b>						
Burglary	Mean	4.61 %	3.81 %	4.01 %	5.20 %	5.18 %
	StDev	(.210)	(.191)	(.196)	(.222)	(.222)
	N	85141	16211	19175	35755	13895
Car Theft	Mean	3.36 %	1.59 %	2.97 %	4.19 %	4.87 %
	StDev	(.180)	(.125)	(.179)	(.200)	(.215)
	N	69226	14953	16599	28183	9413
Motorbike Theft	Mean	5.34 %	2.25 %	4.42 %	7.28 %	8.91 %
	StDev	(.225)	(.148)	(.205)	(.260)	(.285)
	N	10051	2633	2470	3755	1181
Bike Theft	Mean	3.71 %	1.14 %	2.56 %	5.52 %	6.34 %
	StDev	(.189)	(.106)	(.158)	(.228)	(.244)
	N	46321	10974	11687	17877	5730
Home Vandalism	Mean	4.13 %	1.93 %	3.93 %	5.68 %	3.69 %
	StDev	(.199)	(.138)	(.194)	(.231)	(.188)
	N	85142	16214	19177	35751	13895
Car Vandalism	Mean	10.46 %	5.66 %	8.46 %	13.42 %	14.48 %
	StDev	(.306)	(.231)	(.278)	(.341)	(.352)
	N	69192	14955	16593	28170	9396
Car Object Theft	Mean	6.71 %	3.52 %	5.41 %	8.16 %	10.79 %
	StDev	(.250)	(.184)	(.226)	(.274)	(.310)
	N	69227	14953	16598	28186	9412
<b>Individual victimization</b>						
Robbery	Mean	0.95 %	0.29 %	0.54 %	1.13 %	2.09 %
	StDev	(.097)	(.054)	(.073)	(.106)	(.143)
	N	85154	16213	19177	35759	13900
Theft	Mean	3.38 %	2.46 %	2.88 %	3.74 %	4.62 %
	StDev	(.181)	(.155)	(.167)	(.190)	(.210)
	N	85148	16211	19176	35759	13897
Assault	Mean	2.42 %	1.76 %	2.03 %	3.09 %	2.34 %
	StDev	(.154)	(.132)	(.141)	(.173)	(.151)
	N	85142	16212	19171	35758	13896

**Table 2.2:** Sample characteristics: households and individuals characteristics.

<b>Household Characteristics</b>				
	[Min - Max]	Mean	(StDev)	Med
<i>Household monthly income:</i>				
$w \leq 1500$	[0 - 1]	0.313	(0.464)	0
$1500 < w \leq 2500$	[0 - 1]	0.294	(0.456)	0
$w > 2500$	[0 - 1]	0.393	(0.488)	0
<i>Ownership Status:</i>				
Owner	[0 - 1]	0.598	(0.490)	1
Rent in private market	[0 - 1]	0.213	(0.410)	0
Rent in public housing	[0 - 1]	0.144	(0.351)	0
Other	[0 - 1]	0.045	(0.207)	0
<i>Household composition:</i>				
Head with a partner	[0 - 1]	0.569	(0.495)	1
Number of children	[0 - 11]	0.644	(0.993)	0

	<b>Household Head</b>				<b>Individual</b>			
	[Min - Max]	Mean	(StDev)	Med	[Min - Max]	Mean	(StDev)	Med
<i>Age</i>	[15 - 101]	53.25	(17.87)	52	[14-102]	47.19	(19.62)	46
<i>Gender</i>	[0 - 1]	0.622	(0.485)	1	[0 - 1]	0.479	(0.500)	0
<i>Nationality:</i>								
Native French	[0 - 1]	0.908	(0.290)	1	[0 - 1]	0.907	(0.290)	1
Naturalized French	[0 - 1]	0.043	(0.203)	0	[0 - 1]	0.042	(0.201)	0
EU 15	[0 - 1]	0.021	(0.142)	0	[0 - 1]	0.019	(0.136)	0
Other EU (after 2004)	[0 - 1]	0.001	(0.037)	0	[0 - 1]	0.001	(0.037)	0
Maghrebian	[0 - 1]	0.013	(0.115)	0	[0 - 1]	0.014	(0.118)	0
Other African	[0 - 1]	0.005	(0.073)	0	[0 - 1]	0.006	(0.075)	0
Other nationality	[0 - 1]	0.009	(0.092)	0	[0 - 1]	0.010	(0.100)	0
<i>Employment status:</i>								
Employed	[0 - 1]	0.559	(0.496)	1	[0 - 1]	0.488	(0.500)	0
Unemployed	[0 - 1]	0.041	(0.198)	0	[0 - 1]	0.059	(0.235)	0
Inactive	[0 - 1]	0.340	(0.490)	0	[0 - 1]	0.453	(0.498)	0
<i>Socio-economic Category:</i>								
Farmer	[0 - 1]	0.014	(0.117)	0	[0 - 1]	0.012	(0.108)	0
Craftsman, shopkeeper	[0 - 1]	0.046	(0.209)	0	[0 - 1]	0.034	(0.182)	0
Higher occupation	[0 - 1]	0.111	(0.314)	0	[0 - 1]	0.078	(0.269)	0
Intermediate occupation	[0 - 1]	0.150	(0.357)	0	[0 - 1]	0.128	(0.334)	0
Employee	[0 - 1]	0.130	(0.336)	0	[0 - 1]	0.165	(0.371)	0
Factory worker	[0 - 1]	0.155	(0.362)	0	[0 - 1]	0.131	(0.337)	0
Retired	[0 - 1]	0.353	(0.478)	0	[0 - 1]	0.286	(0.452)	0
Other inactive	[0 - 1]	0.041	(0.198)	0	[0 - 1]	0.166	(0.372)	0

Reading: The head of the average household is 53 years and 3 months old. The average surveyed individual is about 47 years and 2 months old. 55.9 % of household have an employed head. 48.8 % of individuals are employed.

**Table 2.3:** Sample characteristics: households living environment

	Contextual Variables			
	[Min - Max]	Mean	(StDev)	Med
<i>IRIS Characteristics:</i>				
Share of immigrants	[0 - 0.794]	0.079	(0.074)	0.055
Median income (log)	[7.69 - 10.98]	9.792	(0.258)	9.784
Unemployment rate	[0 - 0.741]	0.112	(0.056)	0.100
Share single-parent families	[0 - 0.673]	0.137	(0.068)	0.127
Share hh in public housing	[0 - 1]	0.136	(0.187)	0.064
Share of recent movers	[0 - 0.935]	0.129	(0.061)	0.116
Share of 14-18 y.o.	[0 - 0.239]	0.056	(0.017)	0.056
<i>City density (log)</i>	[-1.09 - 10.55]	6.297	(1.955)	6.292
<i>Type of neighborhood:</i>				
Dispersed houses	[0 - 1]	0.176	(0.381)	0
Houses Lot / in cities	[0 - 1]	0.443	(0.497)	0
Apartment block (city)	[0 - 1]	0.231	(0.422)	0
Apartment block (suburbs)	[0 - 1]	0.091	(0.288)	0
Mixed	[0 - 1]	0.059	(0.235)	0
<i>Size of the Urban Unit:</i>				
Rural Areas	[0 - 1]	0.226	(0.418)	0
Less than 50,000	[0 - 1]	0.251	(0.433)	0
More than 50,000	[0 - 1]	0.365	(0.481)	0
Paris Urban Unit	[0 - 1]	0.158	(0.165)	0

Reading: The average household lives in an IRIS where there are 7.9 % of immigrants. 59.8 % of households own their home. The head of household lives with a partner in 56.9 % of households.

**Table 2.4:** Probability that the incident occurs in own's neighborhood

	Mean	(StDev)	N
<b>Household Victimization</b>			
Car theft	0.724	(0.447)	2,052
Motorbike theft	0.643	(0.479)	497
Bike theft	0.755	(0.430)	1,620
Vandalism on the car	0.657	(0.475)	6,581
Theft of object from car	0.289	(0.454)	4,080
<b>Individual Victimization</b>			
Robbery	0.396	(0.489)	633
Theft	0.301	(0.459)	2,308
Assault	0.372	(0.484)	1,725

When at least one offence is reported, more details are asked about the latest event. In particular, the respondent indicates whether the incident happened in one's "own village or neighborhood". Reading: 72.4 % of the latest car theft happened in the owner's neighborhood. 37.2 % of the latest assaults occurred in the victim's neighborhood.

Table 2.5: Occurrence of victimization: role of environment vs household characteristics

	IRIS	Environment	Département fixed effects	IRIS, environment, dép f.e.	Household	Individual	Household, individual	All sets of Controls
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
<b>All hh vict</b>	0.021	0.022	0.012	<b>0.032</b>	0.015	0.018	<b>0.022</b>	0.050
Burglary	0.001	0.003	0.002	<b>0.005</b>	0.001	0.002	<b>0.003</b>	0.007
Car theft	0.004	0.005	0.006	<b>0.009</b>	0.004	0.005	<b>0.007</b>	0.013
Motorbike theft	0.021	0.023	0.005	<b>0.026</b>	0.013	0.012	<b>0.020</b>	0.035
Bike theft	0.018	0.018	0.005	<b>0.021</b>	0.014	0.012	<b>0.019</b>	0.031
Home vandalism	0.007	0.008	0.005	<b>0.013</b>	0.001	0.001	<b>0.003</b>	0.015
Car vandalism	0.029	0.028	0.009	<b>0.036</b>	0.018	0.017	<b>0.025</b>	0.047
Theft of car objects	0.012	0.008	0.005	<b>0.014</b>	0.005	0.006	<b>0.007</b>	0.017
<b>All ind vict</b>	0.007	0.007	0.004	<b>0.009</b>	0.005	0.005	<b>0.007</b>	0.014
Robbery	0.003	0.003	0.001	<b>0.004</b>	0.001	0.002	<b>0.002</b>	0.005
Theft	0.003	0.003	0.003	<b>0.004</b>	0.001	0.001	<b>0.002</b>	0.005
Assault	0.003	0.003	0.000	<b>0.004</b>	0.005	0.004	<b>0.007</b>	0.008

The figures reported in this table are adjusted  $R^2$  from OLS regressions of the occurrence of a given type of victimization (first column) on a given set of controls, as specified in columns (1) to (8). The set of controls are defined as follows. (1) IRIS characteristics: share of immigrants, median income (log), unemployment rate, share of single-parent families, share of household living in a social housing, share of recently arrived (less than two years) households and share of youth (14 to 18 y.o.); (2) Environment characteristics: city density (log), type of neighborhood (5 categories, e.g. houses vs apartment buildings), size of the urban unit (4 categories); (3) Département fixed effects; (4)=(1)+(2)+(3); (5) Household characteristics: income, ownership status, number of children; (6) Individual characteristics (corresponding to the surveyed individual for individual victimization, and to the household head for household victimization): nationality, age (log), gender, employment status, socio-economic category; (7)=(5)+(6); (8)=(4)+(7). In each regression, IRIS level clusters are used.

Table 2.6: Household Victimization: Full Specification (to be continued)

	Burglary (1)	Car Theft (2)	Motorbike Theft (3)	Bike Theft (4)	Home Vandalism (5)	Car Vandalism (6)	Theft of car objects (7)
<i>Neighborhood characteristics</i>							
Share of Immigrants	0.032 (0.023)	0.009 (0.021)	-0.045 (0.060)	0.040 (0.032)	0.008 (0.023)	-0.040 (0.035)	0.041** (0.019)
Median Income (log)	0.009 (0.007)	-0.006 (0.006)	0.019 (0.018)	0.007 (0.010)	-0.008 (0.007)	0.020* (0.011)	-0.002 (0.006)
Unemployment rate	0.116*** (0.028)	0.046* (0.024)	0.218** (0.081)	0.097** (0.035)	0.043 (0.028)	0.261*** (0.044)	0.006 (0.024)
Share Monoparental	0.013 (0.021)	0.008 (0.017)	-0.001 (0.054)	0.008 (0.024)	0.014 (0.021)	0.107*** (0.030)	0.040** (0.016)
Share Public Housing	-0.014* (0.008)	-0.015* (0.008)	0.028 (0.029)	-0.034** (0.011)	0.011 (0.009)	-0.014 (0.013)	-0.014* (0.007)
Share Recent Movers	0.042** (0.018)	-0.014 (0.016)	0.010 (0.055)	0.107*** (0.027)	0.123*** (0.021)	0.167*** (0.030)	0.044** (0.017)
Share 14-18 y.o.	0.124* (0.067)	0.077 (0.056)	-0.110 (0.170)	0.024 (0.084)	0.043 (0.068)	0.080 (0.098)	0.092* (0.053)
City Density (log)	0.002** (0.001)	0.000 (0.001)	0.003 (0.002)	0.006*** (0.001)	0.007*** (0.001)	0.007*** (0.001)	0.001* (0.001)
<i>Type of buildings in the neighborhood (Ref: Dispersed houses)</i>							
Houses Lot / in cities	0.005* (0.003)	0.007*** (0.002)	0.001 (0.004)	-0.000 (0.002)	0.013*** (0.002)	0.016*** (0.002)	0.005*** (0.001)
Apartment block (city)	-0.019*** (0.004)	0.005* (0.003)	0.036*** (0.010)	0.018*** (0.004)	-0.010** (0.004)	0.027*** (0.005)	0.009** (0.003)
Apartment block (suburbs)	-0.026*** (0.005)	0.004 (0.004)	0.026* (0.015)	0.006 (0.006)	-0.011** (0.005)	0.030*** (0.007)	0.007* (0.004)
Mixed	-0.017*** (0.005)	0.012** (0.004)	0.019 (0.012)	0.006 (0.006)	0.003 (0.005)	0.024*** (0.007)	0.012** (0.004)
<i>Size of the Urban Unit (Ref: Rural Areas)</i>							
Less than 50,000	-0.002 (0.003)	0.006** (0.002)	0.001 (0.005)	-0.006** (0.002)	0.004 (0.003)	-0.003 (0.003)	-0.001 (0.002)
More than 50,000	0.005 (0.005)	0.010*** (0.003)	0.000 (0.008)	0.001 (0.004)	0.007* (0.004)	0.010** (0.005)	0.001 (0.003)
Paris Urban Unit	-0.013 (0.011)	-0.000 (0.009)	-0.019 (0.018)	-0.004 (0.010)	-0.005 (0.010)	0.024* (0.013)	0.003 (0.007)
<i>Household income (Ref: bottom 30%)</i>							
Middle 30%	-0.003 (0.002)	-0.003 (0.002)	-0.006 (0.008)	-0.004 (0.003)	0.006** (0.002)	-0.006** (0.003)	-0.003** (0.002)
Top 30%	0.002 (0.003)	-0.001 (0.002)	-0.007 (0.008)	-0.008** (0.003)	0.007** (0.002)	-0.011** (0.003)	-0.003 (0.002)
<i>Household Ownership Status (Ref: Owner)</i>							
Rent in private market	-0.004 (0.003)	0.007** (0.002)	0.011* (0.007)	0.016*** (0.003)	-0.016*** (0.002)	0.015*** (0.004)	0.004** (0.002)
Rent in public housing	-0.002 (0.003)	0.008** (0.003)	0.020* (0.011)	0.013*** (0.004)	-0.002 (0.003)	0.024*** (0.005)	0.007** (0.003)
Other	-0.004 (0.004)	0.000 (0.003)	-0.002 (0.013)	0.002 (0.005)	-0.004 (0.004)	0.008 (0.006)	-0.001 (0.003)
Number of children in hh	0.003** (0.001)	0.003** (0.001)	0.007** (0.002)	0.011*** (0.001)	0.004*** (0.001)	0.008*** (0.001)	0.003*** (0.001)

Standard errors clustered at the IRIS level are reported in parentheses. \* p<0.10, \*\* p<0.05, \*\*\* p<0.001

Table 2.6: Household Victimization: Full Specification (continued)

	Burglary (1)	Car Theft (2)	Motorbike Theft (3)	Bike Theft (4)	Home Vandalism (5)	Car Vandalism (6)	Theft of car objects (7)
<i>Nationality</i> (Ref: Native French)							
Naturalized French	-0.004 (0.004)	0.006 (0.004)	-0.003 (0.015)	0.005 (0.006)	-0.005 (0.004)	0.014** (0.007)	-0.000 (0.004)
EU 15	-0.002 (0.006)	0.009 (0.006)	0.032 (0.021)	-0.009 (0.007)	0.003 (0.006)	0.006 (0.009)	-0.001 (0.005)
Other EU (after 2004)	0.006 (0.027)	0.042 (0.040)	0.076 (0.122)	0.097 (0.069)	-0.034** (0.012)	0.074 (0.056)	0.005 (0.029)
Maghrebian	-0.016** (0.006)	-0.010 (0.007)	0.112* (0.061)	0.039** (0.018)	-0.008 (0.007)	-0.017 (0.013)	-0.001 (0.008)
Other African	-0.020** (0.009)	-0.025* (0.013)	-0.095*** (0.020)	0.059* (0.032)	-0.013 (0.009)	0.021 (0.030)	0.018 (0.019)
Other nationality	-0.002 (0.009)	-0.007 (0.010)	0.087 (0.056)	0.037* (0.019)	-0.014* (0.008)	-0.008 (0.017)	-0.015* (0.008)
<i>Employment Status</i> (Ref: Employed)							
Unemployed	0.024*** (0.005)	0.010** (0.005)	0.021 (0.015)	0.009 (0.007)	0.020*** (0.005)	0.018** (0.008)	0.003 (0.004)
Inactive	0.011 (0.010)	0.013 (0.010)	0.046 (0.036)	0.008 (0.011)	0.011 (0.010)	-0.010 (0.016)	-0.004 (0.010)
Age ( <i>log</i> )	-0.003 (0.003)	-0.018*** (0.003)	0.002 (0.011)	-0.003 (0.005)	-0.004 (0.003)	-0.031*** (0.006)	-0.011*** (0.003)
Male	-0.003 (0.002)	-0.000 (0.002)	-0.013** (0.006)	-0.000 (0.002)	-0.005** (0.002)	-0.003 (0.003)	-0.001 (0.001)
<i>Socio Economic Category</i> (Ref: Higher Occupation)							
Farmer	0.028** (0.010)	0.001 (0.004)	0.029* (0.015)	-0.007 (0.005)	-0.004 (0.006)	-0.019** (0.006)	-0.003 (0.004)
Craftsman	0.016** (0.006)	0.008** (0.004)	-0.009 (0.008)	-0.004 (0.005)	-0.002 (0.005)	-0.004 (0.007)	0.005 (0.004)
Intermediate	-0.009** (0.004)	0.002 (0.003)	-0.001 (0.007)	-0.009** (0.004)	-0.004 (0.003)	-0.009** (0.005)	-0.005* (0.003)
Employee	-0.006* (0.004)	0.007** (0.003)	0.016* (0.009)	-0.007 (0.004)	-0.006 (0.004)	-0.008 (0.005)	-0.006** (0.003)
Factory worker	-0.007* (0.004)	0.006* (0.003)	0.008 (0.007)	-0.012** (0.004)	-0.009** (0.004)	-0.017*** (0.005)	-0.008** (0.003)
Retired	-0.021* (0.011)	-0.010 (0.009)	-0.047 (0.037)	-0.019* (0.011)	-0.016 (0.011)	-0.017 (0.016)	-0.006 (0.010)
Other inactive	-0.019* (0.011)	-0.021** (0.010)	-0.017 (0.039)	-0.024* (0.013)	-0.012 (0.011)	-0.024 (0.018)	-0.012 (0.011)
Intercept	-0.053 (0.074)	0.131** (0.060)	-0.225 (0.185)	-0.096 (0.098)	0.054 (0.071)	-0.142 (0.108)	0.043 (0.059)
<i>Département</i> f.e.	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year f.e.	Yes	Yes	Yes	Yes	Yes	Yes	Yes
N	63586.000	52453.000	7891.000	35665.000	63588.000	52428.000	52455.000
Adj. R <sup>2</sup>	0.007	0.014	0.035	0.031	0.015	0.047	0.032

Standard errors clustered at the IRIS level are reported in parentheses. \* p&lt;0.10, \*\* p&lt;0.05, \*\*\* p&lt;0.001

**Table 2.7:** Individual Victimization: Full Specification (To be continued)

	Robbery (1)	Theft (2)	Assault (3)
<b>Neighborhood characteristics</b>			
Share of Immigrants	0.004 (0.007)	0.022* (0.011)	0.001 (0.010)
Median Income (log)	-0.004** (0.002)	-0.003 (0.004)	0.008** (0.003)
Unemployment rate	0.013 (0.009)	0.000 (0.014)	0.048*** (0.014)
Share Monoparental	0.000 (0.006)	0.018* (0.011)	-0.000 (0.011)
Share Public Housing	-0.008** (0.003)	-0.011** (0.004)	0.006 (0.004)
Share Recent Movers	0.013* (0.007)	0.023** (0.010)	0.021** (0.010)
Share 14-18 y.o.	0.010 (0.023)	-0.030 (0.032)	-0.042 (0.030)
City Density (log)	0.001** (0.000)	0.001 (0.001)	-0.001 (0.000)
<i>Type of buildings in the neighborhood (Ref: Dispersed houses)</i>			
Houses Lot / in cities	-0.000 (0.000)	-0.000 (0.001)	0.001 (0.001)
Apartment block (city)	0.002 (0.001)	0.004** (0.002)	0.004** (0.002)
Apartment block (suburbs)	-0.001 (0.001)	0.002 (0.002)	0.006** (0.002)
Mixed	-0.000 (0.001)	0.001 (0.002)	-0.001 (0.002)
<i>Size of the Urban Unit (Ref: Rural Areas)</i>			
Less than 50,000	-0.001* (0.001)	-0.000 (0.001)	0.000 (0.001)
More than 50,000	-0.000 (0.001)	-0.000 (0.002)	0.003* (0.002)
Paris Urban Unit	0.001 (0.002)	-0.002 (0.004)	0.000 (0.004)
<i>Household income (Ref: bottom 30%)</i>			
Middle 30%	-0.001 (0.001)	-0.001 (0.001)	-0.003** (0.001)
Top 30%	-0.001* (0.001)	-0.002 (0.001)	-0.005*** (0.001)
<i>Household Ownership Status (Ref: Owner)</i>			
Rent in private market	-0.000 (0.001)	-0.001 (0.001)	0.004*** (0.001)
Rent in public housing	0.000 (0.001)	0.001 (0.002)	0.004** (0.002)
Other	0.002 (0.001)	-0.001 (0.002)	0.004* (0.002)
<i>Number of children in hh</i>	0.000 (0.000)	-0.000 (0.000)	0.002*** (0.001)

Standard errors clustered at the IRIS level are reported in parentheses. \* p<0.10, \*\* p<0.05, \*\*\* p<0.001



Table 2.7: Individual Victimization: Full Specification (Continued)

	Robbery (1)	Theft (2)	Assault (3)
<i>Nationality</i> (Ref: Native French)			
Naturalized French	-0.000 (0.001)	0.000 (0.002)	-0.004** (0.002)
EU 15	0.000 (0.002)	-0.006** (0.002)	-0.003 (0.003)
Other EU (after 2004)	0.003 (0.012)	-0.006 (0.012)	-0.004 (0.012)
Maghrebian	-0.003 (0.002)	-0.004 (0.004)	-0.009** (0.003)
Other African	0.004 (0.006)	-0.001 (0.007)	-0.008 (0.006)
Other nationality	-0.001 (0.004)	0.002 (0.006)	-0.007 (0.005)
<i>Employment Status</i> (Ref: Employed)			
Unemployed	0.003** (0.001)	0.003 (0.002)	0.004** (0.002)
Inactive	0.005 (0.004)	0.009* (0.005)	-0.001 (0.006)
<i>Age</i> (log)	-0.004*** (0.001)	-0.006*** (0.002)	-0.008*** (0.002)
<i>Male</i>	0.001 (0.001)	-0.001 (0.001)	0.002** (0.001)
<i>Socio Economic Category</i> (Ref: Higher Occupation)			
Farmer	0.001 (0.001)	0.028*** (0.007)	-0.001 (0.002)
Craftsman	0.006** (0.002)	0.008** (0.003)	0.006** (0.003)
Intermediate	0.001 (0.001)	0.001 (0.002)	0.002 (0.002)
Employee	0.001 (0.001)	0.001 (0.002)	0.003* (0.002)
Factory worker	-0.000 (0.001)	0.001 (0.002)	0.002 (0.002)
Retired	-0.002 (0.004)	-0.004 (0.005)	0.003 (0.006)
Other inactive	-0.002 (0.004)	-0.004 (0.005)	0.005 (0.006)
Intercept	0.049** (0.022)	0.057 (0.039)	-0.040 (0.035)
<i>Département</i> f.e.	Yes	Yes	Yes
<i>Year</i> f.e.	Yes	Yes	Yes
N	63655.000	63653.000	63649.000
Adj. R <sup>2</sup>	0.005	0.005	0.008

Standard errors clustered at the IRIS level are reported in parentheses. \* p<0.10, \*\* p<0.05, \*\*\* p<0.001

**Table 2.8:** Determinants of victimization: including large neighborhood fixed effects

	Household victimization										Individual victimization			
	Burglary (1)	Car Theft (2)	Motorbike Theft (3)	Bike Theft (4)	Home Vandalism (5)	Car Vandalism (6)	Theft of car objects (7)	Robbery (8)	Theft (9)	Assault (10)				
Share of Immigrants	-0.118** (0.043)	0.024 (0.038)	0.049 (0.165)	0.079 (0.056)	0.024 (0.041)	-0.038 (0.063)	0.036 (0.035)	-0.022* (0.013)	0.017 (0.021)	0.006 (0.020)				
Median Income (log)	-0.026* (0.014)	-0.001 (0.012)	0.074 (0.051)	0.024 (0.018)	-0.012 (0.013)	-0.006 (0.020)	0.009 (0.011)	-0.004 (0.004)	0.004 (0.007)	0.011* (0.006)				
Unemployment rate	0.105** (0.044)	0.007 (0.039)	0.181 (0.169)	0.059 (0.056)	-0.003 (0.042)	0.216*** (0.064)	-0.017 (0.035)	0.026* (0.013)	-0.026 (0.022)	0.041** (0.020)				
Share Monoparental	-0.027 (0.035)	0.061** (0.030)	-0.032 (0.126)	0.070 (0.043)	-0.018 (0.034)	0.152** (0.050)	0.048* (0.027)	0.009 (0.011)	0.016 (0.017)	0.021 (0.016)				
Share Public Housing	-0.008 (0.012)	-0.024** (0.011)	0.071 (0.047)	-0.021 (0.016)	0.028** (0.012)	-0.040** (0.018)	-0.004 (0.010)	-0.007* (0.004)	0.001 (0.006)	0.004 (0.006)				
Share Recent Movers	0.042 (0.031)	0.021 (0.026)	0.021 (0.115)	0.060 (0.038)	0.089** (0.030)	0.104** (0.044)	0.050** (0.024)	0.014 (0.010)	0.036** (0.015)	0.003 (0.014)				
Share 14-18 y.o.	0.098 (0.096)	0.087 (0.082)	-0.099 (0.339)	0.040 (0.115)	0.095 (0.092)	0.141 (0.135)	0.076 (0.074)	-0.027 (0.029)	-0.060 (0.047)	-0.038 (0.044)				
N	63,586	52,453	7,891	35,665	65,588	52,428	52,455	63,655	63,653	63,649				

In addition to the IRIS characteristics reported here, the regressions include the following controls: city density (log), type of neighborhood, household characteristics, household head (respectively individual characteristics) in the household (respectively individual) victimization regressions. They also control for year and large neighborhood fixed effects. \* p<0.10, \*\* p<0.05, \*\*\* p<0.001

**Table 2.9:** Correlation between IRIS characteristics

	Share Immigrants	Median Income	Unemployment Rate	Share Single Parent	Share Public Housing	Share Recent Movers	Share 14-18 y.o.
Share Immigrants	1.000						
Median Income	-0.255	1.000					
Unemployment Rate	0.528	-0.645	1.000				
Share Single Parent	0.496	-0.457	0.671	1.000			
Share Public Housing	0.505	-0.527	0.661	0.697	1.000		
Share Recent Movers	0.066	0.049	0.126	0.191	-0.104	1.000	
Share 14-18	0.128	-0.158	0.225	0.161	0.292	-0.151	1.000

These correlations are obtained using one observation per IRIS per year. The numbers in the columns of the first line correspond to the numbers in the lines of the first column. For instance, (1) stands for the Share of Immigrants, so that "-0.255" is the correlation between the share of immigrants in an IRIS in a given year and the median income in the same IRIS and year.

**Table 2.10:** Household victimization: unemployment rate in adjacent neighborhoods.

	Household victimization					Individual victimization				
	Burglary Theft (1)	Car Theft (2)	Motorbike Theft (3)	Bike Theft (4)	Home Vandalism (5)	Car Vandalism (6)	Theft of car objects (7)	Robbery (8)	Theft (9)	Assault (10)
<b>No other controls</b>										
<i>u<sub>iris</sub></i>	0.005 (0.025)	0.049** (0.024)	0.273** (0.115)	0.116** (0.038)	0.096*** (0.026)	0.353*** (0.042)	0.055* (0.023)	0.016* (0.009)	-0.012 (0.013)	0.051*** (0.013)
<i>u<sub>iris</sub></i> 1	0.104* (0.058)	-0.009 (0.054)	0.145 (0.266)	0.054 (0.086)	0.034 (0.059)	0.133 (0.095)	0.286*** (0.052)	0.016 (0.020)	0.041 (0.031)	-0.022 (0.029)
<i>u<sub>iris</sub></i> 2	-0.028 (0.089)	0.049 (0.082)	-0.372 (0.362)	-0.093 (0.127)	-0.055 (0.091)	-0.211 (0.143)	0.140* (0.079)	-0.005 (0.031)	0.014 (0.047)	-0.035 (0.044)
N	45,736	34,949	4,538	21,968	45,733	34,923	34,950	45,744	45,744	45,739
<b>Full specification</b>										
<i>u<sub>iris</sub></i>	0.081** (0.028)	0.021 (0.027)	0.211 (0.128)	0.021 (0.042)	0.124*** (0.029)	0.235*** (0.047)	-0.006 (0.026)	0.013 (0.010)	-0.018 (0.015)	0.029** (0.014)
<i>u<sub>iris</sub></i> 1	0.144** (0.062)	0.035 (0.058)	-0.045 (0.283)	0.030 (0.090)	0.024 (0.063)	0.112 (0.100)	0.226*** (0.055)	0.018 (0.022)	0.051 (0.032)	-0.011 (0.031)
<i>u<sub>iris</sub></i> 2	-0.013 (0.094)	0.052 (0.087)	-0.475 (0.379)	0.166 (0.132)	0.063 (0.096)	-0.189 (0.150)	0.001 (0.083)	0.004 (0.033)	0.028 (0.049)	-0.052 (0.047)
N	41,781	32,191	4,264	20,502	41,780	32,170	32,194	41,826	41,826	41,822

In these regressions, the sample has been restricted to observations living in actual IRISes, so that the remaining IRISes are homogeneous in terms of size. "*u<sub>iris</sub>*" is the local (IRIS) unemployment rate, "*u<sub>iris</sub>* 1" the average unemployment rate in the first ring of adjacent IRISes, and "*u<sub>iris</sub>* 2" the average unemployment rate in the second ring of adjacent IRISes. In the second set of results, the controls are environment characteristics (city density (log), type of neighborhood, size of the urban unit), year fixed effects, household characteristics (income, ownership status, number of children) and individual characteristics, corresponding to the surveyed individual for individual victimization and to the household head for household victimization (nationality, age (log), gender, employment status, socio-economic category). They also include large neighborhood fixed effects. \* p<0.10, \*\* p<0.05, \*\*\* p<0.001

Table 2.11: Unemployment rate in adjacent neighborhoods, excluding Paris, Lyon and Marseille

	Household victimization						Individual victimization			
	Burglary Theft (1)	Car Theft (2)	Motorbike Theft (3)	Bike Theft (4)	Home Vandalism (5)	Car Vandalism (6)	Theft of car objects (7)	Robbery (8)	Theft (9)	Assault (10)
<i>u<sub>iris</sub></i>	0.074** (0.029)	0.042 (0.028)	0.243* (0.129)	0.003 (0.042)	0.126*** (0.030)	0.227*** (0.048)	0.002 (0.026)	0.017* (0.010)	-0.014 (0.014)	0.026* (0.015)
<i>u<sub>iris</sub> 1</i>	0.164** (0.064)	0.057 (0.059)	0.031 (0.288)	0.008 (0.090)	0.053 (0.066)	0.104 (0.103)	0.213*** (0.056)	0.017 (0.021)	0.037 (0.031)	-0.009 (0.032)
<i>u<sub>iris</sub> 2</i>	-0.033 (0.097)	0.073 (0.088)	-0.308 (0.377)	-0.216 (0.132)	-0.076 (0.100)	-0.199 (0.153)	0.011 (0.083)	0.023 (0.032)	0.025 (0.047)	-0.052 (0.049)
N	38,409	30,409	3,945	19,525	38,405	30,390	30,412	38,451	38,450	38,448

In these regressions, the sample has been restricted to observations living in actual IRISes, so that the remaining IRISes are homogeneous in terms of size. "*u<sub>iris</sub>*" is the local (IRIS) unemployment rate, "*u<sub>iris</sub> 1*" the average unemployment rate in the first ring of adjacent IRISes, and "*u<sub>iris</sub> 2*" the average unemployment rate in the second ring of adjacent IRISes. The controls are environment characteristics (city density (log), type of neighborhood, size of the urban unit), year fixed effects, household characteristics (income, ownership status, number of children) and individual characteristics, corresponding to the surveyed individual for individual victimization and to the household head for household victimization (nationality, age (log), gender, employment status, socio-economic category). They also include large neighborhood fixed effects. \* p<0.10, \*\* p<0.05, \*\*\* p<0.001

## **Part II**

# **Diversity and Employment**



## Chapter 3

# Diversity and Employment Prospects: Do Neighbors Matter?

### 1 Introduction

Western economies are facing intensified flows of immigration due to trade development and economic integration, and consequently have to cope with increasingly mixed populations. This feature is particularly salient in the European Union following the recent enlargement process. The economic and social implications of higher heterogeneity are therefore central issues. In particular, public opinion is generally hostile to immigrants who are often perceived as a threat to job security and wages, although this is not clear from empirical research. In the latest paper on the topic, [Ottaviano and Peri \(2012\)](#) adopt a general equilibrium approach and show that the massive immigration to the US over the 1990-2004 period actually increased natives wages, contradicting the influential paper by [Borjas \(2003\)](#). Similarly, [Manacorda et al. \(2012\)](#) show that the large increase in immigration to the UK over the previous 30 years had no effect on natives' wages, due to imperfect substitutability between natives and immigrant workers. In contrast to the large literature studying the economic impact of immigration on natives, papers looking at the labor market effect of diversity *per se* are scarce. Using data on US cities, [Ottaviano and Peri \(2006\)](#) jointly estimate a wage and a rent equations and find that diversity is positively associated to both variables. They conclude that diversity has a net positive impact on US-born workers' productivity. Using a similar setting, [Prarolo et al. \(2009\)](#) replicate these results for European regions. To the best of my knowledge, these are the only two papers in the diversity literature dealing with labor market outcomes, although not directly with employment.

This paper intends to fill this gap by assessing the impact of local diversity on individuals' employment prospects. It asks the following question: to what extent people living in more heterogeneous neighborhoods have different employment probabilities than those



living in more homogeneous areas? In other words, this work looks at how individuals cope with increasing levels of diversity, and in particular how this affects their employment prospects. It is relevant in the current context of high unemployment, especially in high immigration countries such as France, Italy and more recently Spain. At the micro level, if different ethnic or cultural groups are hermetic to each other, in the sense that no interaction takes place across groups, then diversity can act as a barrier to communication and in particular to job information transmission. Given the importance of personal networks in the job search process (see [Ioannides and Datcher Loury, 2004](#)), diversity could therefore reduce the chance of finding a job. On the other hand, if communication across groups is not an issue, then mixing people conveying non-redundant pieces of information (e.g. due to different backgrounds) can certainly improve employment prospects. At a more aggregate level, diversity can affect employment probability through its impact on productivity, which is ambiguous as well. On the bright side, diversity can be beneficial to productivity due to complementarity in workers' skills (see [Lazear, 1999b](#); [Alesina and La Ferrara, 2005](#)). On the downside, heterogeneity can hinder productivity by preventing social capital formation ([Coleman, 1988](#)).<sup>1</sup>

As we see from this brief discussion, the question of the role of diversity on employment is not trivial. By addressing the issue of diversity and employment at a local level, I intend to show how diversity directly affects workers as individuals, in addition to impacting them indirectly via firms' productivity. I am able to deal with this question using detailed geolocalized French employment data that allow me to measure diversity at very low geographic levels. More precisely, I measure diversity using several definitions of origins and at various geographic levels, so as to understand as precisely as possible the mechanisms lying behind the diversity-employment relationship. In addition, I adopt various identification strategies in order to bypass the endogeneity issue that likely flaws any estimate of the impact of diversity. These three methodological elements allowing me to answer the central question of this paper are detailed below.

First, the level of diversity is measured at two different geographic levels. As discussed above, diversity could impact employment prospects locally through networks and on a larger scale through productivity. In order to account for both effects, I compute diversity at a very local neighborhood level and at the local labor market level. As far as I know, it is the first time that various geographic scales for diversity are simultaneously considered. In addition, this and [Algan et al. \(2013\)](#) are the first studies conducted at

---

<sup>1</sup>Not only is the impact of diversity on productivity unclear, but the impact of productivity on employment is ambiguous as well: [Nordhaus \(2005\)](#) finds that more rapid productivity growth leads to increased rather than decreased employment in manufacturing, a sector that recently experienced a large employment decline. On the contrary, [Michelis et al. \(2013\)](#) find a strong negative relationship between TFP growth and labor input.

such disaggregated levels. The results reveal that employment probability is negatively correlated with neighborhood diversity, but positively correlated with employment zone diversity, suggesting a negative effect through networks and a positive one through productivity.

Second, I rely on three alternative definitions of origins to measure diversity, namely nationality, birth country, and parents' origins, while the existing literature mostly relies on ethno-linguistic and religious groups. The measure based on parents' origin encompasses first- and second-generation immigrants and is therefore more closely related to the standard ethnic classification of individuals. By contrast, defining kinship according to nationality introduces the notion of immigrants' assimilation through naturalization. This distinction allows me to draw conclusions on whether the cultural or the ethnic dimension of diversity prevails. An important finding of this paper is that diversity based on nationality has a larger impact than diversity based on birth country, which is itself more relevant than diversity based on parents' origins, suggesting a prominent role of cultural over ethnic diversity.

Third, I tackle the endogeneity issue that is pervasive in the literature on ethnic diversity. An important concern is that individuals have a preference for living close to their co-ethnics and thus tend to gather along ethnic lines, biasing any measure of the effect of diversity. Another issue is that of reverse causality that can arise if immigrants decide to settle in more economically dynamic areas. To deal with the endogeneity of local neighborhoods diversity, I follow [Bayer et al. \(2008\)](#) and assume that although households are able to select the precise area in which they want to live, they are, however, unable to pinpoint an exact neighborhood within this given area. Therefore, after controlling for sorting in a larger area, the assignment of individuals to a specific neighborhood is essentially random and provides a useful source of exogenous variation to identify the effect of diversity. As it turns out, the effect of local diversity on employment is corrected downward, *i.e.* becomes more negative. The endogeneity of employment zones diversity is handled through a more traditional instrumental variable approach, where two different instruments are proposed. Following [Card \(2001\)](#) and [Saiz \(2007\)](#), I first construct the predicted level of diversity in each employment zone based on the past distribution of the various origin group across employment zones, and on the current number of individuals from each origin in France. The second and more innovative instrument is the level of diversity within the public housing tenants of the employment zone. It builds on [Algan et al. \(2013\)](#) who show that the allocation of households across public housing units in France does not take their origins or their preference for diversity into account, so that public housing diversity can be considered as exogenous. Interestingly, once employment zone diversity is instrumented using any of these two variables, its positive relationship with employment is driven down to zero, confirming the intuition that the previous effect

was actually driven by selection.

The rest of the paper proceeds as follows. Section 2 discusses more extensively the channels through which diversity can affect employment prospects. Section 3 presents the data and the various measures of diversity. The relationship between diversity and employment status is investigated in Section 4. Section 5 corrects for endogeneity. Results are interpreted in Section 6. Section 7 concludes.

## 2 Diversity and employment prospects

The interest in the effect of diversity on economic performance and social peace has been rooted in economic research since the seminal paper by [Easterly and Levine \(1997\)](#) showing that Africa's high levels of ethnic diversity help understand its "tragic growth performance". The subsequent literature covers a very broad set of issues. Diversity is generally found to reduce public good provision, because the threat of social sanction to punish defectors is not credible across groups, or because different groups do not share the same preferences and cannot agree on the type of public good to be produced. This result holds in developing countries and developed countries alike (see [Miguel and Gugerty \(2005\)](#) for Kenya, [Alesina et al. \(1999\)](#) for the US and [Algan et al. \(2013\)](#) for France).<sup>2</sup> Another trend of the literature focuses rather on the social impact of diversity, and shows that it is associated with lower participation to civic life and community activities ([Alesina and La Ferrara, 2000](#); [Costa and Kahn, 2003](#)) and reduced trust ([Alesina and La Ferrara, 2002](#)).

The present paper focuses on employment and is therefore more closely related to the branch of the literature that studies productivity. At the macro level, diversity can affect employment through its effect on productivity. A large part of the literature supports the idea that diversity has a positive impact on productivity related to skills complementarity, dominating the negative effects linked to coordination issues. Indeed, workers from different origins are more likely to have been exposed to diverse cultures and distinct school systems (especially if they come from different countries), acquiring various skills and learning different approaches to the same problem, so that their collaboration can increase productivity and facilitate innovation. More formally, [Hong and Page \(2001\)](#) develop a model showing that team work may benefit more from low-skilled but cognitively diverse workers than from homogeneous high-skilled workers. In a different theoretical setting, [Lazear \(1999b\)](#) shows that when multicultural workers are complementary in the

---

<sup>2</sup>An exception is [Glennerster et al. \(2013\)](#) who do not find any particular effect of diversity in Sierra Leone villages.

sense that they can exchange non-redundant and relevant information, the benefits from diversity offset its costs (e.g. barriers to communication).

Several recent papers also bring empirical support to the beneficial impact of diversity on productivity and economic performance more generally. Using data from 160 metropolitan areas in the US, [Ottaviano and Peri \(2006\)](#) jointly estimate a wage and a rent equations and find that diversity, measured in terms of birth countries, is positively associated to both variables. These results are robust to the inclusion of many confounding factors proxying for productivity and amenity shocks across cities, as well as to the instrumentation of diversity to correct for endogeneity. They conclude that diversity has a net positive impact on US-born workers' productivity. A similar methodology is adopted by [Prarolo et al. \(2009\)](#) who reach the same conclusion for European regions. Finally, [Alesina et al. \(2013\)](#) investigate the relationship between birth country diversity and economic development in a cross-section of countries. Potential endogeneity due to reverse causality is addressed through instrumental variable estimation. The authors compute a predicted measure of immigrants diversity by estimating a gravity model based on exogenous geographic and cultural bilateral variables. They find that while standard ethnolinguistic fractionalization is detrimental to economic success, the impact of diversity in terms of birth countries is positive, especially in more developed countries.<sup>3</sup>

At a more micro level, diversity can affect individuals' employment prospects through the channel of networks and job information transmission. There is considerable evidence that information transmission plays a key role on the labor market.<sup>4</sup> Many empirical studies conducted over various time periods and on diverse countries agree that relying on friends and family is a very popular job search method and that on average half of jobs are found through social networks (see for instance [Corcoran et al., 1980](#); [Granovetter, 1995](#); [Holzer, 1988](#); [Wahba and Zenou, 2005](#)). Theoretically as well, [Calvó-Armengol and Jackson \(2004\)](#) show that employment probability increases both with the number of links an agent has, and with the employment rate in the individual's network. In particular, several papers focus on the role of ethnic and immigrant networks. A recent paper by [Battu et al. \(2011\)](#) shows that ethnic minorities in the UK rely extensively on personal networks when searching for a job, although this does not necessarily lead to better employment prospects. The sociology literature also emphasizes the importance of ethnic networks in business relations and entrepreneurship, through an increased capacity to cooperate due to common language and values ([Light and Rosenstein, 1995](#); [Light, 2005](#)).

Because communication across ethnic groups may be hindered by a tendency to self-segregate, by different religious beliefs and culture, and above all by differences in the

<sup>3</sup>This paper also provides a comprehensive review of the literature on the costs and benefits of diversity.

<sup>4</sup>[Ioannides and Datcher Loury \(2004\)](#) and [Ioannides and Topa \(2010\)](#) provide comprehensive surveys on the topic.

languages spoken, diversity may prevent network formation and information transmission, thus having a negative impact on individuals' labor market performances. This effect could be amplified if diversity exists at the neighborhood level, as networks tend to be very local (see for instance [Wellman \(1996\)](#)). In particular, a few recent studies have shown that local social interactions within neighborhoods do affect employment and wage outcomes. For instance, [Weinberg et al. \(2004\)](#) show that a one standard deviation increase in neighborhood employment is associated with a 6.1% increase in annual hours worked for adult males on average. [Bayer et al. \(2008\)](#) estimate that living in the same block increases by more than 33% the probability to work at the same location. In a paper dealing explicitly with ethnic networks, [Patacchini and Zenou \(2012\)](#) show that the individual probability of finding a job increases with the number of ties, but that the magnitude of the effect decreases with distance. To summarize, if individuals are unable to create social ties within their neighborhood because they live in a diverse environment, this might hinder their ability to search and find job through the network.

### 3 Data and descriptive statistics

The main dataset used in this paper is the French Labor Force Survey (*Enquête Emploi*, INSEE, hereafter the LFS), which has been conducted quarterly since 2003. One sixth of the sample is renewed each quarter, so that the survey takes the form of a quasi-panel, as each household is surveyed for six consecutive waves before leaving the sample. Each wave of the survey comprises about 72,000 respondents aged 15 year-old or older. The sampling strategy of the LFS makes it particularly valuable for studying neighborhood effects. To put it simply, France is divided into areas made up of twenty homes on average. The sample is then drawn from a random selection of these areas, in which all the households will be surveyed.<sup>5</sup> As a consequence, I am able to characterize the immediate neighborhood of each surveyed individual. In particular, it is possible to measure the precise level of diversity and the unemployment rate within these twenty-household units.

The LFS contains all the relevant information about individuals' labor market situation: employment status, wage, type of contract, tenure, job search methods and socio-economic category. It also provides detailed individual information, such as age, gender, education and marital status. Individuals' ethnic background can be inferred from their birth country, their nationality, and their parents' origins. Because I am interested in individuals' employment status I restrict the sample to working-age individuals (16 to 65 years old) and I drop the students. In addition, because the information about parents that is used to characterize individuals' origin is mostly missing before 2007, I restrict my

---

<sup>5</sup>Refer to [INSEE documentation](#) for more details on the sample composition and selection.

sample to the 2007 to 2010 period. Table 3.1 summarizes the main employment-related individual characteristics for this sub-sample. I then define three different measures of origins. Two measures are simply based on individuals' nationality and country of birth, while a third one combines the nationality and birth country of both individuals' and their parents'. The survey contains two variables with 28 categories describing individuals' nationality and birth country respectively, that are used as two different indicators of origin of their own.<sup>6</sup> The information about parents' nationality and birth country is given by four variables with 9 categories, for each parent's (mother's and father's) nationality *at birth* and birth country.<sup>7</sup> This enables me to build a measure of origin that takes second generation immigrants into account. More precisely, for an individual to be sorted in a given group, it must be the case that at least one of her or his parents belongs to this group. For instance, a French citizen born in France but whose parents were born with a Maghrebian nationality is allocated to the *Maghreb* group. Note that this third type of classification of origins can contain at most 9 categories given the way parents' origins are defined. As we will see in the next paragraph, measure of diversity is sensitive to the number of categories used. The alternative measures of origins should therefore contain the same number of categories to allow comparison. In addition, some of the considered groups, such as *Northern European*, represent such small shares of the population living in France that I decide to aggregate them further. I eventually build three measures of origins based on nationality, birth country and individual's and parents' origins that are divided into the following 6 categories: France, Southern Europe, other European countries, Maghreb, other African countries and rest of the world. Table 3.2 describes the sample along the three dimensions of origins.

Using these various classifications of origins, I am able to compute three corresponding measures of diversity. The level of diversity in a given area is measured using a standard fractionalization index (see Alesina et al. (2003) for an extensive description). It reflects the probability that two randomly drawn individuals from the population belong to two different groups:

$$DIV_j = \sum_{i=1}^N s_{ij} (1 - s_{ij}) = 1 - \sum_{i=1}^N s_{ij}^2 \quad (3.1)$$

where  $s_{ij}$  is the share of individuals from group  $i$  ( $i=1, \dots, N$ ) in geographic area  $j$ . This

<sup>6</sup>The 28 possible nationalities or birth countries correspond to the following countries: France, Algeria, Tunisia, Morocco, Other African countries, Vietnam / Laos / Cambodia, Italy, Germany, Belgium, Netherlands, Luxembourg, Ireland, Denmark, United Kingdom, Greece, Spain, Portugal, Switzerland, Austria, Poland, Yugoslavia, Turkey, Norway, Sweden, other European countries (including former USSR), USA / Canada, Latin American countries and other countries.

<sup>7</sup>The 9 categories correspond to France, Northern Europe, Southern Europe, Eastern Europe, Maghreb, rest of Africa, Middle East, Vietnam / Laos / Cambodia and rest of the world.

index takes its minimum at 0 when the population living in area  $j$  is fully homogeneous, and it converges to 1 as the population heterogeneity increases. Note that  $DIV_j$  can increase for two reasons: it will increase with the number of groups, and it will increase the more equal the size of the groups.

In the context of this paper, the considered groups are alternatively nationalities, birth countries and parents' origins, as defined above. Looking at various measures of diversity enables me to capture various dimensions of diversity. I argue that parents' origin diversity is the best proxy for *ethnic* diversity as it is more likely to reflect color of skin. For instance, a person whose parents are Senegalese is very likely black, even though s/he is French and born in France. This measure of diversity is therefore the closest to the ethnic diversity used in the literature and in particular in US studies. On the other hand, diversity based on nationality reflects *cultural* rather than *ethnic* diversity. Indeed, two individuals sharing the same nationality are more likely to speak a common language and to share other cultural traits. This can be true for two native individuals, as well as for naturalized French who spent some time in France, learning French before being naturalized. Given what precedes, it is then reasonable to rank birth country diversity in-between.

Let me now present the various types of geographic areas for which I assess diversity. The first type of area considered is the local neighborhood made of around twenty contiguous households (*i.e.* the LFS sampling unit). Measuring diversity at such a low geographic level enables me to indirectly account for local interactions between immediate neighbors and to test whether diversity acts as a barrier to job information transmission. The second type of area used to measure diversity is the employment zone, which is a local labor market. More precisely, it is a geographical area within which most of the labor force lives and works, and in which establishments can find most of the labor force necessary to occupy the offered jobs. To give an idea of the level of aggregation, there are about 300 employment zones in metropolitan France. Measuring diversity at a level corresponding to a consistent local labor market is particularly useful to challenge the idea that diversity is beneficial at more aggregate levels through increased productivity.

Table 3.3 describes the levels of diversity prevailing in individuals' neighborhood and employment zone. It is immediate to see that diversity is the lowest when measured in terms of nationalities, and the highest when computed based on individuals' and parents' origins, which is not surprising. Consider for instance an area made of three French individuals, one born in France from French parents (e.g. native French), one born in France from Vietnamese parents (e.g. second generation immigrant), and one born in Morocco from Moroccan parents (e.g. first generation immigrant). This population is completely homogeneous ( $DIV_j=0$ ) if we consider the individuals' nationality only. However, diver-

sity is larger (0.44) once birth countries are taken into account, and even more (0.67) once parents' origins are considered. Note also that employment zone diversity is on average slightly larger than neighborhood diversity, but that the latter takes more extreme values (larger maxima) than the former.

## 4 Analysis

In this section, diversity is considered as exogenous, and its impact on employment status is estimated through the following equation:

$$EMP_{ijt} = \alpha + \beta DIV_{jt} + \eta Z_{jt} + \gamma X_{it} + \phi_g + \phi_t + \varepsilon_{ij} \quad (3.2)$$

where  $EMP_{ijt}$  is the employment status of individual  $i$  living in area  $j$  at time  $t$ ,  $DIV_{jt}$  is the level of diversity in area  $j$  at time  $t$ ,  $Z_{jt}$  is a vector of characteristics of area  $j$  at time  $t$ , and  $X_{it}$  is a set of individual control variables. I also include geographic fixed effects  $\phi_g$ , generally *départements* fixed effects, along with time fixed effects  $\phi_t$  (quarter dummies). Finally,  $\varepsilon_{ij}$  is an error term. The main coefficient of interest is  $\beta$ . Individuals' employment status can either be *employed*, *unemployed* or *inactive*. In what follows, the dependent variable is a dummy variable equal to 1 if the individual is employed, and 0 otherwise (unemployed or inactive).<sup>8</sup> The results presented in this section derive from OLS estimates, with robust standard errors clustered at the neighborhood level. Logistic regressions lead to qualitatively similar results, but OLS estimates are displayed for simplicity. At some point, multinomial logit estimates will be presented, to take into account the three possible employment statuses, without altering the main result.

The various measures of diversity (nationality-based, birth country-based and parents' origin-based) are included separately, in different regressions. However, both neighborhood and employment zone diversity (based on the same origin groups) are included in a given regression. The set of individual controls  $X_{it}$  comprises the standard socio-demographic variables: age (quadratic form), gender, origin, education, socio-economic category and potential experience (quadratic function). The origin variable can take on 6 different values: France, South Europe, rest of Europe, Maghreb, rest of Africa and rest of the world. Specifically, I alternatively include nationality, birth country and parents-based origin indicators when diversity is measured based on nationality, birth country and parents' origin respectively. The education variable describes the highest degree obtained by the individual, which can be one of the following: *No diploma*, end of junior high

<sup>8</sup>An alternative dummy variable considered takes value 1 if the individual is employed, and 0 if s/he is unemployed, letting aside inactive individuals. Using this alternative dependent variable does not significantly alter the estimated coefficients.



school degree (9th grade) (*BEPC*), early vocational training degree (*CAP*), *Technical degree*, technical or vocational senior high school degree (*Tech. & Pro. Baccalauréat*), general senior high school degree (*Baccalauréat*), *Undergraduate diploma* (two years after the *Baccalauréat*), *Bachelor's degree* (three years after the *Baccalauréat*), *Graduate diploma* (four years after the *Baccalauréat*), and higher degree (*Master's & PhD*).<sup>9</sup> There are 6 possible socio-economic categories: *Farmer, Craftsman or Shopkeeper, Executive or other high position, Intermediate occupation, Employee and (Factory) worker*. Potential experience is measured as the number of years since the highest degree was awarded. Finally, I also control for the unemployment rate prevailing in an individual's neighborhood, so as to account for peer effects. Note that a given individual is excluded when computing the unemployment rate in her or his neighborhood.

Table 3.4 presents the estimates obtained regressing the employment dummy on neighborhood and employment zone diversity based on nationality. Each column corresponds to a different specification, starting from no control in column 1 to the full set of controls in column 4. The sample is restricted to the non-student, working-age population (16 to 65 year-old individuals). In addition, I keep the first observation of each individual, so that an individual appears only once in the sample.<sup>10</sup> The estimates reported in the first column directly reveal that local diversity is negatively associated to the probability of being employed, while the correlation with employment zone diversity is positive.<sup>11</sup> These results are in line with the idea that diversity can have an adverse effect on job finding locally by preventing communication, but that at a more aggregate level, diversity has a positive effect on productivity and hence on employment probability. An alternative interpretation for employment zone diversity could be that when diversity is high, individual's networks lie in a larger area than their direct neighborhood. The estimates presented in column 2 are obtained controlling for the set of individual characteristics. The coefficients for the two measures of diversity are significantly reduced (in absolute terms), but we still have a negative coefficient for neighborhood diversity and a positive one for employment zone diversity. Turning to individuals' nationality, it is quite interesting to see that South European perform better than French in terms of employment, while individuals of any other nationality are more likely unemployed or inactive than French. The positive coefficient of South European can be attributed to Portuguese whose unemployment rate is much lower than France average. Coefficients for education levels, socio-economic categories, gender, age and experience all have the expected signs. Column 3 adds quarters and *départements* fixed effects to the previous specification. The most notable change is

<sup>9</sup>The terms in italic characters stand for the labels that are written in the tables.

<sup>10</sup>The panel aspect of the data is ignored for the time being.

<sup>11</sup>This is also true when local and employment zone diversity are included in separate regressions. In this case, both coefficients are smaller (in absolute terms), but are still significant at the 1% level.

for employment zone diversity which is reduced further. Finally, the results displayed in column 4 are obtained when neighborhood unemployment rate is controlled for. Including this variable significantly decreases (in absolute terms) the coefficients of diversity, especially that of neighborhood diversity. Obviously, the coefficient for local unemployment rate is strongly negative.

The estimated effects of the various types of diversity are summarized in Table 3.5. Each column corresponds to a different specification, as in Table 3.4. The first two lines display the estimates for nationality-based diversity that were already shown in the previous table. The second and third sets of estimates correspond to birth country-based and parents' origin-based diversity. As previously, we observe that the coefficient is always negative for neighborhood diversity, and always positive for employment zone diversity, no matter how diversity is measured. Note also that the negative effect of local neighborhood diversity always dominates the positive effect of employment zone diversity, suggesting that close neighbors are indeed more important than distant individuals when it comes to finding job. A substantial result emerges from comparing the coefficients for the various measures of diversity. In all specifications, the estimated effect of nationality-based diversity is larger (in absolute terms) than that of birth country-based diversity, which is also larger than parents' origin-based diversity. To put it differently, living in a context where people have different nationalities matters more for employment than living in a context where people were born in different countries, and even more than living close to people whose parents are from different origins. As discussed in the previous section, parents' origins more likely reflect ethnicity than nationality does, the latter being rather associated to common values and language. A direct interpretation of these results is therefore that neighborhood diversity lowers the probability of employment because of cultural differences, probably including language differences, rather than ethnic differences. This reinforces the intuition that diversity affects job finding by limiting communication, and in particular job information transmission between neighbors.

Finally, Table 3.6 reports the estimates for the various types of diversity obtained with multinomial logits. This enables me to look at the effect of diversity on the three possible employment statuses. The two results put forward in the preceding tables hold in this case. First, living in a more diverse neighborhood reduces one's employment prospects, while the effect of living in a more diverse employment zone goes in the opposite direction. Second, cultural diversity, embedded in diversity based on nationality, matters more than ethnic diversity, which is embedded in parents' origins diversity. The additional information contained in this table is that when neighborhood diversity is found to decrease employment probability, it corresponds to an increase in both unemployment and inactivity, the rise in the former being two to three times larger than that of latter. On the other

hand, the decrease in the probability to be unemployed or inactive associated to a larger employment zone diversity are comparable.

## 5 Results with endogenous diversity

The analysis presented in the previous section assumes that diversity is exogenous. However, there are several reasons to suspect that diversity might actually be endogenous. First, individuals who have a taste for diversity might both self-select into high diversity areas and be more able or willing to communicate with their neighbors. In this case, people living in more diverse areas would not face difficulty communicating with each other and the negative coefficient found previously would be overestimated (less negative than the true effect). Reverse causality could also be a problem for employment zone diversity if immigrants are attracted to more economically dynamic places, where jobs are more abundant. The issue of endogeneity related to the non-random location of individuals is addressed in this section.

### 5.1 Local neighborhood diversity: a local approach

The first part of this section deals with the endogeneity of local neighborhood diversity. The approach adopted here builds on the very local nature of the data. It follows [Bayer et al. \(2008\)](#) who study the role of neighbors on work location. The idea is that although households are able to select a given area in which they want to live, they are, however, unable to select a precise neighborhood within this given area. This assumption means that even if households are able to choose a given residential area, there will not be any correlation in unobserved factors affecting employment probability among individuals living in the same neighborhood *within the larger selected area*.

Let me now present a few arguments supporting this assumption. First, because the housing market is very tight, it is reasonable to think that an individual targeting a given area is very unlikely to have a choice over the precise neighborhood where s/he will end up in this area. This would indeed require that at least one housing unit satisfying the other decision criteria of the individual (e.g. size) be vacant in each of the neighborhoods within the target area at the time when the individual is looking for a new place. A second consideration is that it may be difficult for prospecting individuals to identify neighborhood-by-neighborhood variation in neighbors and contextual characteristics, prior to moving into the neighborhood. To put it differently, although the individual may have a realistic *ex-ante* view of the characteristics of the target area, it is less likely that s/he is actually able to identify differences in these characteristics across the various neighborhoods of

the area. This is particularly reasonable when the neighborhood considered consists of about twenty households. Finally, the neighborhoods studied here (the labor force survey primary units called *aires*) do not follow any administrative or official frontier. People do not know where the borders are, and more generally do not even know what an *aire* is, as it is only used as the sampling unit of the LFS. For those reasons, it is practically impossible that households purposely decide to live in a given *aire* rather than in the contiguous one.

All these arguments support the validity of the assumption that there should be no correlation in unobserved factors affecting employment among neighbors living in the same neighborhood (*aire*) within the larger targeted area. As a consequence, once we control for the characteristics of the larger area selected by the individual, the remaining spatial variance of diversity across neighborhoods within this larger area is supposed to be exogenous. This is done through the inclusion of fixed effects of larger areas than the neighborhood under study (*the aire*). Yet, one cannot know for sure which is the larger area initially selected by an individual prior to moving in a new home. I therefore run several regressions where I successively control for smaller and smaller areas fixed effects. The results are summarized in Table 3.7, which reports the coefficients for local neighborhood diversity. Each column corresponds to a separate regression, with the full specification, but including fixed effects for different larger areas. Note also that as I focus here on local neighborhood diversity, I exclude employment zone diversity from these regressions.<sup>12</sup> In the first column, I control for *départements* characteristics, as in the regressions presented in the previous section. As *départements* are quite large areas, it is very likely that individuals actually target a more precise location. Hence, I control for employment zone fixed effects in column 2. We can see that the coefficients are slightly more negative than in the *département* fixed effects specification, comforting the idea that the previous estimates of neighborhood diversity were indeed overestimated. Employment zones still being rather large areas, I go one step further and include municipalities (*i.e.* cities) fixed effects in column 3. In particular, the *arrondissements* of Paris, Lyon and Marseille are municipalities of their own. Again, the estimated effects of diversity are even lower than in the previous set of regressions, as we control for the characteristics of a more precise area in which individuals are more likely to self-select. I finally control for the characteristics of the *sector* where the individual lives, which is the smallest identifiable area after the *aire* (the 20 home neighborhoods). More precisely, a sector is

---

<sup>12</sup>The coefficients for local neighborhood diversity are slightly larger when I control for employment zone diversity, but the changes related to the inclusion of alternative larger areas fixed effects are similar. On the other hand, the coefficients for employment zone diversity lose their significance once fixed effects for areas smaller than the *départements* are included. This reinforces the intuition that the naive estimates presented in the previous section were actually upward biased. This is addressed in the following subsection.

an area delimited by topographical elements such as streets, roads, railways and rivers, containing between 120 and 240 homes and hence between 6 and 11 *aires*, out of which 6 are randomly selected to be included in the labor force survey sample to cover its total lifespan. The last column reports the estimates of diversity when *sector* fixed effects are included. The estimates are still significantly negative, but are not lower than with the municipalities fixed effects. All in all, these results confirm that local diversity has indeed a strong negative causal impact on the probability to be employed, and that, if anything, this effect was underestimated in the previous section.

## 5.2 Employment zone diversity: an instrumental variable approach

The second part of this section deals with the endogeneity of employment zones diversity. Because employment zones are quite large areas (there are about 300 employment zone in mainland France ), the assumption made in the previous subsection cannot hold and the above strategy cannot be applied. Instead, I rely on a more standard instrumental variable estimation. A plausible instrument should be correlated with employment zone level of diversity (e.g. employment zone population composition), but uncorrelated with labor market outcomes. In what follows, I propose two different instruments.

The first instrument relies on the "shift-share" methodology initiated by [Card \(2001\)](#) and more recently used by [Saiz \(2007\)](#) and [Ottaviano and Peri \(2006\)](#) in a setting close to this paper's. It builds on the idea that new migrants to a country tend to settle where former migrants from the same origin previously settled, *i.e.* ethnic enclaves ([Munshi, 2003](#); [Winters et al., 2001](#)). Using past settlements of immigrants from various countries across French employment zones, it is possible to construct a predicted measure of current diversity in each employment zone. More precisely, I use the 1968 population census data to compute the distribution of each origin group across French employment zones. Because employment zones did not exist in 1968, and because their frontiers evolve over time, I apply the 2010 employment zones borders to the 1968 population. The origin groups considered are limited by the information contained in the 1968 census. In particular, no information about parents is available. I can still rely on nationalities and birth countries, grouped into the six categories defined previously. Then, for each origin group, I apply the corresponding 1968 distribution across employment zones to the current (2007 to 2010) total population in France. Doing so, I compute the expected number of individuals from each origin in each employment zone, solely based on the ethnic enclaves pull factor. From this, I can deduce the predicted composition of each employment zone population. Once the predicted shares of each group are computed, I can eventually construct the predicted level of diversity in each employment zone over the 2007 to 2010 period. By

construction, the predicted measure of diversity depends only on 1968 population settlements and not on any employment zone-specific shock (e.g. productivity shock), and can thus be used to instrument actual diversity. A more formal description how the predicted level of diversity is computed can be found in Appendix 3.A.

The second instrument is more innovative and builds on [Algan et al. \(2013\)](#). In this paper, the authors provide evidence that the allocation of households across public housing units in France does not take their origins or their preference for diversity into account, so that diversity can be considered as exogenous within the public housing sector. Not only do legal rules prohibit housing allocation based on ethnic backgrounds, but the characteristics of the public housing sector, which is very tight and highly regulated, also make it very complicated to bypass the law in practice. In addition to presenting these general arguments, the authors conduct a variety of formal statistical tests to verify the absence of self-sorting on ethnic characteristics. In particular, they show that the observed spatial distribution of residents across public housing blocks is not statistically different from a random distribution. Now that the exogeneity of diversity is acknowledged, I argue that the level of diversity within the public housing sector of a given area is necessarily correlated to the global level of diversity of this area. Indeed, because the public housing population is part of the total population, fractionalization based only on this sub-population is mechanically proportional to fractionalization based on the population as a whole. Also, it is reasonable to think that individuals living in the private housing market, and who are therefore less constrained upon their location choice, are influenced by the composition of the public housing population. Otherwise stated, people from a given group might be attracted to an area where some of their co-ethnics were located through the public housing allocation process, so that diversity in the area is likely to reflect diversity in the area's public housing sector. In a nutshell, public housing diversity can be considered as exogenous, and it is correlated to total diversity both mechanically and through a magnet effect on immigrants living in the private housing sector. It can therefore reasonably be used to instrument the total level of diversity.

I now estimate the effect of diversity on employment status using a two-stage least square procedure, where the two instrumental variables described above are alternatively employed. Aside from the inclusion of an instrument, the specification corresponds to the full specification presented in Section 4, where I control for individual characteristics, local neighborhood unemployment, and *département* fixed effects. The results are summarized in Table 3.8, where I only report the coefficients and statistics of interest.<sup>13</sup> I report the OLS estimates in the first column to ease their comparison with IV estimates.

---

<sup>13</sup>The coefficients of the other variables are almost unchanged, and are available upon request.

Column 2 displays the results obtained using the predicted measure of employment zone diversity as an instrument for employment zone diversity. More precisely, in the first part of the table, which deals with diversity based on nationalities, the predicted diversity is also based on nationalities. In the second part of the table which focuses on birth country-based diversity, I use the predicted diversity based on birth countries instead.<sup>14</sup> In both cases, we observe that the coefficient for employment zone diversity loses its significance once it is instrumented by the predicted level of diversity. The magnitude of the coefficient drops significantly and is driven down to zero (especially in the birth country regression), so that the lack of significance is not just a consequence of larger standard errors. The first-stage statistics reported at the bottom of Table 3.8 illustrate the strength of the excluded instrument. The F-statistics testing the hypothesis that the excluded instrument is equal to zero in the first stage are much larger than the rule-of-thumb value of 10 indicated in the literature on weak instruments (Staiger and Stock, 1997, *e.g.*). In addition, the partial  $R^2$  confirm the large correlation between the excluded instruments and the endogenous variables. The results obtained using the second instrument, namely diversity within the public housing sector of the employment zones, are reported in column 3. The first-stage statistics also reveal that this instrument is strong. Using it leads to the same results as those obtained with the first instrument: the coefficients for employment zone diversity are basically annihilated in the second stage.<sup>15</sup> To summarize, these results show that employment zone diversity does not have any causal impact on individuals' employment status. This confirms the suspicion that the naive estimates derived in the previous section were upwardly biased, probably due to a sorting of immigrants into more economically dynamic areas.

## 6 Interpretation of the results

So far, I have shown that there is a positive relationship between diversity and individuals' employment probability at the employment zone level, but that it is merely due to self-selection, and does not correspond to any causal relationship from the former to the latter. By contrast, I have also established that living in a diverse neighborhood actually implies a lower employment probability. This section is an attempt to understand why local di-

---

<sup>14</sup>Given that I do not have any information about parents' origins in the 1968 census, I am unable to compute the predicted level of diversity based on this particular measure of origins. As a consequence, I have alternatively instrumented parents' origins-based diversity by the predicted level of diversity based on nationalities and on birth countries. The results are comparable to those reported in the table for the other measures of diversity, and are available upon request.

<sup>15</sup>A comment on local neighborhood diversity is in order here. As noted in the previous subsection, the estimates reported in Table 3.8 are likely biased, as we only control for *département* fixed effects. However, we already know that correcting for this bias by introducing fixed effects for smaller area (*e.g.* municipality) reduces these coefficients further.

iversity reduces individuals' employment prospects. As mentioned at the beginning of the paper, one of the channels that comes to mind when thinking about the relationship between neighborhood diversity and employment is the channel of communication between agents. Specifically, if neighbors from different origins do not communicate, e.g. because they do not speak the same language, then information about job opportunities or about how to register to an employment agency does not circulate across groups. One of the results obtained in this paper, namely that the negative effect of neighborhood diversity is stronger for nationality-based diversity than for birth country- or parents' origin-based diversity is a first evidence supporting this intuition. Indeed, even if the individuals are from different origins, the fact that they share the same nationality imply that they have lived long enough in the corresponding contry to acquire the nationality.

In order to challenge assumptions more formally, I look at the correlation between local diversity and the nature of neighborhood relationships using the 2002 French Housing Survey. Surveyed individuals are asked to qualify the relationships with their neighbors, which can either be *good*, *average*, *bad*, or *nonexistent*. In addition, I know the precise (block level) place where the individuals live, and I am able to match it with representative block level measures of diversity computed using the 1999 population census. The results of multinomial logit regressions of the quality of neighborhood relationships on neighborhood diversity are presented in Table 3.9. Each line corresponds to a separate regression: the first line displays the estimated coefficients of diversity based on nationality at birth, those for birth country-based diversity being reported in the second line. Each regression controls for individual characteristics (age, gender, origin, employment status, education, household income), block level unemployment rate, department fixed effects and a detailed indicator of the social and economic composition of the neighborhood in 27 categories.<sup>16</sup> The results reveal that individuals living in more diverse neighborhoods are less likely to report having good relationships with their neighbors. In particular, they are more likely to report having bad relationships than average relationships than no relationship at all. These simple results tend to support the idea that communication can be hindered in more diverse neighborhoods due to the poor quality of the relationships between neighbors.

An alternative test of this intuition is to see how employment status is affected by the presence of people from the same origin group. Presumably, if the negative effect of neighborhood diversity is due to limited information transmission across groups, then living close to people from the same origin should conversely be related to better employ-

---

<sup>16</sup>The socio-economic classification of French neighborhoods into 27 groups is realized by [Martin-Houssart and Tabard \(2002\)](#).



ment prospects. Using the LFS data, I compute, for each individual, the share of the local neighborhood population belonging to the same origin group (excluding the reference individual from the computation). I then run simple OLS regressions of the employment status dummy used in Sections 4 and 5 on this variable, using the same set of controls as in the full specification. However, because I want to avoid the bias due to endogenous location selection, I include municipalities rather than *département* fixed effects. The results are presented in Table 3.10. The estimates displayed in column 1 show that the larger the share of neighbors from one's own origin group, the higher one's employment probability. This is especially true when the individual's nationality determines her or his origin group. To put it differently, when communication is free from cultural or language barriers with a larger share of individuals, employment prospects are improved. Mechanically, more diversity implies smaller group shares, contributing to the negative effect of diversity. As a matter of fact, once we control for neighborhood diversity in column 2, the estimates of the share of people from the same group are strongly reduced (and lose their significance except for nationality).

A more direct and natural way to dig into the hypothesis that job information transmission bridges the gap between diversity and employment, is to focus on the role of personal networks for job seekers. The LFS data provide information about the methods used by individuals who are looking for a job. Job seekers, whether unemployed or not can indicate which methods they use among 15 possible methods. For the purpose of the present study, I focus on the use of friends and family network as a job search method. I construct two variables: a dummy indicating whether the job seeker relies on personal networks or not, possibly combined with other job search methods, and a dummy equal to one if the person exclusively uses her or his network. Simple OLS regressions including the full set of controls used throughout the paper (individual characteristics including employment status, neighborhood unemployment rate, quarter and municipality fixed effects), reveal that neighborhood diversity does not relate to these variables, as shown in Table 3.11. Yet, individuals' origin matter to some extent in explaining the use of networks to search for a job. People with Mediterranean, Maghrebian and other African origins (taking 2nd generation into account) are more likely to rely on personal networks than natives (column 3). Interestingly, European (other than South European) and African citizens are also more likely to rely exclusively on networks (column 4), revealing a low level of integration for those particular groups. For the minority groups that heavily rely on networks to look for jobs, living in more diverse areas and hence being cut from the bulk of their friends and family might therefore hinder their job search efforts.

The LFS also asks employed workers to indicate the main channel through which

they found their current job. I build a variable equal to 1 if the individual found her or his job through personal contacts, and 0 otherwise, which I regress on diversity using the same specification as for job search methods (obviously excluding employment status). Table 3.12 reports the estimates of neighborhood diversity, which do not significantly differ from zero, suggesting that living in a more diverse environment do not influence the chance to find a job through contacts. However, employed foreigners of any citizenship are more likely to have found their job using networks than French citizens are (column 1). This is especially true for individuals of South European and rest of the world nationalities. The coefficients decrease (or even vanish) as other measures of origins are considered, suggesting that networks are particularly helpful for the least integrated people, i.e. those who are of foreign origin but who have not yet been naturalized. Although these results are to be interpreted with caution because they do not correct for selection and do not control for the search methods that were actually used, they suggest that friends and family network is an important vector of employment for foreign individuals. Therefore, even if diversity is not directly involved in the use of networks to search and find jobs, it might still be an issue for minorities if they live in diverse areas, isolated from the core of their network.

## 7 Conclusion

By exploring employment prospects and diversity at very low geographic levels, the findings of this paper bring new insights to the literature on diversity. First, measuring diversity at different geographic levels reveals that this effect is not independent from the level of observation. Neighborhood diversity reduces employment prospects, while employment zone diversity is neutral, after correcting for endogenous sorting. This implies that the mechanisms through which diversity hinders employment at a local level are counterbalanced at a more aggregate level. In particular, job seekers might be unable to develop efficient networks in their own neighborhood because of diversity, but they might instead rely on a network established in a larger area. Second, measuring diversity based on various definitions of origins reveals that diversity in terms of nationality matters more than diversity in terms of parents origins. This is a key result, as it means that diversity of origins plays a role through the variety of cultures and languages rather than through ethnic diversity *per se*. This speaks in favor of the idea that diversity affects employment prospects by altering job information transmission. More generally, this work calls for a new approach to the literature on diversity, as it shows that (i) the notion of diversity hides various aspects that can influence the considered outcome in different ways, and that (ii) the effect of diversity can vary according to the considered geographical level.

Although part of this paper is devoted to test the hypothesis that the negative impact of

local diversity on employment prospects is related to job information transmission, much remains to be done in this direction. In addition, a natural subsequent question is that of the quality of the job found in terms of tenure or wage for instance. These issues remain open for future research.

## Tables

**Table 3.1:** Sample Description: 16-65 y.o. individuals, 2007-2010

	[Min-Max]	Mean	(Std Dev)
Male	[0-1]	49.30	(50.00)
Age	[16-65]	42.92	(12.71)
Experience (years)	[0-63]	24.00	(13.82)
<b>Employment Status</b>			
Employed	[0-1]	0.714	(0.452)
Unemployed	[0-1]	0.066	(0.249)
Inactive	[0-1]	0.220	(0.414)
<b>Socio-Economic Category</b>			
Farmer	[0-1]	0.018	(0.133)
Craftsman, shopkeeper	[0-1]	0.059	(0.236)
Executive or other high position	[0-1]	0.143	(0.350)
Intermediate occupation	[0-1]	0.225	(0.417)
Employee	[0-1]	0.304	(0.460)
(Factory) worker	[0-1]	0.240	(0.427)
Unemployed never employed	[0-1]	0.011	(0.106)
<b>Level of Education</b>			
Master, PhD, schools	[0-1]	0.077	(0.267)
Graduate (bac+4)	[0-1]	0.030	(0.170)
Under-graduate (bac+3)	[0-1]	0.037	(0.188)
Lower under-grad (bac+2)	[0-1]	0.125	(0.331)
General Baccalaureat	[0-1]	0.081	(0.272)
Techno. / Pro. Baccalaureat	[0-1]	0.075	(0.263)
bretech	[0-1]	0.020	(0.140)
cap	[0-1]	0.253	(0.435)
bepc	[0-1]	0.082	(0.274)
No Diploma	[0-1]	0.222	(0.415)
<b>Employed workers characteristics</b>			
Hourly wage (log)	[-5.02-6.828]	2.276	(0.445)
Tenure (months)	[0-792]	136.0	(125.6)
Public servant	[0-1]	0.280	(0.449)
Part time job	[0-1]	0.165	(0.371)
Permanent contract	[0-1]	0.835	(0.371)

These figures are obtained using a sample of 920,388 individuals aged between 16 and 65 years old. It consists in the observations from the 16 successive waves of the labor Force Survey from 2007 to 2010.

**Table 3.2:** Distribution of individuals' origins, 2007-2010 (in %)

	Nationality	Birth Country	Parents
France	93.65	87.09	80.31
Southern Europe	1.51	2.18	6.30
Rest of Europe	1.07	1.81	3.31
Maghreb	1.86	4.86	5.66
Rest of Africa	0.81	1.79	1.80
Rest of the World	1.10	2.27	2.62
N	920,235	920,346	905,241

Reading: among the 15-65 y.o. individuals living in France, 1.86 % are of Maghrebian nationality, 4.86 % are born in Maghreb and 5.66 % have a Maghrebian origin, either by their nationality or through their parents'.

**Table 3.3:** Diversity in individuals' living environment

	[Min-Max]	Mean	(Std Dev)	Median
<b>Neighborhood diversity</b>				
Nationality	[0-0.771]	0.089	(0.131)	0.034
Birth Country	[0-0.803]	0.175	(0.162)	0.132
Parents	[0-0.818]	0.280	(0.207)	0.246
<b>Employment Zone diversity</b>				
Nationality	[0-0.559]	0.099	(0.080)	0.078
Birth Country	[0-0.731]	0.190	(0.117)	0.166
Parents	[0-0.735]	0.325	(0.171)	0.317

Reading: Individuals live in neighborhoods where diversity in terms of nationality amounts to 8.9 % on average. They live in employment zones where diversity in terms of birth country amounts to 19 % on average. Alternatively: there is a 32.5 % chance that two individuals living in the same employment zone are from different origin background.

**Table 3.4:** Employment Status and Diversity by Nationality

	No controls	Individual Characteristics	Ind. charac., Time & Geo. FE,	Ind. charac., Time & Geo. FE, Local Unemployment
	(1)	(2)	(3)	(4)
<b>Diversity by Nationality</b>				
Local Neighborhood	-0.364*** (0.016)	-0.172*** (0.013)	-0.181*** (0.013)	-0.105*** (0.011)
Employment Zone	0.447*** (0.024)	0.189*** (0.018)	0.122*** (0.028)	0.090*** (0.025)
<b>Nationality (Ref.: French)</b>				
South European		0.103*** (0.009)	0.098*** (0.009)	0.087*** (0.009)
Other European		-0.125*** (0.013)	-0.124*** (0.013)	-0.132*** (0.013)
Maghrebian		-0.143*** (0.011)	-0.139*** (0.011)	-0.130*** (0.011)
Other African		-0.122*** (0.017)	-0.126*** (0.017)	-0.125*** (0.017)
Other nationality		-0.111*** (0.014)	-0.115*** (0.014)	-0.117*** (0.014)
<b>Education (Ref: Baccalauréat)</b>				
Master, PhD & schools		0.049*** (0.006)	0.048*** (0.006)	0.047*** (0.006)
Graduate (bac+4)		0.034*** (0.007)	0.033*** (0.007)	0.034*** (0.007)
Under-graduate (bac+3)		0.044*** (0.006)	0.045*** (0.006)	0.045*** (0.006)
Lower under-grad (bac+2)		0.039*** (0.005)	0.039*** (0.005)	0.038*** (0.005)
Techno. & Pro. Baccalauréat		0.037*** (0.005)	0.038*** (0.005)	0.037*** (0.005)
Technical degree		0.013* (0.008)	0.012 (0.008)	0.012 (0.008)
cap		-0.004 (0.005)	-0.004 (0.005)	-0.003 (0.005)
bepc		-0.039*** (0.006)	-0.037*** (0.006)	-0.035*** (0.006)
No diploma		-0.082*** (0.005)	-0.080*** (0.005)	-0.073*** (0.005)

**Table 3.4:** Employment Status and Diversity by Nationality (C'ed)

	No controls	Individual Characteristics	Ind. charac., Time & Geo. FE,	Ind. charac., Time & Geo. FE, Local Unemployment
	(1)	(2)	(3)	(4)
<b>Socio-economic category (Ref: )</b>				
Craftsman, shopkeeper		0.204*** (0.010)	0.206*** (0.010)	0.205*** (0.010)
Executive or other high position		0.191*** (0.010)	0.191*** (0.010)	0.190*** (0.010)
Intermediate occupation		0.162*** (0.010)	0.163*** (0.010)	0.163*** (0.010)
Employee		0.150*** (0.010)	0.152*** (0.010)	0.154*** (0.009)
(Factory) worker		0.098*** (0.010)	0.101*** (0.009)	0.105*** (0.009)
<b>Other individual characteristics</b>				
Male		0.087*** (0.002)	0.087*** (0.002)	0.087*** (0.002)
Age		0.068*** (0.002)	0.068*** (0.002)	0.067*** (0.001)
Age <sup>2</sup>		-0.001*** (0.000)	-0.001*** (0.000)	-0.001*** (0.000)
Experience		0.007*** (0.001)	0.007*** (0.001)	0.007*** (0.001)
Experience <sup>2</sup>		-0.000*** (0.000)	-0.000*** (0.000)	-0.000*** (0.000)
<b>Local unemployment rate</b>				
				-0.272*** (0.014)
Intercept	0.682*** (0.003)	-0.651*** (0.027)	-0.608*** (0.031)	-0.570*** (0.030)
Quarter F.E.	No	No	Yes	Yes
Department F.E.	No	No	Yes	Yes
N	162,097	151,053	151,053	150,913
Adj. R <sup>2</sup>	0.009	0.254	0.257	0.260

The dependent variable indicates the employment status of an individual in a given quarter. It takes value 1 if the individual is employed, and 0 otherwise (unemployed or inactive). It is regressed on diversity by nationality based on the 6-categories nationality variable. Each column corresponds to a different specification. In column (1), the employment dummy is regressed on neighborhood and employment zone diversity, without any other control. Column (2) controls for individual characteristics: origin group (6 categories), gender, quadratic function of age, education (10 categories), socio-economic category (6 categories), quadratic function of experience. Column (3) = (2) + quarter fixed effects + *département* fixed effects. Column (4) = (3) + unemployment rate in the local neighborhood (excluding the individual). The sample is made of the first observation of each individual. Standard errors clustered at the neighborhood level are reported in parentheses. \* p<0.10, \*\* p<0.05, \*\*\* p<0.001

**Table 3.5:** Employment status and diversity: summary of the results (OLS)

	No controls (1)	Individual Characteristics (2)	Ind. charac., Time & Geo. FE, (3)	Ind. charac., Time & Geo. FE, Local Unemployment (4)
<b>1. Diversity by Nationality</b>				
Local Neighborhood	-0.364*** (0.016)	-0.172*** (0.013)	-0.181*** (0.013)	-0.105*** (0.011)
Employment Zone	0.447*** (0.024)	0.189*** (0.018)	0.122*** (0.028)	0.090*** (0.025)
<b>2. Diversity by Birth Country</b>				
Local Neighborhood	-0.346*** (0.014)	-0.149*** (0.011)	-0.156*** (0.011)	-0.092*** (0.010)
Employment Zone	0.381*** (0.018)	0.151*** (0.014)	0.120*** (0.023)	0.088*** (0.020)
<b>3. Diversity by Parents Origins</b>				
Local Neighborhood	-0.234*** (0.011)	-0.089*** (0.009)	-0.096*** (0.009)	-0.053*** (0.008)
Employment Zone	0.215*** (0.013)	0.091*** (0.010)	0.056*** (0.016)	0.041** (0.014)
Individual controls	No	Yes	Yes	Yes
Quarter F.E.	No	No	Yes	Yes
Département F.E.	No	No	Yes	Yes
Local unemployment rate	No	No	No	Yes

The dependent variable indicates the employment status of an individual in a given quarter. It takes value 1 if the individual is employed, and 0 otherwise (unemployed or inactive). It is regressed on diversity by nationality in the first set of regressions (1.), on diversity by birth country and by parents' origins in the second (2.) and third (3.) sets of regressions respectively. Fractionalization indices are based on the 6-categories origin variables. Each column corresponds to a different specification. In column (1), the employment dummy is regressed on neighborhood and employment zone diversity, without any other control. Column (2) controls for individual characteristics: origin group (6 categories), gender, quadratic function of age, education (10 categories), socio-economic category (6 categories), quadratic function of experience. Column (3) = (2) + quarter fixed effects + *département* fixed effects. Column (4) = (3) + unemployment rate in the local neighborhood (excluding the individual). The sample is made of the first observation of each individual. Standard errors clustered at the neighborhood level are reported in parentheses. \* p<0.10, \*\* p<0.05, \*\*\* p<0.001

**Table 3.6:** Employment status and diversity: summary of the results (Multinomial logit)

Dep Var:	Employment Status (Ref: Employed)	
	Unemployed	Inactive
<b>1. Diversity by Nationality</b>		
Local Neighborhood	0.921*** (0.090)	0.463*** (0.092)
Employment Zone	-0.566** (0.228)	-0.628** (0.207)
<b>1. Diversity by Birth Country</b>		
Local Neighborhood	0.929*** (0.083)	0.351*** (0.078)
Employment Zone	-0.580** (0.187)	-0.603*** (0.167)
<b>1. Diversity by Parents' Origins</b>		
Local Neighborhood	0.687*** (0.069)	0.201** (0.062)
Employment Zone	-0.387** (0.129)	-0.273** (0.114)
<hr/>		
Individual controls		Yes
Local unemployment rate		Yes
Quarter <i>dep.</i> F.E.		Yes

The dependent variable indicates the employment status of an individual in a given quarter. It takes value 1 if the individual is employed (reference category), 2 if s/he is unemployed and 3 if s/he is inactive. It is regressed on diversity by nationality based on the 6-categories origin variables. It is regressed on diversity by nationality in the first regression (1.), on diversity by birth country and by parents' origins in the second (2.) and third (3.) regressions respectively. The results come from a multinomial logit estimation, using the full specification. In each regression, the following controls are included: individual characteristics (origin group, gender, quadratic function of age, education, socio-economic category, quadratic function of experience), local unemployment rate, and quarter and *département* fixed effects. The sample is made of the first observation of each individual. Standard errors clustered at the neighborhood level are reported in parentheses. \* p<0.10, \*\* p<0.05, \*\*\* p<0.001



**Table 3.7:** Employment status and local diversity: considering within area variation

	Département FE (1)	Employment Zone FE (2)	Municipality FE (3)	Sector FE (4)
<b>Local Neighborhood Diversity</b>				
1. By Nationality	-0.094*** (0.010)	-0.109*** (0.010)	-0.140*** (0.013)	-0.110*** (0.017)
2. By Birth Country	-0.079*** (0.009)	-0.094*** (0.009)	-0.120*** (0.011)	-0.080*** (0.016)
3. By Parents Origins	-0.047*** (0.007)	-0.054*** (0.007)	-0.075*** (0.009)	-0.060*** (0.012)

The dependent variable indicates the employment status of an individual in a given quarter. It takes value 1 if the individual is employed, and 0 otherwise (unemployed or inactive). The sample is made of the first observation of each individual. Fractionalization indices are based on the 6-categories origin variables. Each regression controls for the full set of individual characteristics, quarter and department fixed effects and local neighborhood unemployment rate. However, compared to the previous specification, employment zone diversity is not included so as to focus on the changes of local neighborhood diversity to the inclusion of the alternative fixed effects. Standard errors clustered at the neighborhood level are reported in parentheses.

\* p<0.10, \*\* p<0.05, \*\*\* p<0.001

**Table 3.8:** Effect of diversity on employment status: IV regressions

Instrument used:	OLS	IV: Expected Diversity	IV: Public Housing Diversity
	(1)	(2)	(3)
<b>1. Diversity by Nationality</b>			
Local Neighborhood	-0.105*** (0.011)	-0.092*** (0.014)	-0.104*** (0.012)
Employment Zone	0.090*** (0.025)	-0.028 (0.070)	0.025 (0.054)
<b>First stage</b>			
Expected Diversity		0.468*** (0.020)	
Public Housing Diversity			0.197*** (0.006)
F-stat (excl. instr.)		537.20	937.02
Partial $R^2$ (excl. instr.)		0.133	0.252
<b>2. Diversity by Birth Country</b>			
Local Neighborhood	-0.092*** (0.010)	-0.081*** (0.011)	-0.092*** (0.011)
Employment Zone	0.088*** (0.020)	0.006 (0.044)	0.046 (0.046)
<b>First stage</b>			
Expected Diversity		0.723*** (0.024)	
Public Housing Diversity			0.209*** (0.007)
F-stat (excl. instr.)		922.28	848.12
Partial $R^2$ (excl. instr.)		0.227	0.230

The dependent variable indicates the employment status of an individual in a given quarter. It takes value 1 if the individual is employed, and 0 otherwise (unemployed or inactive). The sample is made of the first observation of each individual. Fractionalization indices are based on the 6-categories origin variables. Each regression controls for the full set of individual characteristics, quarter and department fixed effects and local neighborhood unemployment rate. Standard errors clustered at the neighborhood level are reported in parentheses. \*  $p < 0.10$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.001$

**Table 3.9:** Quality of the relationships with neighbors and diversity

Dep. Var.:	Quality of Neighborhood Relationships (Ref: Good)		
	Average	Bad	No relationship
1. Diversity by nationality	1.434*** (0.235)	1.708*** (0.506)	0.882*** (0.179)
2. Diversity by birth country	1.617*** (0.273)	1.958** (0.610)	1.206*** (0.207)

Each line reports the coefficients from a separate multinomial logit regression. The dependent variable indicates opinion about the relationships with the neighbors. It takes value 1 if the surveyed individual declares having good relationships with his/her neighbors (reference category), 2 if the relationships are average, 3 if they are bad, and 4 if there is no relationship at all. The main variable of interest is the level of diversity, computed at the block level, based on nationalities in the first regression and on birth countries in the second one. In each specification, the following controls are included: individual characteristics (age, gender, origin, employment status, education, household income), block level unemployment rate, department fixed effects and a detailed indicator of the social and economic composition of the neighborhood (27 categories). The data come from the 2002 French Housing Survey and the 1999 population census (INSEE). Standard errors clustered at the neighborhood level are reported in parentheses. \* p<0.10, \*\* p<0.05, \*\*\* p<0.001

**Table 3.10:** Share of neighbors from own origin group and employment probability

Dep. Var.: Employment status (employed vs unemployed or inactive)		
	(1)	(2)
<b>1. Origins: Nationality</b>		
Neighborhood share of same origin	0.154*** (0.016)	0.046** (0.023)
Neighborhood Diversity		-0.115*** (0.018)
<b>2. Origins: Birth country</b>		
Neighborhood share of same origin	0.073*** (0.013)	-0.014 (0.163)
Neighborhood Diversity		-0.127*** (0.014)
<b>3. Origins: Parents</b>		
Neighborhood share of same origin	0.037*** (0.008)	0.002 (0.010)
Neighborhood Diversity		-0.074*** (0.011)

Each column and each set of results (1., 2. and 3.) report the coefficients from a separate OLS regression. The dependant variable indicates whether the individual is employed (1) or unemployed or inactive (0). In each specification, the following controls are added to the variables displayed: individual characteristics (origin group, gender, quadratic function of age, education, socio-economic category, quadratic function of experience), neighborhood unemployment rate (excluding the individual), quarter dummies and municipalities fixed effects. The sample is made of the first observation of each individual. Standard errors clustered at the neighborhood level are reported in parentheses. \* p<0.10, \*\* p<0.05, \*\*\* p<0.001

Table 3.11: Job search method: use of friends and family networks

Dep Var:	Use of networks			Exclusive use of networks		
	Nationality (1)	Birth country (2)	Parents (3)	Nationality (4)	Birth country (5)	Parents (6)
<b>Neighborhood diversity</b>	-0.027 (0.025)	-0.035 (0.023)	-0.005 (0.019)	-0.007 (0.007)	-0.006 (0.006)	-0.004 (0.005)
<b>Origin Group: (Ref: France)</b>						
South Europe	0.024 (0.024)	0.027 (0.021)	0.033** (0.011)	0.008 (0.007)	-0.003 (0.006)	-0.003 (0.003)
Rest of Europe	-0.006 (0.021)	0.018 (0.017)	0.017 (0.014)	0.026*** (0.006)	0.022*** (0.005)	0.012** (0.004)
Maghreb	0.018 (0.014)	0.015 (0.010)	0.028** (0.009)	0.005 (0.004)	0.000 (0.003)	0.001 (0.002)
Rest of Africa	0.022 (0.019)	0.027* (0.014)	0.030** (0.014)	0.012** (0.005)	0.003 (0.004)	0.007* (0.004)
Rest of World	0.049** (0.019)	0.039** (0.015)	0.033** (0.013)	0.018*** (0.005)	0.009** (0.004)	0.008** (0.004)

In the first three columns, the dependent variable indicates whether the individual relied on personal networks to search for a job, possibly combined with other job search methods. In the last three columns, the dependent variable indicates whether the individual relied on personal networks to search for a job, excluding the use of any other job search method. Fractionalization indices are based on the 6-categories origin variables. The estimates come from OLS regressions. In addition to diversity and origin group (which differ in each column), each regression controls for employment status, gender, age, age squared, education, SEC, experience, experience squared, neighborhood unemployment rate, quarter dummies and municipality fixed effect are also included. The sample is made of the first observation of each individual. \* p<0.10, \*\* p<0.05, \*\*\* p<0.001

**Table 3.12:** Main method through which the job was found: networks

Dep Var:	Job found through networks		
	Nationality	Birth country	Parents
Origins:	(1)	(2)	(3)
<b>Neighborhood diversity</b>	0.012 (0.016)	-0.003 (0.015)	-0.012 (0.012)
<b>Origin Group:</b> (Ref: France)			
South Europe	0.141*** (0.011)	0.103*** (0.009)	0.057*** (0.006)
Rest of Europe	0.040** (0.015)	0.023** (0.011)	0.034*** (0.008)
Maghreb	0.031** (0.012)	0.006 (0.008)	-0.003 (0.007)
Rest of Africa	0.038** (0.016)	-0.000 (0.011)	-0.002 (0.011)
Rest of World	0.178*** (0.014)	0.114*** (0.010)	0.103*** (0.009)

The dependent variable indicates whether the currently employed individual found his/her job through personal network. The sample is made of the first observation of each employed individual, and excludes civil servants. Fractionalization indices are based on the 6-categories origin variables. The estimates come from OLS regressions. In addition to diversity and origin group (which differ in each column), each regression controls for gender, age, age squared, education, SEC, experience, experience squared, neighborhood unemployment rate, quarter dummies and municipality fixed effect are also included. \*  $p < 0.10$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.001$

### 3.A Construction of the predicted level of employment zone diversity

In this appendix, I describe more formally the construction of the "shift-share" instrumental variable. Denote  $N_{France,1968}^g$  the number of individuals from origin group  $g = 1, \dots, g_{max}$  in France in 1968 and  $N_{EZ_j,1968}^g$  the number of individuals from origin group  $g = 1, \dots, N_g$  in employment zone  $j = 1, \dots, N_j$  in 1968. Then, the share of group  $g$  individuals living in employment zone  $j$  in 1968 (out of the total number of group  $g$  individuals in France in 1968) can be computed as follows:

$$S_{EZ_j,1968}^g = \frac{N_{EZ_j,1968}^g}{N_{France,1968}^g} \quad (3.3)$$

with  $\sum_{j=1}^{N_j} S_{EZ_j,1968}^g = 1$ , for any group  $g$ .

Then, the expected number of group  $g$  individuals living in employment zone  $j$  in year  $t = 2007, \dots, 2010$  is given by:

$$\widehat{N_{EZ_j,t}^g} = S_{EZ_j,1968}^g * N_{France,t}^g \quad (3.4)$$

From this, we can deduce the expected share of group  $g$  individuals in employment zone  $j$  in year  $t$  (out of the total number of individuals living in employment zone  $j$  in  $t$ , all groups included):

$$\widehat{s_{EZ_j,t}^g} = \frac{\widehat{N_{EZ_j,t}^g}}{\sum_{g=1}^{N_g} \widehat{N_{EZ_j,t}^g}} \quad (3.5)$$

with  $\sum_{g=1}^{N_g} \widehat{s_{EZ_j,t}^g} = 1$ , for any employment zone  $j$ .

Finally, the predicted measure of diversity in employment zone  $j$  in  $t$  is obtained as follows:

$$\widehat{DIV_{EZ_j,t}} = 1 - \sum_{g=1}^{N_g} \widehat{s_{EZ_j,t}^g}^2 \quad (3.6)$$



# Chapter 4

## Ethnic Networks and the Informal Labor Market

### 1 Introduction

Studies on economic integration and labor market success of immigrants in western economies all reveal that ethnic minorities face adverse labor market conditions. Both [Chiswick \(1980\)](#) and [Borjas \(1994\)](#), working respectively on UK and US data, show that workers from ethnic minorities earn significantly lower wages than majority workers. More recently, [Pendakur and Pendakur \(2011\)](#) report that Canadian-born individuals belonging to visible minorities face a substantial earning gap, which has not eroded over the past twenty years. In a study comparing France, Germany and the UK, [Algan et al. \(2010\)](#) reveal that not only do first and second generation immigrants have lower wages than natives, but they also face higher unemployment rates.<sup>1</sup> In another UK study, [Manacorda et al. \(2012\)](#) even find that recent immigration waves to the UK negatively affect former immigrants' wages. The literature provides various explanations to these low performances, from discrimination ([Becker, 1971](#); [Aeberhardt et al., 2010](#)), spatial mismatch ([Selod and Zenou, 2006](#); [Zenou, 2013](#)), or cultural differences ([Senik and Verdier, 2011](#)), to ethnic identity ([Bisin et al., 2011](#)).

A natural question to ask is how minority workers, especially low-skilled ones, cope with these adverse labor market conditions. The sociological literature argues that one way to compensate for low wages and low labor demand is to enter the informal labor market ([Tienda and Raijman, 2000](#)). [Portes and Haller \(2005\)](#) also highlight the cushioning function of informal employment, that they describe as a common additional source of income when unemployment is high and social benefits meager. In economics as well,

---

<sup>1</sup>Also refer to [Verdugo \(2009\)](#) for a more detailed description of immigrants labor market situation in France.



Lemieux et al. (1994), using Quebec City data, document that the participation rate of students and unemployed persons in the underground sector amounts to 28 percent, and rises to 32.4 percent for welfare recipients. All those studies suggest that informal economy is a way of improving one's economic condition when facing adverse labor market conditions.

More specifically, I acknowledge that turning to undeclared activities in response to precarious economic situation is particularly relevant for ethnic minorities. Several studies, mostly in sociology, provide empirical support to an over representation of ethnic minorities in the informal economy: in an OECD report, Lubell (1991) gives a few examples illustrating the fact that immigrants are to a large extent part of the informal sector. More recently, Baganha (2000) argues that working in the informal market is the prevailing mode of economic integration for immigrants in Portugal, Light and Gold (2000) estimate that 10 percent of the average American ethnic group's workers are employed in the informal sector of ethnic ownership economy, and Tienda and Raijman (2000) find that 38 percent of Mexican immigrant households in Chicago work in the underground economy.

I now turn to more formal arguments supporting the idea that immigrants and ethnic minorities are likely over-represented in the underground economy. First, and as mentioned earlier, minorities face higher unemployment rates and lower wages than the majority group, and hence are more likely to use informal work as an economic safety net. Then, because informal jobs are characterized by the absence of any written contract, and more broadly because they are not submitted to any labor legislation, engaging in informal activities allows to bypass red tape.<sup>2</sup> Given that immigrants are more subject to red tape than natives (e.g. obtaining a work permit), they can be more inclined to turn to informality for this reason as well. In addition to these structural factors, opportunity factors can be brought forward:<sup>3</sup> first, there is evidence that immigrants and ethnic minorities are on average less educated than natives and majority group, and hence are more likely to perform undeclared jobs, that are usually characterized by low-skilled workers (Boeri and Garibaldi, 2002).<sup>4</sup> In addition, many studies stress the role of networks in the informal economy, showing that ethnic networks are particularly powerful in fostering underground economies.

Alternatively, the extensive participation of ethnic minorities in the informal economy

---

<sup>2</sup>Note that avoiding red tape is one of the main reasons for the existence of informal jobs, after tax evasion. See European Commission report on the Eurobarometer pilot survey on undeclared work (2007).

<sup>3</sup>This classification of factors leading to informality is due to Renooy (1990), who distinguishes structural (financial pressure, institutional constraints) and opportunity (skills, education, contacts, environment...) factors.

<sup>4</sup>See Enste and Schneider (2000) and Gërxhani (2004) for comprehensive surveys on the informal economy.

can be explained by what is known in sociology as the *double disadvantage theory*.<sup>5</sup> A standard theory in the literature on ethnic economies, is that economically disadvantaged groups are likely to turn to entrepreneurship and self-employment. This *single disadvantage theory* explains why entrepreneurship rates are so high among ethnic minorities. However, when disadvantaged individuals also lack essential resources to become self-employed (e.g. skills, language, social capital), they turn instead to the informal economy (Light, 1979; Light and Rosenstein, 1995; Light, 2005).<sup>6</sup>

From what precedes, one can reasonably think that immigrants and ethnic minorities are more inclined to engage into undeclared activities.<sup>7</sup> To put it another way, current estimates of the labor market situation of ethnic minorities may underestimate their true "activity" rate as well as their income. In this paper, I provide a theoretical background for the fact that ethnic minorities circumvent adverse labor market outcomes by undertaking undeclared jobs. More precisely, this paper develops a dynamic model in which agents belonging to two different ethnic groups randomly meet by pair at each period, and exchange information on two competing activities: formal and informal labor. Formal workers pay taxes, while informal workers evade them. Workers from each group are similar in terms of skills (unskilled workers are more relevant for the considered setting). In this model, networks play a central role, as information about formal and informal job opportunities can only be obtained through word of mouth communication. There is indeed extensive evidence, both in the economic and in the sociological literature that social networks are primordial when it comes to finding a job.<sup>8</sup> The population is characterized by homophily, *i.e.* the tendency of individuals to associate with others who are similar to themselves, so that information about job offers circulates imperfectly across groups.

The setting of this paper is close to that of Calvó-Armengol *et al.* (2007) who study the role of the network and in particular of the strength of interpersonal relationships on the decision to enter criminality versus employment. Yet, the two models differ in three important aspects. First, I consider two different groups instead of a single one in order to introduce some asymmetry in the population in terms of network' size. Second, because I want to focus on the role of information transmission, I assume that both formal and informal job offers circulate only through word of mouth communication. By contrast,

---

<sup>5</sup>Also known as the *resource-constraint variant* of the disadvantage theory

<sup>6</sup>This quote from Light and Rosenstein (1995) is self-explanatory: "A typical firm in the informal economy requires few resources. For example, fruit vendors at freeway entrances are self-employed, but their informal enterprise did not require extensive resources of money, skill, and knowledge. Anyone could do it. On the other hand, when people with resources confront [labor market] disadvantage, they mobilize those resources to produce a *bona fide* business firm."

<sup>7</sup>Note also that even individuals turning to formal self-employment represent a non negligible share of the informal sector: using data from Denmark, Kleven *et al.* (2011) show that the tax evasion rate is equal to 41.6% for total self-reported net income, while it is close to zero when income is reported by a third-party.

<sup>8</sup>This evidence is presented in section 2.1 on information transmission.

Calvó-Armengol et al. (2007) suppose that information is diffused through the network for only one of the two activities (crime). Third, I let aside the notion of strong versus weak ties, therefore abandoning the dyadic structure of the network that prevails in their model. This simplifies the model that is already complicated by the first two additional features. Besides, the idea that the relationships are not the same between every two pairs of agents survives through the homophily assumption.

I analyze the flows of individuals between the various occupations and characterize all the steady-states of this dynamic economy. For this purpose, I solve for the endogenous individual decisions to accept or reject formal and informal job offers. I then derive the conditions under which these steady-states arise in equilibrium. Interestingly, when the net formal wage is larger than the net informal wage, all equilibria involving participation in the informal economy arise for low enough values of the unemployment benefit. This reveals that when the unemployment benefit is too low, individuals are more likely to immediately accept an informal job offer when they receive one rather than to decline and wait for a formal job opportunity.

In particular, an equilibrium in which both groups are formally employed but one also engages in the informal sector arises, and I show that this group is the minority group. This equilibrium is in line with the intuition that the informal sector may be used as an economic safety net when formal employment prospects are low. Indeed, the minority group faces such a low rate of formal job offers arrival (due to the size of the network) that its members are better off immediately accepting an informal job offer than waiting longer for a formal one. On the contrary, unemployed members of the majority group, whose network is more developed and hence who are more likely to receive formal job offers, can afford declining unattractive informal offers.

Simple comparative statics show that when communication across groups is improved, *i.e.* when the degree of homophily of the network decreases, then the formal employment rate increases for both groups, while the informal employment rate is reduced. That is, even if better information transmission can benefit both formal and informal sectors, individuals react by turning down more informal offers. Similarly, I find that increasing the rate at which informal workers are caught by fiscal authorities has a positive impact on the formal employment rate for both groups and a negative effect on the informal employment rate. However, because such a policy does not directly increase the rate of entry toward formal employment, it comes at the expense of increased unemployment rates. On the contrary, reducing homophily has a global negative impact on unemployment.

The remainder of the paper is organized as follows. Section 2 presents the model. In particular, it describes the network structure, information transmission mechanisms

and individual incentives. Section 3 defines and characterizes the steady-states, while the conditions for their existence in equilibrium are provided in Section 4. Finally, additional results and some comparative statics are derived in Section 5. Section 6 concludes.

## 2 The model

The population of size  $N$  is composed of two groups, the majority ( $M$ ) and the minority ( $m$ ).  $N_m$  and  $N_M$  respectively denote the size of the minority group and of the majority group in the population, with  $N_m + N_M = N$ , and  $N_M > N_m$ . We also denote  $n_m$  and  $n_M$  the share of each group in the population, such that  $\frac{N_M}{N} + \frac{N_m}{N} = n_M + n_m = 1$ .

The agents are assumed to be homogenous in terms of skills. We can for instance think of low-skilled workers, who are more concerned by undeclared jobs (Boeri and Garibaldi, 2005). The European Commission report on the Eurobarometer pilot survey on undeclared work (2007) also concludes that the main three economic sectors into which respondents admitted to do undeclared work are household services (including housekeeping, gardening, baby-sitting, elderly-sitting), construction services and personnel services. Lemieux et al. (1994) as well report that in Quebec City, workers with a declared job in primary industries have the highest participation rate in the underground sector (17.2 percent), followed by workers with a job in the construction industry (9.4 percent). They also reveal that two thirds of the jobs in the underground economy are in construction and services. Moreover, in response to poor labor market conditions, low-skilled workers are less likely to become self-employed as they lack the necessary resources.

In this model, individuals can engage in three competing activities: declared work, unemployment and undeclared work.<sup>9</sup> There are two main differences between declared and undeclared jobs: undeclared (or informal) jobs are assumed to be more flexible than declared (or formal) jobs,<sup>10</sup> and formal wages are subject to taxation, while undeclared wages are not. However, tax evaders face a positive probability of being detected by fiscal authorities, in which case they have to pay a fine. We denote  $F_{j,t}$  the number of formal workers in group  $j \in \{m, M\}$  at time  $t$ , and  $f_{j,t} \in \frac{F_{j,t}}{N_{j,t}}$  group  $j$ 's formal employment rate at time  $t$ . Similarly,  $U_{j,t}$ , and  $I_{j,t}$ ,  $\forall j \in \{m, M\}$  respectively denote the number of unemployed and informal workers in group  $j \in \{m, M\}$ , while  $u_{j,t}$  and  $i_{j,t}$  stand for group  $j$ 's unemployment and informal employment rates at time  $t$ . They are linked by the following relationship:

$$f_{j,t} + u_{j,t} + i_{j,t} = 1 \quad (4.1)$$

<sup>9</sup>This model is inspired by Calvó-Armengol et al. (2007), in which agents chose between employment, unemployment and crime.

<sup>10</sup>Section 2.1 provides more details on that issue.

In this economy, agents belong to a network through which they hear about job opportunities, be they formal or informal. There are two reasons why I allow formal job offers to grow through word of mouth communication. First, there is a large and growing literature providing a theoretical basis as well as empirical evidence of the fact that social contacts are a very important channel for finding jobs. I will come back on this literature in section 2.1 on labor market flows, but I refer thereupon to [Ioannides and Datcher Loury \(2004\)](#) and [Topa \(2011\)](#) for extensive surveys of the literature. Second, this paper focuses on the role played by network structure on employment patterns in both regular and underground labor markets. Hence, I do not want to ignore the fact that formal job offers circulate through the network.

## 2.1 Social interactions and job information transmission

**Social interactions.** The pattern of social interactions is close to [Calvó-Armengol et al. \(2007\)](#) or [Montgomery \(1994\)](#), in that individuals randomly meet by pair repeatedly over time. However, while these models are based on the idea that relationships between agents are characterized by their strength ([Granovetter, 1973](#)), I let this notion aside, and abandon the dyadic structure of the network that prevails in their models. That is, agents are not infinitely related to a given other agent who is able to influence their decision. Note however the idea that relationships are not necessarily the same between every two pairs of agents is still there, by assuming ethnic homophily (this will be established below). Therefore, any agent can be matched to an individual from his own ethnic group or from the other group, according to the size of each group in the population and with a bias toward his own group.<sup>11</sup> Similarly, any agent can meet a formal, an informal or an unemployed worker, depending on the various employment rates. Once a pair of agents is randomly formed, the individuals can exchange information about job opportunities. This process is explained below.

**The importance of networks in job information transmission.** The literature offers a plethora of empirical studies supporting the idea that social networks play a key role in finding jobs. As emphasized in [Ioannides and Datcher Loury \(2004\)](#) survey, the first stylized fact emerging from the literature is precisely that workers extensively use social contacts when searching for a job: [Holzer \(1988\)](#) finds that more than 85% of workers use their personal network for job search, while [Corcoran et al. \(1980\)](#) and [Granovetter \(1995\)](#) both report that more than 50% of jobs are found through social contacts. More recently,

---

<sup>11</sup>A possible way to reconcile this simple setting with real world is to think of this population as inhabitants of an ethnically diverse neighborhood: individuals can meet at the bus stop or at the bakery, they identify themselves as neighbors but do not necessarily talk about job opportunities, especially across groups. Another way to consider this set up, in the framework of previous papers, is to say that it focuses on the role of weak ties, *i.e.* random encounters (by opposition to friends and relatives).

Bayer et al. (2008) estimate that living in the same block increases by more than 33% the probability to work at the same location. Most importantly perhaps in the context of this paper is the fact that finding job through referrals is particularly prominent for low-skilled workers (see Corcoran et al., 1980; Granovetter, 1995; Topa, 2001; Wahba and Zenou, 2005; Battu et al., 2011).<sup>12</sup> In light of this empirical evidence, I assume without loss of generality that formal jobs opportunities circulate only through the network, and hence I neglect the fact that declared jobs can be obtained through direct formal methods (e.g. employment agencies).<sup>13</sup> Given the illegal nature of undeclared activities, underground job offers cannot be posted by employment agencies or be too widely spread (on the internet for instance) because they incur a risk of sanction by the authorities. Therefore, I also assume that informal jobs opportunities only circulate through the network.

**Transmission of information and labor market flows.** I now turn to a more precise description of the job information transmission mechanisms. As social network is the only channel through which job information circulates, an unemployed worker can hear about a formal (respectively informal) job offer only if s/he is matched with a formal (resp. informal) employed worker who is aware of an offer.<sup>14</sup> Formal workers are aware of a formal job opportunity at an exogenous rate  $\alpha_F$ , and will transmit this information if they meet unemployed workers.<sup>15</sup> Similarly, underground job offers circulate through undeclared workers who hear about an opportunity at an exogenous rate  $\alpha_I$ , and who transmit it to unemployed workers if they are randomly matched to one. I assume that the informal sector is more flexible than the formal one, so that  $\alpha_I > \alpha_F$ . As already mentioned, I assume that this economy features homophily, that is, individuals tend to form links and gather with people similar to themselves, and in particular along ethnic lines (see McPherson et al. (2001) for a review of the literature on homophily, and Currarini et al. (2009) and Bramoullé et al. (2012) for model of network formation featuring homophily). As a consequence, (formal or informal) employed workers will systematically transmit information to individuals of their own ethnic group, but they will communicate a job offer to someone of the other ethnic group with a probability  $\beta \in (0, 1)$ . In other words, the network is characterized by a bias making it more difficult to get information from someone of a different group (*inbreeding bias*), or to meet someone (*meeting bias*) of the other ethnic group.<sup>16</sup> As emphasized by Fearon and Laitin (1996), "ethnic groups are often characterized by relatively dense social networks and low-cost access to information".

<sup>12</sup>Also see Montgomery (1991), Calvó-Armengol and Jackson (2004) and Galenianos (2010) for theoretical work on the role of networks in the labor market.

<sup>13</sup>By contrast, Calvó-Armengol et al. (2007) assume that jobs can only be found through direct formal methods and that the information going through social contacts concerns crime opportunities only.

<sup>14</sup>The general idea is that workers are aware of job creations or vacancies in their own company.

<sup>15</sup>There is no on-the-job search in this model.

<sup>16</sup>Montgomery (1991) also assumes inbreeding bias, but along skills rather than ethnicity.

Bertrand et al. (2000) also provide empirical evidence of strong information transmission (about welfare benefits) within language group, at the neighborhood level, supporting the assumption that information circulate better within than across ethnic groups. We can also interpret  $\beta$  as an inverse measure of the degree of spatial segregation between the minority and the majority groups (or as the degree of integration of the minority group): the lower  $\beta$ , the higher the segregation, hence the lower the probability to find a job through someone from a different ethnic group.<sup>17</sup>

I finally close the description of labor market flows with job destruction, which is assumed exogenous in both sectors. I denote  $\delta_F$  the rate at which formal workers lose their job, and  $\delta_I$  the rate at which an undeclared worker is fired. I assume that  $\delta_I > \delta_F$  as the informal worker is not protected by any job contract, making it easier to lay him off. In addition, informal jobs can also be terminated if they are audited by fiscal authorities, which happens at a rate  $p$ . In the end, informal workers lose their job at rate  $\delta_I + p$ .

## 2.2 Flows of dyads between states

From what precedes, I characterize the net flows of the formal employment, unemployment and informal employment rates between  $t$  and  $t + \Delta t$  for each group  $j \in \{m, M\}$  as follows.

$$\begin{cases} \dot{F}_{j,t} = U_{j,t} \left( \frac{F_{j,t}}{N} + \beta \frac{F_{-j,t}}{N} \right) \alpha_F \phi_j - \delta_F F_{j,t} \\ \dot{U}_{j,t} = \delta_F F_{j,t} + (\delta_I + p) I_{j,t} - U_{j,t} \left[ \left( \frac{F_{j,t}}{N} + \beta \frac{F_{-j,t}}{N} \right) \alpha_F \phi_j + \left( \frac{I_{j,t}}{N} + \beta \frac{I_{-j,t}}{N} \right) \alpha_I \psi_{j,t} \right] \\ \dot{I}_{j,t} = U_{j,t} \left( \frac{I_{j,t}}{N} + \beta \frac{I_{-j,t}}{N} \right) \alpha_I \psi_j - (\delta_I + p) I_{j,t} \end{cases} \quad (4.2)$$

Let us explain the first equation, to get the general intuition for this system:  $\dot{F}_{j,t}$  is the change in the number of formally employed workers between  $t$  and  $t + \Delta t$ . At each point in time, formal jobs are destroyed at rate  $\delta_F$ . On the contrary, a formal job is created when an unemployed worker hears of a formal job opportunity. This happens if he is matched with a formal worker from any ethnic group (there is a probability  $\frac{F_j}{N}$  to meet a formal worker of group  $j$ ), who is aware of a job offer (which happens at rate  $\alpha_F$ ), and who transmits the information (with probability 1 if they belong to the same ethnic group, and  $\beta$  otherwise due to the in-group bias).<sup>18</sup>

<sup>17</sup>Zenou (2013) develops a similar model with two ethnic groups and one activity, embedded in a urban framework. He also introduces separation, but in a different manner: spatial segregation implies that agents further away from the business district meet less weak ties, whether belonging to the majority or to the minority group.

<sup>18</sup>I can also put it in the following way: he has a  $\frac{N_j}{N} f_{I_j}$  probability to meet a formal worker of his own group and a  $\beta \frac{N_{-j}}{N} f_{I_{-j}}$  probability to meet one of the other ethnic group. In this case, I can also view  $\beta$  as

In addition, once an unemployed worker hears of a formal job offer, he may still decide not to take the job and stay unemployed. This is embodied by the choice variable  $\phi_j$ , which is a dummy variable equal to one if the unemployed agent decides to accept the formal job offer, and equal to zero otherwise. Similarly, in the third equation,  $\psi_j$  is the dummy variable representing the decision to accept an undeclared job opportunity. These decision variables will be endogenized afterwards, when individuals' incentives are introduced.

Noting that  $\frac{F_{j,t}}{N} = \frac{F_{j,t}}{N} \frac{N_j}{N_j} = f_{j,t} n_j$ , and dividing both sides of each equation by  $N_j$ , system (4.2) can be rewritten as follows.

$$\begin{cases} \dot{f}_{j,t} = u_{j,t}(f_{j,t} n_j + \beta f_{-j,t} n_{-j}) \alpha_F \phi_j - \delta_F f_{j,t} \\ \dot{u}_{j,t} = \delta_F f_{j,t} + (\delta_I + p) i_{j,t} - u_{j,t} [(f_{j,t} n_j + \beta f_{-j,t} n_{-j}) \alpha_F \phi_j + (i_{j,t} n_j + \beta i_{-j,t} n_{-j}) \alpha_I \psi_j] \\ \dot{i}_{j,t} = u_{j,t} (i_{j,t} n_j + \beta i_{-j,t} n_{-j}) \alpha_I \psi_j - (\delta_I + p) i_{j,t} \end{cases} \quad (4.3)$$

Finally, using equation (4.1), this system reduces to the following two dynamic equations system.

$$\begin{cases} \dot{f}_{j,t} = u_{j,t}(f_{j,t} n_j + \beta f_{-j,t} n_{-j}) \alpha_F \phi_j - \delta_F f_{j,t} \\ \dot{i}_{j,t} = u_{j,t} (i_{j,t} n_j + \beta i_{-j,t} n_{-j}) \alpha_I \psi_j - (\delta_I + p) i_{j,t} \end{cases} \quad (4.4)$$

### 2.3 Incentives

As mentioned earlier, the choice variables  $\phi_j$  and  $\psi_j$  are endogenously determined taking into account the incentives faced by individuals in their decision making. Declared workers earn a wage  $w_F$ , and pay a unit tax  $t$  so that their net wage is  $w_F(1-t)$ . Unemployed workers receive a financial payoff  $b$ , which can be considered as the unemployment insurance, as social welfare or as the value of leisure.<sup>19</sup> I further assume that the net formal wage is larger than the unemployment benefit:  $w_F(1-t) > b$ . Finally, undeclared workers earn a wage  $w_I$ , and evade from paying taxes. However, they can be detected by fiscal authorities with probability  $p$ , which implies paying a fine  $f$  defined as a share of the wage perceived. Therefore, the expected informal wage is given by:  $w_I(1-pf)$ .

Agents make their decisions on the basis of their future employment prospects. In the long-run, individual values for each possible labor market status are given by the

a *meeting* bias, as in [Currarini et al. \(2009\)](#). This has a slightly different interpretation: in the case of the *inbreeding* bias,  $\beta$  means that one is less willing to transmit information to a member of a different group, while in the case of the *meeting* bias,  $\beta$  means that one is less likely to meet a member of the other ethnic group. In both cases, the consequence is that information circulates better within than across ethnic groups.

<sup>19</sup>The last two may be a better interpretation of  $b$ , as individuals are unlikely to receive unemployment insurance after performing undeclared work.



following Bellman equations, where  $V_j^S$  denotes the expected lifetime utility of a member of group  $j \in \{m, M\}$  involved in sector  $S \in \{F, I, U\}$ .

$$rV_j^F = w_F(1-t) + \delta_F(V_j^U - V_j^F) \quad (4.5)$$

$$rV_j^U = b + (f_{j,t}n_j + \beta f_{-j,t}n_{-j})\alpha_F\phi_j(V_j^F - V_j^U) + (i_{j,t}n_j + \beta i_{-j,t}n_{-j})\alpha_I\psi_j(V_j^I - V_j^U) \quad (4.6)$$

$$rV_j^I = w_I(1-pf) + (\delta_I + p)(V_j^U - V_j^I) \quad (4.7)$$

In these equations,  $r$  denotes the interest rate. Equation (4.5) tells that the lifetime expected utility of a formally employed worker is equal to his wage, plus the net value of losing his job (*i.e.* the value of being unemployed compared to the value of being in a formal job). The other equations have a similar interpretation.

### 3 Steady-state analysis

At a steady-state  $(f_j^*, u_j^*, i_j^*)$  the number of workers in each sector (and by extension the various employment rates) is stable, (*i.e.* the net flows in each employment status is equal to zero), and I get the following relationships using the set of dynamic equations (4.4).

$$f_j^* = \frac{(f_j^*n_j + \beta f_{-j}^*n_{-j})\alpha_F\phi_j}{\delta_F} u_j^* \quad (4.8)$$

$$i_j^* = \frac{(i_j^*n_j + \beta i_{-j}^*n_{-j})\alpha_I\psi_j}{\delta_I + p} u_j^* \quad (4.9)$$

$$u_j^* = 1 - f_j^* - i_j^* \quad (4.10)$$

There exist sixteen different types of steady-states depending on the values of the formal and informal employment rates (e.g. null or strictly positive). However, they can be reduced to the following ten steady-states by symmetry between  $j$  and  $-j$ ,  $\forall j \in \{m, M\}$ :

(i)	Full unemployment	$f_j^* = f_{-j}^* = 0$	$i_j^* = i_{-j}^* = 0$
(ii)	One group in the formal sector only	$f_j^* > 0, f_{-j}^* = 0$	$i_j^* = i_{-j}^* = 0$
(iii)	Both groups in the formal sector only	$f_j^* > 0, f_{-j}^* > 0$	$i_j^* = 0, i_{-j}^* = 0$
(iv)	One group in the informal sector only	$f_j^* = f_{-j}^* = 0$	$i_j^* > 0, i_{-j}^* = 0$
(v)	One group in formal and informal sectors	$f_j^* > 0, f_{-j}^* = 0$	$i_j^* > 0, i_{-j}^* = 0$
(vi)	One group in each sector	$f_j^* = 0, f_{-j}^* > 0$	$i_j^* > 0, i_{-j}^* = 0$
(vii)	Both groups in the formal, only one in the informal	$f_j^* > 0, f_{-j}^* > 0$	$i_j^* > 0, i_{-j}^* = 0$
(viii)	Both groups in the informal sector only	$f_j^* = f_{-j}^* = 0$	$i_j^* > 0, i_{-j}^* > 0$
(ix)	Both groups in the informal, only one in the formal	$f_j^* > 0, f_{-j}^* = 0$	$i_j^* > 0, i_{-j}^* > 0$
(x)	Both groups in both sectors	$f_j^* > 0, f_{-j}^* > 0$	$i_j^* > 0, i_{-j}^* > 0$

I now characterize each of the ten steady-states presented above: I indicate the values of the choice variables  $\phi_j$  and  $\varphi_j$  compatible with each steady-state and derive expressions of the steady-state employment rates. Appendix 4.A proves the results presented in this section.

- (i) A full unemployment steady-state ( $f_j^* = f_{-j}^* = 0$  and  $i_j^* = i_{-j}^* = 0$ ) when  $\phi_j, \phi_{-j}, \psi_j$  and  $\psi_{-j} \in \{0, 1\}$ .
- (ii) Steady-states where one group is fully unemployed, while the other takes part in the formal economy only ( $f_j^* > 0, f_{-j}^* = 0$  and  $i_j^* = i_{-j}^* = 0$ ) when  $\phi_j = 1, \phi_{-j} = 0$  and  $\psi_j, \psi_{-j} \in \{0, 1\}$ . In this case,  $f_j^* = 1 - \frac{1}{n_j} \frac{\delta_F}{\alpha_F}$ .<sup>20</sup>
- (iii) A no-informal economy steady-state ( $f_j^* > 0, f_{-j}^* > 0, i_j^* = 0$  and  $i_{-j}^* = 0$ ) when  $\phi_j = \phi_{-j} = 1$  and  $\psi_j, \psi_{-j} \in \{0, 1\}, \forall j$ . In this case, the equilibrium values of the formal employment rates are characterized by  $f_j^* = \frac{(f_j^* n_j + \beta f_{-j}^* n_{-j}) \alpha_F}{\delta_F + (f_j^* n_j + \beta f_{-j}^* n_{-j}) \alpha_F} < 1, \forall j \in \{m, M\}$ .
- (iv) Steady-states where one group is fully unemployed, while the other takes part in the informal economy only ( $f_j^* = f_{-j}^* = 0, i_j^* > 0$  and  $i_{-j}^* = 0$ ) when  $\phi_j, \phi_{-j} \in \{0, 1\}, \psi_j = 1$  and  $\psi_{-j} = 0$ . In this case,  $i_j^* = 1 - \frac{1}{n_j} \frac{\delta_I + p}{\alpha_I}$ .
- (v) Steady-states where one group is fully unemployed while the other takes part in both employment sectors ( $f_j^* > 0, f_{-j}^* = 0, i_j^* > 0$  and  $i_{-j}^* = 0$ ) when  $\phi_j = 1, \phi_{-j} = 0, \psi_j = 1$  and  $\psi_{-j} = 0$ . In this steady state,  $f_j^*$  and  $i_j^*$  are characterized by  $1 - f_j^* - i_j^* = \frac{1}{n_j} \frac{\delta_F}{\alpha_F}$  and  $1 - f_j^* - i_j^* = \frac{1}{n_j} \frac{\delta_I + p}{\alpha_I}$ , implying that the relative destruction rate is the same in both sectors ( $\frac{\delta_F}{\alpha_F} = \frac{\delta_I + p}{\alpha_I}$ ).
- (vi) Steady-states where one group works in the informal sector only while the other takes part in the formal labor market only ( $f_j^* = 0, f_{-j}^* > 0, i_j^* > 0$  and  $i_{-j}^* = 0$ ) when  $\phi_j = 0, \phi_{-j} = 1, \psi_j = 1, \psi_{-j} = 0$ . In this case, we have the following values:

$$f_j^* = 0, i_j^* = 1 - \frac{1}{n_j} \frac{\delta_I + p}{\alpha_I} \text{ and } u_j^* = \frac{1}{n_j} \frac{\delta_I + p}{\alpha_I}$$

$$f_{-j}^* = 1 - \frac{1}{n_{-j}} \frac{\delta_F}{\alpha_F}, i_{-j}^* = 0 \text{ and } u_{-j}^* = \frac{1}{n_{-j}} \frac{\delta_F}{\alpha_F}$$

- (vii) Steady-states where one group is involved in both sectors, while the other works in the formal sector only ( $f_j^* > 0, f_{-j}^* > 0, i_j^* > 0$  and  $i_{-j}^* = 0$ ) when  $\phi_j = \phi_{-j} = 1, \psi_j = 1,$

<sup>20</sup>Obviously, because job offers circulate through the network, the larger the size of the group  $n_j$ , the larger  $f_j^*$ . Note then that if the group that has a positive formal employment rate ( $j$ ) is the minority group, its formal employment rate is always smaller than the formal employment rate of the majority group in the case where the majority group would be the one to have a positive formal employment rate.

and  $\psi_{-j} = 0$ . In this case, the employment rates are determined as follows.

$$f_j^* = \beta \left[ \frac{n_{-j}}{n_j} \frac{\alpha_F(\delta_I + p)}{\alpha_I \delta_F - \alpha_F(\delta_I + p)} - \frac{1}{n_j} \frac{\delta_F(\delta_I + p)}{\alpha_I \delta_F - (1 - \beta^2)\alpha_F(\delta_I + p)} \right] \quad (4.11)$$

$$f_{-j}^* = 1 - \frac{1}{n_{-j}} \frac{\delta_F}{\alpha_F} \frac{\alpha_I \delta_F - \alpha_F(\delta_I + p)}{\alpha_I \delta_F - (1 - \beta^2)\alpha_F(\delta_I + p)} \quad (4.12)$$

$$i_j^* = 1 - \frac{1}{n_j} \frac{\delta_I + p}{\alpha_I} - f_j^* \quad (4.13)$$

(viii) An informal economy steady-state ( $f_j^* = f_{-j}^* = 0$  and  $i_j^* > 0$ ,  $i_{-j}^* > 0$ ) when  $\phi_j, \phi_{-j} \in \{0, 1\}$ ,  $\psi_j = 1$  and  $\psi_{-j} = 1$ . In this case, the equilibrium values of the informal employment rates are characterized by  $i_j^* = \frac{(i_j^* n_j + \beta i_{-j}^* n_{-j})\alpha_I}{\delta_I + p + (i_j^* n_j + \beta i_{-j}^* n_{-j})\alpha_I} < 1$ ,  $\forall j \in \{m, M\}$ .

(ix) Steady-states where one group is involved in both sectors, while the other works in the informal sector only ( $f_j^* > 0$ ,  $i_j^* > 0$  and  $f_{-j}^* = 0$ ,  $i_{-j}^* > 0$ ) when  $\phi_j = 1$ ,  $\phi_{-j} = 0$ ,  $\psi_j = 1$  and  $\psi_{-j} = 1$ . In this case, the employment rates are determined as follows.

$$i_j^* = \beta \left[ \frac{n_{-j}}{n_j} \frac{\alpha_I \delta_F}{\alpha_F(\delta_I + p) - \alpha_I \delta_F} - \frac{1}{n_j} \frac{\delta_F(\delta_I + p)}{\alpha_F(\delta_I + p) - (1 - \beta^2)\alpha_I \delta_F} \right]$$

$$i_{-j}^* = 1 - \frac{1}{n_{-j}} \frac{\delta_I + p}{\alpha_I} \frac{\alpha_F(\delta_I + p) - \alpha_I \delta_F}{\alpha_F(\delta_I + p) - (1 - \beta^2)\alpha_I \delta_F}$$

$$f_j^* = 1 - \frac{1}{n_j} \frac{\delta_F}{\alpha_F} - i_j^*$$

(x) A mixed steady-state ( $f_j^* > 0$ ,  $f_{-j}^* > 0$ ,  $i_j^* > 0$  and  $i_{-j}^* > 0$ ) when  $\phi_j = \phi_{-j} = 1$  and  $\psi_j = \psi_{-j} = 1$ . In this case, the employment rates are characterized as follows,  $\forall j$ .

$$f_j^* = \frac{(f_j^* n_j + \beta f_{-j}^* n_{-j})\alpha_F}{\delta_F} (1 - f_j^* - i_j^*)$$

$$i_j^* = \frac{(i_j^* n_j + \beta i_{-j}^* n_{-j})\alpha_I}{\delta_I + p} (1 - f_j^* - i_j^*)$$

However, some of these steady-states never arise in equilibrium. The next section is devoted to define and characterize the various equilibria. In particular, I derive the conditions under which the steady-states presented above arise in equilibrium.

## 4 Equilibria

### 4.1 Unemployed workers decisions

The conditions of existence of the various equilibria are derived from the conditions determining the value of the two choice variables  $\phi_j$  and  $\psi_j$ , *i.e.* the conditions under which unemployed workers are willing to accept or reject a formal or an informal job offer. To do so, I simply compare the lifetime expected utility of an unemployed worker to the one he would get from accepting a formal or an informal job offer. Combining equations (4.5), (4.6) and (4.7), I get:

$$V_j^U - V_j^F = \frac{[b - w_f(1-t)][r + \delta_I + p + (i_j n_j + \beta i_{-j} n_{-j})\alpha_I \psi_j] + [w_I(1-pf) - b][(i_j n_j + \beta i_{-j} n_{-j})\alpha_I \psi_j]}{(r + \delta_F)[r + \delta_I + p + (i_j n_j + \beta i_{-j} n_{-j})\alpha_I \psi_j] + (r + \delta_I + p)[(f_j n_j + \beta f_{-j} n_{-j})\alpha_F \phi_j]} \quad (4.14)$$

and

$$V_j^I - V_j^U = \frac{[w_I(1-pf) - b][r + \delta_F + p + (f_j n_j + \beta f_{-j} n_{-j})\alpha_F \phi_j] + [b - w_f(1-t)][(f_j n_j + \beta f_{-j} n_{-j})\alpha_F \phi_j]}{(r + \delta_F)[r + \delta_I + p + (i_j n_j + \beta i_{-j} n_{-j})\alpha_I \psi_j] + (r + \delta_I + p)[(f_j n_j + \beta f_{-j} n_{-j})\alpha_F \phi_j]} \quad (4.15)$$

Equation (4.14) helps determine the conditions under which an unemployed agent will decide to accept a formal job offer: he will accept such an offer if the lifetime expected value of this declared job is larger than the lifetime expected value of unemployment, *i.e.* if  $V_j^U - V_j^F \leq 0$ . Similarly, an unemployed worker will decide to accept an informal job offer when  $V_j^I - V_j^U \geq 0$ . Equation (4.15) is useful to find the conditions under which this inequality holds. More precisely, after noting that the denominator of the above equations is always positive and re-arranging the numerators, we can derive the following conditions.

$$\phi_j = 1 \Leftrightarrow V_j^F \geq V_j^U \Leftrightarrow (r + \delta_I + p)[w_f(1-t) - b] \geq [w_I(1-pf) - w_f(1-t)](i_j n_j + \beta i_{-j} n_{-j})\alpha_I \psi_j \quad (4.16)$$

and

$$\psi_j = 1 \Leftrightarrow V_j^I \geq V_j^U \Leftrightarrow (r + \delta_F)[w_I(1-pf) - b] \geq [w_f(1-t) - w_I(1-pf)](f_j n_j + \beta f_{-j} n_{-j})\alpha_F \phi_j \quad (4.17)$$

Both results are rather intuitive: from equation (4.16), we see that an unemployed worker will accept a formal job offer if the net financial gain from becoming a declared worker is large enough compared to the informal wage net of the formal wage. Similarly, (4.17) tells us that for an unemployed worker to accept an informal job offer, the net financial gain from becoming an undeclared worker must be large enough compared to the formal wage net of the informal wage. Otherwise stated, an unemployed agent compares his immediate net gain (left hand side terms) to the gain he would make in the outside option, would he refuse the offer in one sector to wait for an offer in the other sector (right hand side terms).

## 4.2 Equilibria

I now examine the conditions under which the various steady-states arise in equilibrium. The results are proved in Appendix 4.B. Before taking a closer look at the equilibria, it is important to underline that four steady-states never arise in equilibrium:

- Steady-states (ii) where one group is fully unemployed, while the other takes part in the formal economy only ( $f_j^* > 0, f_{-j}^* = 0$  and  $i_j^* = i_{-j}^* = 0, \forall j$ ),
- Steady-states (iv) where one group is fully unemployed, while the other takes part in the informal economy only ( $f_j^* = f_{-j}^* = 0, i_j^* > 0$  and  $i_{-j}^* = 0, \forall j$ ),
- Steady-states (v) where one group is fully unemployed while the other takes part in both employment sectors ( $f_j^* > 0, f_{-j}^* = 0, i_j^* > 0$  and  $i_{-j}^* = 0, \forall j$ ),
- Steady-states (vi) where one group works in the informal sector only while the other takes part in the formal labor market only ( $f_j^* = 0, f_{-j}^* > 0, i_j^* > 0$  and  $i_{-j}^* = 0, \forall j$ ).

The first three cases correspond to situations where one group takes part in at least one of the two employment sectors, while the other group is fully unemployed. The reason why this type of situation cannot happen is very intuitive and related to the information transmission mechanisms. Because there is always at least some information transmission across groups (except in the very extreme case where  $\beta = 0$ ), it is not possible that a group never hears about any job offer (and hence end up fully unemployed) when part of the other group is occupied.<sup>21</sup> The intuition for the last case (vi) not arising in equilibrium is a bit less direct, and is related to incentives in addition to information transmission mechanisms. The idea is that both groups have similar incentives, as they face the same wages and the same exogenous transition rates between employment status. So there is no reason why a group would engage in formal work while the other would go toward informality. Let us now turn to the equilibria. We have two different sets of equilibria arising according to the value of the relative net wage.

---

<sup>21</sup>In an earlier version of this model, I assume that individuals from different ethnic group do not share informal job information, *i.e.* that  $\beta = 0$  in the informal sector. The idea behind this assumption was the following. Because undeclared workers act beyond the law and face a risk of detection and punishment for evading taxes, they do not want to spread this kind of information too widely. Otherwise stated, they share undeclared job opportunities with people they really trust, and I assumed this could not happen outside one's own ethnic group. With this assumption, steady-state (iv) could arise in equilibrium.

• **Suppose first that  $w_F(1-t) > w_I(1-pf)$ :**

- The full unemployment steady-state (i), *i.e.*  $f_j^* = 0$  and  $i_j^* = 0$ ,  $\forall j \in \{m, M\}$  arises in equilibrium if  $b \leq w_F(1-t)$ , that is always.
- The no informal sector steady-state (iii), *i.e.*  $f_j^* > 0$  and  $i_j^* = 0$ ,  $\forall j \in \{m, M\}$  arises in equilibrium if  $b \leq w_F(1-t)$ , that is always.
- Steady-states (vii) where one group is involved in both sectors, while the other works in the formal sector only, *i.e.*  $f_j^* > 0$ ,  $i_j^* > 0$ ,  $f_{-j}^* > 0$  and  $i_{-j}^* = 0$ ,  $\forall j \in \{m, M\}$  arises in equilibrium as long as

$$\begin{aligned} w_I(1-pf) - [w_F(1-t) - w_I(1-pf)](n_{-j}f_{-j} + \beta n_j f_j) \frac{\alpha_F}{r + \delta_F} < b \\ \leq w_I(1-pf) - [w_F(1-t) - w_I(1-pf)](n_j f_j + \beta n_{-j} f_{-j}) \frac{\alpha_F}{r + \delta_F} \end{aligned}$$

with  $j$  and  $-j$  such that  $n_j f_j < n_{-j} f_{-j}$ .

- The no formal sector steady-state (viii), *i.e.*  $f_j^* = 0$  and  $i_j^* > 0$ ,  $\forall j \in \{m, M\}$  arises in equilibrium if  $b \leq w_I(1-pf)$ .
- The mixed steady-state (x), *i.e.*  $f_j^* > 0$  and  $i_j^* > 0$ ,  $\forall j \in \{m, M\}$  arises in equilibrium if

$$b \leq w_I(1-pf) - [w_F(1-t) - w_I(1-pf)](n_{-j}f_{-j} + \beta n_j f_j) \frac{\alpha_F}{r + \delta_F}$$

with  $j$  and  $-j$  is such that  $n_j f_j < n_{-j} f_{-j}$ .

• **Suppose now that  $w_F(1-t) \leq w_I(1-pf)$ :**

- The full unemployment steady-state (i), where  $f_j^* = 0$  and  $i_j^* = 0$ ,  $\forall j \in \{m, M\}$  arises in equilibrium if  $b \leq w_F(1-t)$ , that is always.
- The no informal sector steady-state (iii), where  $f_j^* > 0$  and  $i_j^* = 0$ ,  $\forall j \in \{m, M\}$  arises in equilibrium if  $b \leq w_F(1-t)$ , that is always.
- The no formal sector steady-state (viii), *i.e.*  $f_j^* = 0$  and  $i_j^* > 0$ ,  $\forall j \in \{m, M\}$  arises in equilibrium if  $b \leq w_F(1-t)$ , that is always.
- Steady-states (ix) where one group is involved in both sectors, while the other works in the informal sector only, *i.e.*  $f_j^* > 0$ ,  $i_j^* > 0$ ,  $f_{-j}^* = 0$  and  $i_{-j}^* > 0$ ,  $\forall j \in \{m, M\}$  arises in equilibrium as long as

$$\begin{aligned} w_F(1-t) - [w_I(1-pf) - w_F(1-t)](n_{-j}i_{-j} + \beta n_j i_j) \frac{\alpha_I}{r + \delta_I + p} < b \\ \leq w_F(1-t) - [w_I(1-pf) - w_F(1-t)](n_j i_j + \beta n_{-j} i_{-j}) \frac{\alpha_I}{r + \delta_I + p} \end{aligned}$$

with  $j$  and  $-j$  are such that  $n_j i_j < n_{-j} i_{-j}$ .

### 4.3 Intuitions and comments

A few comments are in order here. We can first observe that the full-unemployment steady-state always arises in equilibrium, whatever the relative net wage. This feature directly comes from the set up of the model, and is more precisely due to the simplifying assumption that job offers go only through the network. Under this assumption, it is indeed impossible that an unemployed worker hears about a job offer when there are no workers in the economy, so that a situation where no agent is employed in any sector is sustainable. Note that the same argument applies to equilibria (iii) and (viii) in which the employment rate in a given sector is null for both groups. Of course, if the model allowed for part of the job offers to arrive exogenously and independently of the network -as is the case in [Calvó-Armengol et al. \(2007\)](#)- this steady-state would never be an equilibrium.

Interestingly, the no-informal steady-state (iii) always arises in equilibrium, whether the net formal wage is larger than the net informal wage or not. The former situation is easy to understand: the formal wage is attractive enough to make unemployed workers systematically accept formal job offers when they find one. In addition, because job offers only circulate through word of mouth communication, the absence of workers in the informal sector is sustainable as explained above. The case where the informal wage is larger holds for the same reasons, even though slightly less intuitive: the absence of informal workers prevents informal job finding, and workers always accept formal job offers which are still more attractive than unemployment. However, this situation is somehow realistic as one could hardly assume the existence of exogenous informal job offers (e.g. employment agencies or adverts).

Let me now jump to the no-formal-sector steady-state (viii), which is symmetric to case (iii) in terms of activity. As expected, it always arises as an equilibrium when the net wage is larger in the informal sector, for the same reasons that (iii) always arises when the net wage is larger in the formal sector (*i.e.* appropriate financial incentives combined with the lack of outside options due to the absence of worker in the other sector). Yet, when the formal wage is larger than the informal one, the no-formal-sector equilibrium only exists if  $b$  is sufficiently low, and more precisely lower than the informal wage  $w_I(1 - pf)$ . The intuition is that with such a low unemployment benefit, individuals prefer accepting an informal job offer rather than staying unemployed and waiting for a better job offer. It is even trivial in this peculiar equilibrium where more financially attractive offers (formal job offers) are never to arrive.<sup>22</sup>

---

<sup>22</sup>This idea matches the notion of unemployment as a "waiting room" developed in [Calvó-Armengol et al. \(2007\)](#).

Let me now turn to the asymmetric equilibrium (vii) in which one group is engaged in both sectors while the other specializes in the formal sector.<sup>23</sup> The first thing to note is that this equilibrium arises only when the formal wage is larger than the informal wage. Thus, even if the informal sector is less attractive in terms of (net) wages than the formal sector, one of the two groups engages into it. This equilibrium therefore illustrates the intuition that the informal sector may be used as an economic safety net when formal employment prospects are low.

In particular, this equilibrium arises for a given range of  $b$ , which is such that  $b < w_I(1 - pf)$ .<sup>24</sup> Therefore, because the unemployment benefit is so low, some unemployed individuals are more likely to accept an informal job offer when they receive one (even if they would rather turn to the formal sector) instead of staying unemployed longer. However, even if the wage differential is the same for both groups, we know from Section 3 that they do not take the same decision *vis-à-vis* informal job offers: group  $j$  always accepts ( $\psi_j = 1$ ), group  $-j$  always declines ( $\psi_{-j} = 0$ ). This divergent behavior can be explained by differences in the relative rate of access to a given sector, which depends on the job opening and job destruction rates ( $\alpha_S$  and  $\delta_S$ ,  $S \in \{F, I\}$ ), but also on the size of the network. Re-arranging relation (4.17) indicates that a group  $j$  individual accepts an informal job offer if  $[w_I(1 - pf) - b] \geq [w_F(1 - t) - w_I(1 - pf)](f_j n_j + \beta f_{-j} n_{-j}) \frac{\alpha_F}{r + \delta_F}$ , that is if the present net gain from accepting this offer is larger than the expected net gain from formal employment, would he wait for a formal offer instead. So, in this equilibrium, group  $j$ 's probability to receive a formal job offer is so low that its members always accept informal job offers, while group  $-j$  members receive formal job offers frequently enough to afford declining an informal offer and wait for a formal one.

Crucially, I show that in this equilibrium, the group that engages into the informal sector is always the minority group. This is proved in Appendix 4.C. Once again, this is in line with the informal sector acting as an economic safety net for individuals belonging to the minority: because their network is less developed, they face adverse labor market conditions, that is they have a low probability of finding a formal job so that they will accept informal job offers in order to avoid staying unemployed.

I finally want to insist on the role of the unemployment benefit in defining the equilibria. Let us consider the case where the formal wage is larger than the informal wage. We can see that all the equilibria in which at least one group engages in the informal sector ((vii), (viii) and (x)) are equilibria in which the unemployment benefit is under a given

<sup>23</sup>Equilibrium (ix) in which one group is engaged in both sectors while the other specializes in the informal sector perfectly mirrors case (vii) in terms of sectors. Thus, all the comments made for (vii) apply by symmetry to (ix).

<sup>24</sup>It is direct to show that the upper bound for  $b$  displayed in subsection (4.2) is strictly lower than  $w_I(1 - pf)$  when  $w_F(1 - t) > w_I(1 - pf)$ .



threshold. More generally, all the conditions under which these equilibria emerge are such that  $b \leq w_I(1 - pf)$ . As explained above, in this situation unemployed workers are more likely to accept any job offer, even an informal one. Therefore, if unemployment benefits were larger, unemployed workers incentives would be affected such that they could afford declining informal job offers and wait until they hear about a formal one (*i.e.* switch to a different equilibrium).

## 5 Further results and comparative statics

In this section, I look more precisely at the equilibria characteristics in terms of group composition, focusing on the case where the net formal wage is larger than the net informal wage. The alternative case is not considered here, but most of the results and similar reasoning can be applied by symmetry. I also let the trivial full unemployment equilibrium (i) aside, and concentrate on equilibria (iii) and (vii).

An important feature of equilibrium (iii) is that the majority group always has a larger employment rate than the minority group ( $f_M^* \geq f_m^*$ ), due to the size of the network. To put it differently, the minority group always faces a larger unemployment rate. A formal proof of this result is given in Appendix 4.D. The intuition is simply that the wider one's network, the larger one's probability to meet another individual, hence the larger the probability to hear about a job offer.<sup>25</sup> Because the network features homophily, the transmission of information is imperfect across groups, and hence members of the majority group always have access to a larger network than the minority group.

A subsequent intuition is that the employment rates should increase with the degree of integration between the two groups, *i.e.* be larger for smaller values of  $\beta$ . This equilibrium is quite difficult to solve analytically, as  $f_j^*$  and  $f_{-j}^*$  are determined by two equations of two unknowns, and are the roots of third order polynomials (see Appendix 4.A for more details). Yet, it is easy and informative to look at the two extreme sub-cases of complete segregation ( $\beta = 0$ ) and no segregation ( $\beta = 1$ ). When  $\beta = 0$ , *i.e.* when there is no transmission of information across groups, we obtain  $f_m^* = 1 - \frac{\delta_F}{\alpha_F} \frac{1}{n_m}$  and  $f_M^* = 1 - \frac{\delta_F}{\alpha_F} \frac{1}{n_M}$ . Obviously,  $f_M^* > f_m^*$ . As job information can only come from one's own group,  $f_j$  is directly proportional to  $n_j$ . When  $\beta = 1$ , *i.e.* when information circulates perfectly between ethnic groups, we get  $f_m^* = f_M^* = 1 - \frac{\delta_F}{\alpha_F} \frac{1}{(n_m + n_M)} = 1 - \frac{\delta_F}{\alpha_F}$ . The equality between the employment rates of the two groups is very intuitive as information circulates perfectly across groups so that both groups form a unique large group. Note that the no-segregation employment rate is larger than the full-segregation employment rate for both groups, because each individual not only has access to information from his own group, but also from the other one. More

<sup>25</sup>This result is consistent with one of those in Zenou (2013) (proposition 9).

importantly, the increase in the employment rate between the full segregation and the no segregation cases is larger for the minority group. This is due to the fact that when segregation decreases, the new information network to which the minority group has access is relatively larger than that of the majority group (which has already access to a large information network in the full-segregation case). In other words, the increase in the size of the majority group information network is marginal. Finally, these two simple cases reveal that the employment rates decrease with the relative exit rate out of employment  $\frac{\delta_F}{\alpha_F}$ , as foreseeable.<sup>26</sup>

Let me now turn to equilibrium (vii), in which one group specializes in the formal sector, while the other has to engage in the informal sector as a result of low formal employment prospects. This equilibrium is more convenient to study: the various employment rates can be analytically solved for, as shown in Section 3. Using equations (4.11) to (4.13), I can show that:

$$\frac{\partial f_{-j}^*}{\partial \beta} > 0, \frac{\partial f_j^*}{\partial \beta} > 0 \text{ and } \frac{\partial i_j^*}{\partial \beta} < 0$$

Appendix 4.E proves these results. These findings confirm the intuition developed in the previous paragraph that less homophily or less segregation (*i.e.* a larger  $\beta$ ) increases the formal employment rate of each group. Once again, the underlying mechanism is that improving communication across groups enlarges individuals' network, as they gain access to the other group's information. Then information about formal job offers circulates more easily, raising formal employment prospects. More interestingly, these results reveal that reducing segregation actually lowers informal employment, although informal job offers also transit through word of mouth communication. This is because in this equilibrium, increasing  $\beta$  does not only enlarge the network, but also increases the probability that workers from group  $j$  (the group engaged in the informal sector) get access to formal job offers by interacting with members of group  $-j$ . Conversely, even if group  $-j$  members also become more exposed to informal job offers by interacting more with group  $j$  members, they are still in a situation where the incentives to decline such offers dominate.

This result is at odds with one of the findings of [Calvó-Armengol et al. \(2007\)](#) in a comparable setting. In their model, agents all belong to the same group and can either be employed, unemployed or criminals. Each individual is related to a unique other agent with whom s/he is *strongly* tied, and is otherwise weakly related to any other agent. At each period, agents are randomly matched by pair with their dyad partner or with any

<sup>26</sup>By symmetry with case (iii), it is also possible to show that in equilibrium (viii) the informal employment rate is always larger for the majority group.

other agent, according to some given probability. One of the findings of the paper is that increasing the frequency of matching with weak ties increases the crime rate. This result stems from the fact that only crime opportunities circulate through the network: job offers arrive exogenously (e.g. employment agencies) and are not transmitted through word of mouth. By contrast, I have shown that if the network plays a role in both sectors (employment and crime or formal and informal employment), then increasing social interactions can lower the participation in the "bad" sector provided that the alternative activity is attractive.

I finally look at how a policy aiming at directly fighting informal employment actually affects the various employment rates. In this model, such a policy corresponds to an increase in the audit rate  $p$ . Unsurprisingly, I can show that:

$$\frac{\partial f_{-j}^*}{\partial p} > 0, \frac{\partial f_j^*}{\partial p} > 0 \text{ and } \frac{\partial i_j^*}{\partial p} < 0$$

However, the mechanisms behind these results differ from those explaining the effect of  $\beta$ . Variations in  $\beta$  alter the job information transmission process, and hence impact the various employment rates by influencing the probability of unemployed to hear about a job offer. On the other hand, variations in  $p$  directly affect the incentives to engage in the informal economy in two ways. An increase in the apprehension rate both increases the rate at which informal jobs are terminated and decreases the net informal wage, affecting both the transition rates and the financial incentives. The effect of a rise in the audit rate on the formal employment rate is therefore indirect, and comes from the reduced incentives to informal work, while the effect of lower segregation directly improves the formal employment rate through facilitated information transmission. Hence, an important difference between an increase in  $p$  and an increase in  $\beta$  is that the former increases the unemployment rate of the targeted group (the one participating in the informal economy), while the latter does not. This implies that policies improving social cohesion and communication across groups could be more efficient and in any case less costly in terms of unemployment than direct policies aiming at destroying informal jobs. This is particularly relevant in this equilibrium where unemployment is the least remunerating activity.

## 6 Conclusion and discussion

The simple theoretical framework used in this paper helps rationalizing the idea of the informal economy as an economic safety net and the over-representation of minorities in the underground economy. The model predicts that even when the informal sector is less attractive than the formal one in terms of wages, equilibria where some agents

engage in undeclared work may yet arise. More specifically, I show that all equilibria involving the informal sector arise when unemployment benefits are too low. In this case, some unemployed workers are better off accepting informal job offers than staying unemployed in order to wait for a formal job offer. In particular, in the equilibrium where only one group turns informal, this group is the minority group. The rationale for this directly lies in the process of job offers transmission and hence in the nature of social interactions. Because majority workers have access to a large network, they expect to receive formal job offers frequently enough to afford turning down informal job offers and staying unemployed a little longer. Conversely, minority workers receive too few formal offers to decline informal ones.

This model also proves insightful in terms of policy implications, would the planner's objective be to reduce the size of the informal sector. It suggests that increasing the level of the unemployment benefit (or more generally social welfare) could allow workers to remain unemployed and wait for a formal job offer rather than immediately accepting informal job offers. The interplay between formal employment, unemployment and informal employment would thus deserve some attention when designing traditional employment policies. The model also shows that policies reducing homophily, that is enhancing inter-groups relationships and improving immigrants integration tend to favor formal over informal employment. More importantly, such policies turn out to be neutral in terms of unemployment, as they directly amplify the flow toward the formal sector. By contrast, policies directly combating undeclared jobs through more frequent audits seem less efficient: they do limit the size of the informal economy, but at the expense of a larger unemployment rate. Indeed, such policies boost the flow out of informality, but have no direct impact on the flow into formal employment.

Naturally, this model is not exempt from drawbacks. First, it does not feature a proper network in the sense of a graph with well-defined links between any two agents. Instead, in my model, any two agents can be connected or not at each period, and in this sense all agents are symmetric. Therefore, I am unable to draw conclusions regarding the place of an agent in the network (e.g. whether s/he is at the center or at the periphery of the graph), which has been proved quite relevant in the networks literature. Although this would undoubtedly refine the findings of the model, I am confident that the main results would be unaltered. Another concern may come from the fact that job offers, and in particular formal ones circulate only through the network. As already discussed, this assumption is highly unrealistic, and dropping it should significantly alter the dynamics of the model. In particular the exogenous rate of arrival formal job offers could be large enough to allow the minority workers to decline informal offers and remain unemployed until they receive a formal one. This of course could be mitigated if the model included

the demand side of the labor market, which is obviously an important limitation of the model. Yet, even in this case, a possible and credible refinement would be to introduce discrimination, so that minority workers would receive fewer exogenous offers or be less often hired following a match. I am therefore convinced that most of the previous results would hold in this setting.

Finally, an interesting refinement would be to allow for different modes of communication across employment sectors. In particular, transmission of informal job offers could depend on trust. As pointed out by [Portes and Haller \(2005\)](#), trust is a key element in informal transactions: in as much as the informal economy is characterized by the lack of any written contract, trust is the only way to sustain informally established relationships and to overcome commitment problems. As it turns out, sharing a common identity, and especially belonging to the same ethnic group is one of the best ways to generate such trust in informal exchanges.<sup>27</sup> Indeed, not only sanctions threats are more credible and hence more efficient within ethnic groups ([Miguel and Gugerty, 2005](#)), but migrants to a country also tend to be more socially cohesive: established contacts in the destination country is often a pre-requisite to immigration, and, once they arrive in a given country, migrants strengthen their links by developing community-based institutions ([Munshi, 2003](#)). Therefore, I could introduce a trust parameter that would allow informal job offers to circulate more easily among minority workers than among majority workers. This would most likely reinforce my results.

## 4.A Characterization of the steady-states

Using equation (4.10) in equations (4.8) and (4.9), we get:

$$f_j^* = \frac{(f_j^* n_j + \beta f_{-j}^* n_{-j}) \alpha_F \phi_j}{\delta_F} (1 - f_j^* - i_j^*) \quad (4.18)$$

and

$$i_j^* = \frac{(i_j^* n_j + \beta i_{-j}^* n_{-j}) \alpha_I \psi_j}{\delta_I + p} (1 - f_j^* - i_j^*) \quad (4.19)$$

The values of  $\phi_j$  and  $\psi_j$  for which these two relationships are verified are derived below, for each steady-state.

(i) For  $f_j^* = 0$  and  $i_j^* = 0$ ,  $\forall j \in \{m, M\}$ , it is immediate to see that equations (4.18) and (4.19) are always verified for any value of  $\phi_j$  and  $\psi_j$  ( $\forall \phi_j, \psi_j \in \{0, 1\}$ ),  $\forall j \in \{m, M\}$ .

<sup>27</sup>[Greif \(1993\)](#) also established that the 11th century Maghribi traders organized into a coalition to overcome the commitment problem.

(ii) For  $f_j^* > 0$ ,  $f_{-j}^* = 0$  and  $i_j^* = i_{-j}^* = 0$ , equation (4.18) is verified for  $j$  if  $\phi_j = 1$ , and for  $-j$  if  $\phi_j = 0$ , and (4.19) is verified  $\forall \psi_j \in \{0, 1\}$ . Solving for equation (4.18) then leads to  $f_j^* = 1 - \frac{\delta_F}{\alpha_F} \frac{1}{n_j} < 1$ . Note that we must have  $\frac{\delta_F}{\alpha_F} < n_j$  to ensure that  $f_j^* > 0$ .

(iii) For  $f_j^* > 0$ ,  $f_{-j}^* > 0$  and  $i_j^* = i_{-j}^* = 0$ , equation (4.19) is satisfied  $\forall \psi_j \in \{0, 1\}$ ,  $\forall j \in \{m, M\}$ , and we must have  $\phi_j = \phi_{-j} = 1$  for equation (4.18) to hold. We can then rewrite equation (4.18) as follows,  $\forall j \in \{m, M\}$ :

$$f_j^* = \frac{(n_j f_j^* + \beta n_{-j} f_{-j}^*) \alpha_F}{\delta_F + (n_j f_j^* + \beta n_{-j} f_{-j}^*) \alpha_F} < 1 \quad (4.20)$$

In this equilibrium  $f_j^*$  and  $f_{-j}^*$  are particularly difficult to obtain analytically as they are determined by two equations of two unknowns (equation 4.20 for  $j$  and for  $-j$ ). Combining the two equations reveal that  $f_j^*$  actually solves  $\Phi(f_j^*) = 0$ , where  $\Phi(x)$  is a third order polynomial of the form:

$$\Phi(x) = A(1-x)^2 x + B(1-x)^2 + C(1-x)x + D(1-x) + Ex + F,$$

where  $A, B, C, D, E$  and  $F$  are constants made of the parameters  $n_j, n_{-j}, \beta, \alpha_F$  and  $\delta_F$ .

(iv) For  $f_j^* = 0$ ,  $f_{-j}^* = 0$ ,  $i_j^* > 0$  and  $i_{-j}^* = 0$ ,  $\forall j \in \{m, M\}$ : as in case (i), equation (4.18) is verified  $\forall \phi_j \in \{0, 1\}$ ,  $\forall j \in \{m, M\}$ . Then, equation (4.19) holds when  $\psi_j = 1$  and  $\psi_{-j} = 0$ . By symmetry with case (ii), we have  $i_j = 1 - \frac{\delta_{I+P}}{\alpha_I} \frac{1}{n_j}$ .

(v) For  $f_j^* > 0$ ,  $f_{-j}^* = 0$ ,  $i_j^* > 0$  and  $i_{-j}^* = 0$ ,  $\forall j \in \{m, M\}$ , we must have  $\phi_j = 1$ ,  $\phi_{-j} = 0$ ,  $\psi_j = 1$  and  $\psi_{-j} = 0$  for equations (4.18) and (4.19) to hold. Re-arranging these equations leads to  $1 - f_j^* - i_j^* = \frac{\delta_F}{\alpha_F} \frac{1}{n_j}$  and  $1 - f_j^* - i_j^* = \frac{\delta_{I+P}}{\alpha_I} \frac{1}{n_j}$ , which in turn implies that  $\frac{\delta_F}{\alpha_F} = \frac{\delta_{I+P}}{\alpha_I}$ .

(vi) For  $f_j^* = 0$ ,  $f_{-j}^* > 0$ ,  $i_j^* > 0$  and  $i_{-j}^* = 0$ ,  $\forall j \in \{m, M\}$ , we must have  $\phi_j = 0$ ,  $\phi_{-j} = 1$ ,  $\psi_j = 1$  and  $\psi_{-j} = 0$  for equations (4.18) and (4.19) to be satisfied. Then, we are back to case (iv) for group  $j$  with  $i_j = 1 - \frac{\delta_{I+P}}{\alpha_I} \frac{1}{n_j}$ , and to case (ii) for group  $-j$  with  $f_{-j}^* = 1 - \frac{\delta_F}{\alpha_F} \frac{1}{n_{-j}}$ .

(vii) For  $f_j^* > 0$ ,  $f_{-j}^* > 0$ ,  $i_j^* > 0$ ,  $i_{-j}^* = 0$ ,  $\forall j \in \{m, M\}$ , equations (4.18) and (4.19) are verified with  $\phi_j = 1$ ,  $\phi_{-j} = 1$ ,  $\psi_j = 1$  and  $\psi_{-j} = 0$ . In this case, equation (4.19) for group  $j$  becomes:  $i_j^* = \frac{n_j i_j^* \alpha_I}{\delta_{I+P}} (1 - f_j^* - i_j^*)$ . Dividing both sides by  $i_j^*$  and re-arranging leads to  $i_j^* = 1 - f_j^* - \frac{\delta_{I+P}}{\alpha_I} \frac{1}{n_j}$ . We can then plug this expression of  $i_j^*$  in equation (4.18) for group  $j$  to obtain:  $f_j^* = \frac{(n_j f_j^* + \beta n_{-j} f_{-j}^*) \alpha_F}{\delta_F + (n_j f_j^* + \beta n_{-j} f_{-j}^*) \alpha_F} \frac{\delta_{I+P}}{\alpha_I} \frac{1}{n_j}$ . From this, we can express  $f_j^*$  as a function of

$f_{-j}^*$ :

$$f_j^* = \beta \frac{n_{-j}}{n_j} \frac{\alpha_F(\delta_I + p)}{\alpha_I \delta_F - \alpha_F(\delta_I + p)} f_{-j}^* \quad (4.21)$$

Note that because both sides must be positive, it imposes that  $\frac{\alpha_I}{\delta_I + p} > \frac{\alpha_F}{\delta_F}$ , *i.e.* that the net job arrival rate is larger in the informal sector. Finally, using the previous expression of  $f_j^*$  in equation (4.18) for group  $-j$  and re-arranging, allow me to derive the value of  $f_{-j}^*$  at this steady-state:

$$f_{-j}^* = 1 - \frac{\delta_F}{\alpha_F} \frac{\alpha_I \delta_F - \alpha_F(\delta_I + p)}{\alpha_I \delta_F - (1 - \beta^2) \alpha_F(\delta_I + p)} \frac{1}{n_{-j}}$$

Hence the value of  $f_j^*$ :

$$f_j^* = \beta \left[ n_{-j} \frac{\alpha_F(\delta_I + p)}{\alpha_I \delta_F - \alpha_F(\delta_I + p)} - \frac{\delta_F(\delta_I + p)}{\alpha_I \delta_F - (1 - \beta^2) \alpha_F(\delta_I + p)} \right] \frac{1}{n_j}$$

We can also deduce the value of  $i_j^*$  from the following relationship:  $i_j^* = 1 - \frac{\delta_I + p}{\alpha_I} \frac{1}{n_j} - f_j^*$ .

(viii) For the no formal sector steady state  $f_j^* = 0$ ,  $f_{-j}^* = 0$ ,  $i_j^* > 0$  and  $i_{-j}^* > 0$ ,  $\forall j \in \{m, M\}$  to exist, equations (4.18) and (4.19) hold  $\forall \phi_j, \phi_{-j} \in \{0, 1\}$  and require that  $\psi_j = \psi_{-j} = 1$ ,  $\forall j$ . By symmetry with case (iii) we deduce the following relationship:  $i_j^* = \frac{(n_j i_j^* + \beta n_{-j} i_{-j}^*) \alpha_I}{\delta_I + p + (n_j i_j^* + \beta n_{-j} i_{-j}^*) \alpha_I}$ ,  $\forall j \in \{m, M\}$ .

(ix) For  $f_j^* > 0$ ,  $f_{-j}^* = 0$ ,  $i_j^* > 0$  and  $i_{-j}^* > 0$   $\forall j \in \{m, M\}$ , equations (4.18) and (4.19) are satisfied if  $\phi_j = 1$ ,  $\phi_{-j} = 0$  and  $\psi_j = \psi_{-j} = 1$ ,  $\forall j$ . This case is symmetric to case (vii) in terms of activities. Re-arranging equation (4.18) for group  $j$  leads to  $f_j^* = 1 - i_j^* - \frac{\delta_F}{\alpha_F} \frac{1}{n_j}$ . Then, plugging this expression of  $f_j^*$  in equation (4.19) for group  $j$  allows us to express  $i_j^*$  as a function of  $i_{-j}^*$ :  $i_j^* = \beta \frac{n_{-j}}{n_j} \frac{\alpha_I \delta_F}{\alpha_F(\delta_I + p) - \alpha_I \delta_F} i_{-j}^*$ . Note that because both sides must be positive, it imposes that  $\frac{\alpha_I}{\delta_I + p} < \frac{\alpha_F}{\delta_F}$ , *i.e.* that the net job arrival rate is larger in the formal sector. Finally, using the previous expression of  $i_j^*$  in equation (4.19) for group  $-j$  and re-arranging, allow me to derive the value of  $i_{-j}^*$  at this steady-state:

$$i_{-j}^* = 1 - \frac{\delta_I + p}{\alpha_I} \frac{\alpha_F(\delta_I + p) - \alpha_I \delta_F}{\alpha_F(\delta_I + p) - (1 - \beta^2) \alpha_I \delta_F} \frac{1}{n_{-j}}$$

Hence the value of  $i_j^*$ :

$$i_j^* = \beta \left[ n_{-j} \frac{\alpha_I \delta_F}{\alpha_F(\delta_I + p) - \alpha_I \delta_F} - \frac{\delta_F(\delta_I + p)}{\alpha_F(\delta_I + p) - (1 - \beta^2) \alpha_I \delta_F} \right] \frac{1}{n_j}$$

We can also deduce the value of  $f_j^*$  from the following relationship:  $f_j^* = 1 - \frac{\delta_F}{\alpha_F} \frac{1}{n_j} - i_j^*$ .

(x) For the mixed steady-state  $f_j^* > 0$  and  $i_j^* > 0$ ,  $\forall j \in \{m, M\}$  to arise, equations (4.18) and (4.19) require that  $\phi_j = \phi_{-j} = 1$  and  $\psi_j = \psi_{-j} = 1$ . Then, we obtain the following relationships.

$$f_j^* = (n_j f_j^* + \beta n_{-j} f_{-j}^*) \frac{\alpha_F}{\delta_F} (1 - f_j^* - i_j^*), \quad \forall j \in \{B, W\}$$

and

$$i_j^* = (n_j i_j^* + \beta n_{-j} i_{-j}^*) \frac{\alpha_I}{\delta_I + p} (1 - f_j^* - i_j^*), \quad \forall j \in \{B, W\}$$

## 4.B Equilibria

Remember that  $w_F(1-t) > b$  by assumption. Using equations (4.16) and (4.17), I now provide the conditions for the different values of the choice variables to arise in equilibrium. In what follows, I denote  $w_F(1-t)$  as  $w_F$  and  $w_I(1-pf)$  as  $w_I$  to simplify the expressions.

(i) The full unemployment steady-state where  $f_j^* = 0$  and  $i_j^* = 0$ ,  $\forall j \in \{m, M\}$ , exists  $\forall \phi_j, \psi_j \in \{0, 1\}$ ,  $\forall j \in \{m, M\}$ . Given the assumption  $w_F > b$ , I cannot have  $\phi_j = 0$ , which requires that  $(r + \delta_I + p)[w_F - b] < 0$  from equation (4.16). So this steady-state cannot arise in equilibrium for  $\phi_j = 0$ ,  $\forall j \in \{m, M\}$ . Conversely, the condition for  $\phi_j = 1$  when  $i_j^* = i_{-j}^* = 0$  is given by:  $(r + \delta_I + p)[w_F - b] \geq 0$ , equivalent to  $w_F \geq b$ , which is always true. Note that the condition for  $\phi_{-j} = 1$  is exactly the same, hence I cannot have  $\phi_j \neq \phi_{-j}$ . So, we must have  $\phi_j = \phi_{-j} = 1$  for steady-state (i) to arise in equilibrium. I then consider two different cases according to the value of  $\psi_j$ :

- ◇ Case  $\psi_j = 1$  requires that  $(r + \delta_F)[w_I - b] \geq 0$  as  $f_j^* = f_{-j}^* = 0$ , equivalent to  $w_I \geq b$ . Note that we cannot have  $\psi_j \neq \psi_{-j}$  either. Combining what precedes, we can state that the full unemployment steady-state (i) arises with  $\phi_j = \phi_{-j} = \psi_j = \psi_{-j} = 1$ ,  $\forall j \in \{m, M\}$  if  $b \leq \min\{w_F, w_I\}$ .
- ◇ Case  $\psi_j = \psi_{-j} = 0$  requires  $w_I(1-pf) < b$ . Combining what precedes, I can state that the full unemployment steady-state (i) arises with  $\phi_j = \phi_{-j} = 1$ , and  $\psi_j = \psi_{-j} = 0$  if  $w_I < b \leq w_F$ .

To summarize, the full-unemployment steady-state (i) always arises in equilibrium:

- ◇ when  $w_F > w_I$ , (i) arises in equilibrium:
  - if  $b \leq w_I$ , with  $\phi_j = \phi_{-j} = \psi_j = \psi_{-j} = 1$ ,  $\forall j \in \{m, M\}$ , or
  - if  $w_I < b \leq w_F$ , with  $\phi_j = \phi_{-j} = 1$  and  $\psi_j = \psi_{-j} = 0$ ,  $\forall j \in \{m, M\}$ .
- ◇ when  $w_F \leq w_I$ , (i) arises in equilibrium if  $b \leq w_F$ , with  $\phi_j = \phi_{-j} = \psi_j = \psi_{-j} = 1$ ,  $\forall j \in \{m, M\}$ .



(ii) The steady-state where one group is fully unemployed while the other takes part in the formal labor market only ( $f_j^* > 0$ ,  $f_{-j}^* = 0$  and  $i_j^* = i_{-j}^* = 0$ ,  $\forall j \in \{m, M\}$ ) requires that  $\phi_j = 1$ ,  $\phi_{-j} = 0$  (and  $\forall \psi_j \in \{0, 1\}$ ). Given that  $i_j^* = 0$ , the condition for  $\phi_{-j} = 0$  (equation 4.16) becomes  $(r + \delta_I + p)[w_F - b] < 0$ , which is impossible. Hence this steady-state never arises in equilibrium.

(iii) The no informal sector steady-state ( $f_j^* > 0$ ,  $f_{-j}^* > 0$  and  $i_j^* = i_{-j}^* = 0$ ), exists for  $\phi_j = \phi_{-j} = 1$  and  $\forall \psi_j \in \{0, 1\}$ ,  $\forall j \in \{m, M\}$ . The condition for  $\phi_j = \phi_{-j} = 1$  is always satisfied as  $i_j^* = i_{-j}^* = 0$ ,  $\forall j \in \{m, M\}$ . We can then consider three different cases according to the values of  $\psi_j$  and  $\psi_{-j}$ :

- ◇ If  $\psi_j = \psi_{-j} = 1$ , the conditions from equation (4.16) are given by  $b \leq w_I - (w_F - w_I)(n_j f_j^* + \beta n_{-j} f_{-j}^*) \frac{\alpha_F}{r + \delta_F}$ ,  $\forall j$ . Combining with the condition on  $\phi_j$  (*i.e.*  $b \leq w_F$ ), we get the following condition:

$$b \leq \min \left\{ \begin{array}{l} w_I - (w_F - w_I)(n_j f_j^* + \beta n_{-j} f_{-j}^*) \frac{\alpha_F}{r + \delta_F} \\ w_I - (w_F - w_I)(n_{-j} f_{-j}^* + \beta n_j f_j^*) \frac{\alpha_F}{r + \delta_F} \\ w_F \end{array} \right\}$$

It is easily shown that when  $w_F \leq w_I$ ,  $w_F$  is the minimum of the right-hand side term (henceforth RHS) in the above expression, while when  $w_F > w_I$ , the minimum is  $w_I - (w_F - w_I)(n_j f_j^* + \beta n_{-j} f_{-j}^*) \frac{\alpha_F}{r + \delta_F}$ , where  $j$  and  $-j$  are such that  $n_j f_j^* > n_{-j} f_{-j}^*$ .

- ◇ If  $\psi_j = \psi_{-j} = 0$ , the various conditions become:

$$\max \left\{ \begin{array}{l} w_I - (w_F - w_I)(n_j f_j^* + \beta n_{-j} f_{-j}^*) \frac{\alpha_F}{r + \delta_F} \\ w_I - (w_F - w_I)(n_{-j} f_{-j}^* + \beta n_j f_j^*) \frac{\alpha_F}{r + \delta_F} \end{array} \right\} < b \leq w_F$$

We can deduce from the previous case that  $w_F > w_I$ . The maximum of the left-hand side term (henceforth LHS) is  $w_I - (w_F - w_I)(n_{-j} f_{-j}^* + \beta n_j f_j^*) \frac{\alpha_F}{r + \delta_F}$ , where  $j$  and  $-j$  are such that  $n_j f_j^* > n_{-j} f_{-j}^*$ .

- ◇ If  $\psi_j = 0$  and  $\psi_{-j} = 1$ , all the conditions put together give:

$$w_I - (w_F - w_I)(n_j f_j^* + \beta n_{-j} f_{-j}^*) \frac{\alpha_F}{r + \delta_F} < b \leq \min \left\{ \begin{array}{l} w_I - (w_F - w_I)(n_{-j} f_{-j}^* + \beta n_j f_j^*) \frac{\alpha_F}{r + \delta_F} \\ w_F \end{array} \right\}$$

From the previous cases, we know that when  $w_F > w_I$ , then  $w_F$  is the maximum of the RHS. On the contrary, when  $w_F \leq w_I$ , we get a contradiction as  $w_F$  cannot be the minimum of the RHS and larger than the LHS at the same time.

To summarize, steady-state (iii) always arises in equilibrium:

◇ when  $w_F > w_I$ :

- if  $b \leq w_I - (w_F - w_I)(n_j f_j^* + \beta n_{-j} f_{-j}^*) \frac{\alpha_F}{r + \delta_F}$ , with  $\phi_j = \phi_{-j} = 1$  and  $\psi_j = \psi_{-j} = 1$ ,
  - if  $w_I - (w_F - w_I)(n_j f_j^* + \beta n_{-j} f_{-j}^*) \frac{\alpha_F}{r + \delta_F} < b \leq w_I - (w_F - w_I)(n_{-j} f_{-j}^* + \beta n_j f_j^*) \frac{\alpha_F}{r + \delta_F}$ , with  $\phi_j = \phi_{-j} = 1$ ,  $\psi_j = 0$  and  $\psi_{-j} = 1$ ,
  - if  $w_I - (w_F - w_I)(n_{-j} f_{-j}^* + \beta n_j f_j^*) \frac{\alpha_F}{r + \delta_F} < b \leq w_F$ , with  $\phi_j = \phi_{-j} = 1$  and  $\psi_j = \psi_{-j} = 0$ ,
- where  $j$  and  $-j$  are such that  $n_j f_j^* > n_{-j} f_{-j}^*$ .

◇ when  $w_F \leq w_I$ : if  $b \leq w_F$  (always true) with  $\phi_j = \phi_{-j} = 1$  and  $\psi_j = \psi_{-j} = 1$ .

(iv) The steady-state where one group is fully unemployed while the other takes part in the informal labor market only ( $f_j^* = 0$ ,  $f_{-j}^* = 0$ ,  $i_j^* > 0$  and  $i_{-j}^* = 0$ ,  $\forall j \in \{m, M\}$ ) exists for  $\psi_j = 1$ ,  $\psi_{-j} = 0$  and  $\forall \phi_j, \phi_{-j} \in \{0, 1\}$ . The condition (equation 4.17) for  $\psi_j = 1$  writes  $(r + \delta_F)(w_I - b) \geq 0$ , while that for  $\psi_{-j} = 0$  writes  $(r + \delta_F)(w_I - b) < 0$ , hence a contradiction. We can then conclude that this steady-state never arises in equilibrium.

(v) The steady-state where one group is fully unemployed while the other work on both formal and informal markets ( $f_j^* > 0$ ,  $f_{-j}^* = 0$ ,  $i_j^* > 0$  and  $i_{-j}^* = 0$ ,  $\forall j \in \{m, M\}$ ) exists for  $\phi_j = 1$ ,  $\phi_{-j} = 0$ ,  $\psi_j = 1$  and  $\psi_{-j} = 0$ . Given that  $i_{-j}^* = 0$ , the condition for  $\phi_{-j} = 0$  (equation 4.16) cannot be satisfied under the assumption that  $w_F > b$ . So this steady-state never arises in equilibrium.

(vi) The steady-state where one group specializes in the formal economy while the other specializes in the informal economy ( $f_j^* = 0$ ,  $f_{-j}^* > 0$ ,  $i_j^* > 0$  and  $i_{-j}^* = 0$ ,  $\forall j \in \{m, M\}$ ) exists for  $\phi_j = 0$ ,  $\phi_{-j} = 1$ ,  $\psi_j = 1$ , and  $\psi_{-j} = 0$ . The conditions for the previous four equalities to be satisfied can be summarized as follows:

$$\max \left\{ \begin{array}{l} w_F - (w_I - w_F) n_j i_j^* \frac{\alpha_I}{r + \delta_I + p} \\ w_F - (w_I - w_F) n_{-j} i_{-j}^* \frac{\alpha_I}{r + \delta_I + p} \end{array} \right\} < b \leq \min \left\{ \begin{array}{l} w_I \\ w_F \end{array} \right\}$$

It is easy to check that for any of the LHS term to be smaller than  $w_I$ , we must have  $w_F < w_I$ , while for any of the LHS term to be smaller than  $w_F$ , we must have  $w_F > w_I$ , hence a contradiction. Therefore, we can conclude that this steady-state never arises in equilibrium.

(vii) Steady-states where one group is involved in both sectors, while the other works in the formal sector only ( $f_j^* > 0$ ,  $f_{-j}^* > 0$ ,  $i_j^* > 0$ , and  $i_{-j}^* = 0$ ,  $\forall j \in \{m, M\}$ ), exist for

$\phi_j = \phi_{-j} = \psi_j = 1$  and  $\psi_{-j} = 0$ . These four conditions can be summarized as follows.

$$w_I - (w_F - w_I)(n_{-j}f_{-j}^* + \beta n_j f_j^*) \frac{\alpha_F}{r + \delta_F} < b \leq \min \left\{ \begin{array}{l} w_F - (w_I - w_F)n_j i_j^* \frac{\alpha_I}{r + \delta_I + p} \\ w_F \\ w_I - (w_F - w_I)(n_j f_j^* + \beta n_{-j} f_{-j}^*) \frac{\alpha_F}{r + \delta_F} \end{array} \right\}$$

It is easily checked that for the LHS term to be smaller than any of the RHS terms, we must have  $w_F > w_I$ , and  $j$  and  $-j$  such that  $n_j f_j^* < n_{-j} f_{-j}^*$ . Then, under the previous condition that  $w_F > w_I$ , it is easy to show that  $w_I - (w_F - w_I)(n_j f_j^* + \beta n_{-j} f_{-j}^*) \frac{\alpha_F}{r + \delta_F}$  is the minimum of the RHS.

To summarize, steady-state (vii) arises in equilibrium with  $\phi_j = \phi_{-j} = \psi_j = 1$  and  $\psi_{-j} = 0$ , when  $w_F > w_I$ , if

$$w_I - (w_F - w_I)(n_{-j}f_{-j}^* + \beta n_j f_j^*) \frac{\alpha_F}{r + \delta_F} < b \leq w_I - (w_F - w_I)(n_j f_j^* + \beta n_{-j} f_{-j}^*) \frac{\alpha_F}{r + \delta_F}$$

where  $j$  and  $-j$  are such that  $n_j f_j^* < n_{-j} f_{-j}^*$ .

**(viii)** The no formal sector steady state  $f_j^* = 0$ ,  $f_{-j}^* = 0$ ,  $i_j^* > 0$  and  $i_{-j}^* > 0$ ,  $\forall j \in \{m, M\}$  exists  $\forall \phi_j, \phi_{-j} \in \{0, 1\}$  and for  $\psi_j = \psi_{-j} = 1$ . The condition for  $\psi_j = \psi_{-j} = 1$  implies that  $w_I \geq b$ . We can then consider three different cases:

$$\diamond \text{ If } \phi_j = \phi_{-j} = 1, \text{ the conditions are given by } b \leq \min \left\{ \begin{array}{l} w_F - (w_I - w_F)(n_j i_j^* + \beta n_{-j} i_{-j}^*) \frac{\alpha_I}{r + \delta_I + p} \\ w_F - (w_I - w_F)(n_{-j} i_{-j}^* + \beta n_j i_j^*) \frac{\alpha_I}{r + \delta_I + p} \\ w_I \\ w_F \end{array} \right\}$$

It is easily shown that when  $w_F \leq w_I$ , the minimum of the RHS is  $w_F$ . Conversely, when  $w_F > w_I$ ,  $w_I$  is the minimum of the RHS.

$\diamond$  If  $\phi_j = \phi_{-j} = 0$ , the condition becomes

$$\max \left\{ \begin{array}{l} w_F - (w_I - w_F)(n_j i_j^* + \beta n_{-j} i_{-j}^*) \frac{\alpha_I}{r + \delta_I + p} \\ w_F - (w_I - w_F)(n_{-j} i_{-j}^* + \beta n_j i_j^*) \frac{\alpha_I}{r + \delta_I + p} \end{array} \right\} < b \leq \min \left\{ \begin{array}{l} w_I \\ w_F \end{array} \right\}$$

Here we have a contradiction as the LHS terms cannot be smaller than  $w_I$  and  $w_F$  at the same time. Hence this steady-state cannot arise if  $\phi_j = \phi_{-j} = 0$ .

$\diamond$  If  $\phi_j = 0$  and  $\phi_{-j} = 1$ , all the conditions put together give:

$$w_F - (w_I - w_F)(n_j i_j^* + \beta n_{-j} i_{-j}^*) \frac{\alpha_I}{r + \delta_I + p} < b \leq \min \left\{ \begin{array}{l} w_F - (w_I - w_F)(n_{-j} i_{-j}^* + \beta n_j i_j^*) \frac{\alpha_I}{r + \delta_I + p} \\ w_I \\ w_F \end{array} \right\}$$

From what precedes, we know that this leads to a contradiction. Hence this steady-state cannot arise if  $\phi_j = 0$  and  $\phi_{-j} = 1$ .

To summarize, steady-state (viii) arises in equilibrium:

◇ when  $w_F > w_I$ : if  $b \leq w_I$  with  $\phi_j = \phi_{-j} = 1$  and  $\psi_j = \psi_{-j} = 1$ .

◇ when  $w_F \leq w_I$ : if  $b \leq w_F$  (always true), with  $\phi_j = \phi_{-j} = 1$  and  $\psi_j = \psi_{-j} = 1$ .

(ix) Steady-states where one group is involved in both sectors, while the other works in the informal sector only ( $f_j^* > 0$ ,  $f_{-j}^* = 0$ ,  $i_j^* > 0$  and  $i_{-j}^* > 0, \forall j \in \{m, M\}$ ), exist for  $\phi_j = 1$ ,  $\phi_{-j} = 0$ , and  $\psi_j = \psi_{-j} = 1, \forall j$ . These four conditions, combined with the condition that  $b \leq w_F$  sum up to the following condition:

$$w_F - (w_I - w_F)(n_{-j}i_{-j}^* + \beta n_j i_j^*) \frac{\alpha_I}{r + \delta_I + p} < b \leq \min \left\{ \begin{array}{l} w_F - (w_I - w_F)(n_j i_j^* + \beta n_{-j} i_{-j}^*) \frac{\alpha_I}{r + \delta_I + p} \\ w_I - (w_F - w_I)n_j f_j^* \frac{\alpha_F}{r + \delta_F} \\ w_I \\ w_F \end{array} \right\}$$

We can check that for the LHS term to be smaller than any of the RHS terms, we must have  $w_F \leq w_I$ , and  $j$  and  $-j$  such that  $n_j i_j^* < n_{-j} i_{-j}^*$ . Then, we can immediately deduce that  $w_F - (w_I - w_F)(n_j i_j^* + \beta n_{-j} i_{-j}^*) \frac{\alpha_I}{r + \delta_I + p}$  is the minimum of the RHS.

To summarize, steady-state (ix) arises in equilibrium with  $\phi_j = 1$ ,  $\phi_{-j} = 0$ , and  $\psi_j = \psi_{-j} = 1$ , when  $w_F \leq w_I$ , if

$$w_F - (w_I - w_F)(n_{-j} i_{-j}^* + \beta n_j i_j^*) \frac{\alpha_I}{r + \delta_I + p} < b \leq w_F - (w_I - w_F)(n_j i_j^* + \beta n_{-j} i_{-j}^*) \frac{\alpha_I}{r + \delta_I + p}$$

where  $j$  and  $-j$  are such that  $n_j i_j^* < n_{-j} i_{-j}^*$ .

(x) The mixed steady-state ( $f_j^* > 0$  and  $i_j^* > 0, \forall j \in \{m, M\}$ ) exists for  $\phi_j = \phi_{-j} = 1$  and  $\psi_j = \psi_{-j} = 1$ . This leads to the following condition:

$$b \leq \min \left\{ \begin{array}{l} w_F - (w_I - w_F)(n_j i_j^* + \beta n_{-j} i_{-j}^*) \frac{\alpha_I}{r + \delta_I + p} \\ w_F - (w_I - w_F)(n_{-j} i_{-j}^* + \beta n_j i_j^*) \frac{\alpha_I}{r + \delta_I + p} \\ w_I - (w_F - w_I)(n_j f_j^* + \beta n_{-j} f_{-j}^*) \frac{\alpha_F}{r + \delta_F} \\ w_I - (w_F - w_I)(n_{-j} f_{-j}^* + \beta n_j f_j^*) \frac{\alpha_F}{r + \delta_F} \\ w_F \end{array} \right\}$$

The terms  $w_I - (w_F - w_I)(n_j f_j^* + \beta n_{-j} f_{-j}^*) \frac{\alpha_F}{r + \delta_F}, \forall j$  are the only terms that can be smaller to any other without getting a contradiction, when  $w_F > w_I$ . We can finally state that the minimum is given by  $w_I - (w_F - w_I)(n_j f_j^* + \beta n_{-j} f_{-j}^*) \frac{\alpha_F}{r + \delta_F}$ , with  $n_j f_j^* > n_{-j} f_{-j}^*$  and  $n_j i_j^* < n_{-j} i_{-j}^*$ .

## 4.C The minority group turns informal

When  $w_F(1-t) > w_I(1-pf)$ , the group that engages in the informal sector in equilibrium (vii) is the minority group:  $f_j^* > 0$ ,  $i_j^* > 0$ ,  $f_{-j}^* > 0$  and  $i_{-j}^* = 0$  with  $n_{-j} > n_j$ .

*Proof.* Using the expression of  $f_{-j}^*$  (4.12) and the fact that  $f_{-j}^* > 0$ , we can easily show that

$$n_{-j} > \frac{\delta_F}{\alpha_F} \frac{\alpha_I \delta_F - \alpha_F (\delta_I + p)}{\alpha_I \delta_F - (1 - \beta^2) \alpha_F (\delta_I + p)} \quad (4.22)$$

Note that this inequality can also be obtained using expression (4.11) and  $f_j^* > 0$  for instance.

The proof then consists in showing that the RHS in (4.22) is larger than one half, implying that  $n_{-j} > \frac{1}{2}$  and hence that the group working in the formal sector only is the majority group.

Using expression (4.21) and the fact that  $n_{-j} f_{-j}^* > n_j f_j^*$ , we find that  $\frac{\delta_F}{\alpha_F} > (1 + \beta) \frac{\delta_I + p}{\alpha_I}$ , which is equivalent to

$$2 \frac{\delta_F}{\alpha_F} > 2 + 2\beta \frac{\delta_I + p}{\alpha_I} \quad (4.23)$$

We can then show that the following relationship holds:

$$2 + 2\beta \frac{\delta_I + p}{\alpha_I} > 1 + \beta^2 \frac{\alpha_F (\delta_I + p)}{\alpha_I \delta_F - \alpha_F (\delta_I + p)} \quad (4.24)$$

Indeed, we know from (4.21) and  $n_{-j} f_{-j}^* > n_j f_j^*$  that  $\beta \frac{\alpha_F (\delta_I + p)}{\alpha_I \delta_F - \alpha_F (\delta_I + p)} < 1$ . This implies that  $\beta \frac{\alpha_F (\delta_I + p)}{\alpha_I \delta_F - \alpha_F (\delta_I + p)} - 2 \frac{\delta_I + p}{\alpha_I} < 1 - 2 \frac{\delta_I + p}{\alpha_I} < 1$ . Hence,  $\beta [\beta \frac{\alpha_F (\delta_I + p)}{\alpha_I \delta_F - \alpha_F (\delta_I + p)} - 2 \frac{\delta_I + p}{\alpha_I}] < 1$  as  $0 < \beta < 1$  which is equivalent to  $2 + 2\beta \frac{\delta_I + p}{\alpha_I} > 1 + \beta^2 \frac{\alpha_F (\delta_I + p)}{\alpha_I \delta_F - \alpha_F (\delta_I + p)}$ .

By transitivity, inequalities (4.23) and (4.24) finally imply that:

$$\begin{aligned} 2 \frac{\delta_F}{\alpha_F} &> 1 + \beta^2 \frac{\alpha_F (\delta_I + p)}{\alpha_I \delta_F - \alpha_F (\delta_I + p)} \\ \Leftrightarrow 2 \frac{\delta_F}{\alpha_F} &> \frac{\alpha_I \delta_F - (1 - \beta^2) \alpha_F (\delta_I + p)}{\alpha_I \delta_F - \alpha_F (\delta_I + p)} \\ \Leftrightarrow \frac{\delta_F}{\alpha_F} \frac{\alpha_I \delta_F - \alpha_F (\delta_I + p)}{\alpha_I \delta_F - (1 - \beta^2) \alpha_F (\delta_I + p)} &> \frac{1}{2} \end{aligned} \quad (4.25)$$

And therefore  $n_{-j} > \frac{1}{2} < n_j$  □

## 4.D Comparison of majority and minority employment rates

In equilibrium (iii),  $f_j^* > f_{-j}^*$  with  $n_j > n_{-j}$ .

*Proof.* In order to prove this result, we proceed in two steps. We first establish which group is the majority group and then show that the employment rate corresponding to this group is larger than the other. More precisely, we prove that  $n_j > n_{-j}$  and  $f_j^* > f_{-j}^*$ , when  $j$  and  $-j$  are such that  $n_j f_j^* > n_{-j} f_{-j}^*$  as established when defining equilibrium (iii) in Appendix 4.B.<sup>28</sup>

Let us first determine the majority group. Rearranging equation (4.20), we can write  $f_j^*$  as a function of  $f_{-j}^*$  and then express the relative employment rate as follows:

$$f_j^* = \frac{\delta_F - \alpha_F n_{-j} (1 - f_{-j}^*)}{\beta \alpha_F n_j} \frac{f_{-j}^*}{1 - f_{-j}^*} \Leftrightarrow \frac{f_j^*}{f_{-j}^*} = \frac{\delta_F - \alpha_F n_{-j} (1 - f_{-j}^*)}{\beta \alpha_F n_j (1 - f_{-j}^*)} \quad (4.26)$$

Because  $j$  and  $-j$  are such that  $n_j f_j^* > n_{-j} f_{-j}^*$ , we establish that

$$\frac{\delta_F - \alpha_F n_{-j} (1 - f_{-j}^*)}{\beta \alpha_F n_j} \frac{1}{1 - f_{-j}^*} > \frac{n_{-j}}{n_j}$$

which, after some simplification, is equivalent to:

$$n_{-j} (1 - f_{-j}^*) < \frac{\delta_F}{\alpha_F} \frac{1}{1 + \beta} \quad (4.27)$$

By symmetry between  $j$  and  $-j$ , we also have:

$$f_{-j}^* = \frac{\delta_F - \alpha_F n_j (1 - f_j^*)}{\beta \alpha_F n_{-j}} \frac{f_j^*}{1 - f_j^*} \Leftrightarrow \frac{f_{-j}^*}{f_j^*} = \frac{\beta \alpha_F n_{-j} (1 - f_j^*)}{\delta_F - \alpha_F n_j (1 - f_j^*)} > \frac{n_{-j}}{n_j} \quad (4.28)$$

$$\Leftrightarrow n_j (1 - f_j^*) > \frac{\delta_F}{\alpha_F} \frac{1}{1 + \beta} \quad (4.29)$$

Combining equations (4.27) and (4.29) I deduce that  $n_j (1 - f_j^*) > n_{-j} (1 - f_{-j}^*)$ , which is equivalent to  $n_j - n_{-j} > n_j f_j^* - n_{-j} f_{-j}^*$ . I know that the RHS is strictly positive, hence the LHS is positive, and  $n_j > n_{-j}$ .

<sup>28</sup>This "assumption" is just fixed so as to be able to distinguish between the two groups. Of course, we could derive different (but symmetric) conditions of existence of this equilibrium where  $j$  and  $-j$  would be such that  $n_j f_j^* < n_{-j} f_{-j}^*$ . Then, we would obtain the same general result, *i.e.* that the majority group has a larger employment rate, but where the majority group would be  $-j$  instead of  $j$  in the present demonstration.

Let us now show that  $f_j^* > f_{-j}^*$ , where  $j$  corresponds to the majority group. From equation (4.20) for  $j$  and  $-j$  I can write the relative employment rate as follows:

$$\begin{aligned} \frac{f_j^*}{f_{-j}^*} &= \frac{(n_j f_j^* + \beta n_{-j} f_{-j}^*) \alpha_F \delta_F + (n_j f_j^* + \beta n_{-j} f_{-j}^*) \alpha_F}{(n_{-j} f_{-j}^* + \beta n_j f_j^*) \alpha_F \delta_F + (n_{-j} f_{-j}^* + \beta n_j f_j^*) \alpha_F} \\ \Leftrightarrow \frac{f_j^*}{f_{-j}^*} &= \frac{\delta_F (n_j f_j^* + \beta n_{-j} f_{-j}^*) + (n_j f_j^* + \beta n_{-j} f_{-j}^*) (n_{-j} f_{-j}^* + \beta n_j f_j^*) \alpha_F}{\delta_F (n_{-j} f_{-j}^* + \beta n_j f_j^*) + (n_j f_j^* + \beta n_{-j} f_{-j}^*) (n_{-j} f_{-j}^* + \beta n_j f_j^*) \alpha_F} \end{aligned}$$

As the second term of the numerator's sum is the same as the second term of the denominator's sum and as they are both positive, I just need to compare the first terms of the numerator and of the denominator:

$$(n_j f_j^* + \beta n_{-j} f_{-j}^*) - (n_{-j} f_{-j}^* + \beta n_j f_j^*) = (1 - \beta)(n_j f_j^* - n_{-j} f_{-j}^*)$$

Because  $\beta \in (0, 1)$  and  $n_j f_j^* > n_{-j} f_{-j}^*$ , we have  $n_j f_j^* + \beta n_{-j} f_{-j}^* > n_{-j} f_{-j}^* + \beta n_j f_j^*$ . From equation (4.D), we then deduce that  $\frac{f_j^*}{f_{-j}^*} > 1 \Leftrightarrow f_j^* > f_{-j}^*$ , where  $j$  is the majority group.  $\square$

## 4.E Comparative statics

In equilibrium (vii),  $\frac{\partial f_{-j}^*}{\partial \beta} > 0$ ,  $\frac{\partial f_j^*}{\partial \beta} > 0$  and  $\frac{\partial i_j^*}{\partial \beta} < 0$

*Proof.* Using equation 4.12, we can compute the partial derivative of  $f_{-j}$  with respect to  $\beta$ . After simplification, we get:

$$\frac{\partial f_{-j}^*}{\partial \beta} = \frac{2\beta \delta_F (\delta_I + p) [\alpha_I \delta_F - \alpha_F (\delta_I + p)]}{n_{-j} [\alpha_I \delta_F - (1 - \beta^2) \alpha_F (\delta_I + p)]^2}$$

From the intermediate computations used to characterize (vii) in Appendix 4.A, we know that  $\alpha_I \delta_F - \alpha_F (\delta_I + p) > 0$ . All the other terms being positive, the result  $\frac{\partial f_{-j}^*}{\partial \beta} > 0$  directly follows.

In this same appendix, we expressed  $f_j^*$  as  $f_j^* = \beta \frac{n_{-j}}{n_j} \frac{\alpha_F (\delta_I + p)}{\alpha_I \delta_F - \alpha_F (\delta_I + p)} f_{-j}^*$ . Differentiating with respect to  $\beta$  gives:

$$\frac{\partial f_j^*}{\partial \beta} = \frac{n_{-j}}{n_j} \frac{\alpha_F (\delta_I + p)}{\alpha_I \delta_F - \alpha_F (\delta_I + p)} [f_{-j}(\beta) + \beta \frac{\partial f_{-j}^*}{\partial \beta}]$$

All terms being positive we conclude that  $\frac{\partial f_j^*}{\partial \beta} > 0$ .

Finally,  $\frac{\partial i_j^*}{\partial \beta} < 0$  immediately derives from equation (4.13) and the previous result.  $\square$

# Conclusion

As a conclusion, I would simply like to emphasize a few facts, share some thoughts, and give a few lines that I would like to follow in future research.

About one hundred and fifty pages ago, I asked the following question. Do individuals in more diverse neighborhoods fare better than individuals in less diverse neighborhoods? If I had to reply in one word and sticking strictly to the results, my answer would be "No": increased local origin-diversity is shown to lower the quality of local public goods (chapter 1) and to hinder employment prospects (chapter 3), while spatial economic heterogeneity seems to induce economic crimes (chapter 2). Yet, if one were to ask me - as many others did before - if I would consequently advocate for no diversity, I would answer likewise. Indeed, a critical by-product of this work is to shed some light on the underlying mechanisms, allowing me to moderate the core findings. It is particularly worth stressing that all the results related to diversity of origins seem to be driven by the lack of communication across groups and the inability to develop social norms within diverse communities. In chapter 1, we have seen that anomie can lead to lower housing conditions for two reasons. First, the absence of common rules and the inefficiency of social sanctions threats cannot prevent vandalism. Second, tenants are unable to undertake collective action that would pressure the housing owners to improve the quality of the basic housing facilities. The results in chapter 3 also suggests that the absence of communication across groups prevents job information transmission hence lowering employment prospects. In addition, because cultural diversity proves more important than ethnic diversity in explaining this outcome, it points further toward the importance of sharing a common culture and a common language. This idea is supported theoretically as well in chapter 4 which surmises that better communication across groups is beneficial.

Therefore, rather than promoting less diversity, this thesis speaks in favor of more integration. That is, individuals should be encouraged to be part of a society, accepting and conforming to its rules and values (usually those of the host country for immigrants), learning a common language, while preserving their differences. Allowing foreign residents to vote for local or even national elections could for instance encourage integra-



tion.<sup>29</sup> In addition, the process of assimilation is largely dependent on the structure and the composition of the population. [Lazear \(1999a\)](#) shows for instance that smaller minority groups will integrate faster because they cannot rely only on their co-ethnic and thus benefit more from acquiring the language and the culture of the host country. In an empirical study on UK data, [Manning and Roy \(2010\)](#) also find that some groups assimilate faster, especially when they come from countries that are the most dissimilar to the UK.

On the other hand, a large literature on segregation have revealed the potentially detrimental effects of the absence of diversity. [Cutler and Glaeser \(1997\)](#) performed the first analysis establishing a causal relationship between segregation and personal achievement. They find that segregation is costly for African-Americans, in terms of educational attainment, employment, wages, or single-motherhood. These effects may however be balanced by the educational composition of the group, as suggested by [Borjas \(1995\)](#) and shown by [Cutler et al. \(2008\)](#).

All in all, this discussion calls for a careful interpretation of the results, and an even more careful thinking regarding the potential policies induced. In particular, I should probably try to account for both diversity of origins and diversity in terms of educational attainments and social statuses in the neighborhood. In addition, this also raises the question of a tipping point, that should deserve more attention in my future research.

On a different matter, and as emphasized in chapter 2, a more systematic spatial approach could help understanding better some findings. This would apply to the study of criminal victimization, but also to that of employment prospects. Taking into account the urban structure of the neighborhoods, with the public transportation or road networks could also teach us a lot. This is part of what I plan to do next.

---

<sup>29</sup>In France, E.U. citizens only are allowed to vote at municipal and European elections.

# Bibliography

- Aeberhardt, R., Fougère, D., Pouget, J., and Rathelot, R. (2010). Wages and employment of french workers with african origin. *Journal of Population Economics*, 23(3):881–905.
- Alesina, A., Baqir, R., and Easterly, W. (1999). Public goods and ethnic divisions. *The Quarterly Journal of Economics*, 114(4):1243–1284.
- Alesina, A., Devleeschauwer, A., Easterly, W., Kurlat, S., and Wacziarg, R. (2003). Fractionalization. *Journal of Economic Growth*, 8(2):155–94.
- Alesina, A., Harnoss, J., and Rapoport, H. (2013). Birthplace diversity and economic prosperity. NBER Working Papers 18699, National Bureau of Economic Research, Inc.
- Alesina, A. and La Ferrara, E. (2000). Participation in heterogeneous communities. *The Quarterly Journal of Economics*, 115(3):847–904.
- Alesina, A. and La Ferrara, E. (2002). Who trusts others? *Journal of Public Economics*, 85(2):207–234.
- Alesina, A. and La Ferrara, E. (2005). Ethnic diversity and economic performance. *Journal of Economic Literature*, 43(3):762–800.
- Alesina, A. and Zhuravskaya, E. (2011). Segregation and the quality of government in a cross section of countries. *American Economic Review*, 101(5):1872–1911.
- Algan, Y., Bisin, A., Manning, A., and Verdier, T. (2012). *Cultural integration of immigrants in Europe*. Oxford University Press.
- Algan, Y., Dustmann, C., Gritz, A., and Manning, A. (2010). The economic situation of first and second-generation immigrants in france, germany and the united kingdom. *Economic Journal*, 120(542):F4–F30.
- Algan, Y., Hémet, C., and Laitin, D. (2013). The social effects of ethnic diversity at the local level: A natural experiment with exogenous residential allocation. Working Papers 2013-38, AMSE.
- Ashraf, Q. and Galor, O. (2013). The 'out of africa' hypothesis, human genetic diversity, and comparative economic development. *American Economic Review*, 103(1):1–46.
- Baganha, M. I. (2000). *The Economic Incorporation of the Immigrants in the Portuguese Informal Economy*, volume Southern Europe, vol. 2 of Towards Emerging Ethnic Classes in Europe? Eliseo Aja et al. Weinheim Germany : Freudenberg Stiftung.

- Battu, H., Seaman, P., and Zenou, Y. (2011). Job contact networks and the ethnic minorities. *Labour Economics*, 18(1):48–56.
- Bayer, P., Ross, S. L., and Topa, G. (2008). Place of work and place of residence: Informal hiring networks and labor market outcomes. *Journal of Political Economy*, 116(6):1150–1196.
- Becker, G. S. (1968). Crime and punishment: An economic approach. *Journal of Political Economy*, 76:169–217.
- Becker, G. S. (1971). *The Economics of Discrimination*. Number 9780226041162 in University of Chicago Press Economics Books. University of Chicago Press, 2nd edition.
- Bell, B., Machin, S., and Fasani, F. (2010). Crime and immigration: Evidence from large immigrant waves. IZA Discussion Papers 4996, Institute for the Study of Labor (IZA).
- Bernardot, M. (2008). *Loger les immigrés: La Sonacotra, 1956-2006*. Collection "Terra". Editions du Croquant.
- Bernasco, W. and Luykx, F. (2003). Effects of attractiveness, opportunity and accessibility to burglars on residential burglary rates of urban neighborhoods. *Criminology*, 41(3):981–1002.
- Bertrand, M., Luttmer, E. F. P., and Mullainathan, S. (2000). Network effects and welfare cultures. *The Quarterly Journal of Economics*, 115(3):1019–1055.
- Bianchi, M., Buonanno, P., and Pinotti, P. (2012). Do immigrants cause crime? *Journal of the European Economic Association*, 10(6):1318–1347.
- Bisin, A., Patacchini, E., Verdier, T., and Zenou, Y. (2011). Ethnic identity and labour market outcomes of immigrants in Europe. *Economic Policy*, 26(65):57–92.
- Boeri, T. and Garibaldi, P. (2002). Shadow activity and unemployment in a depressed labour market. CEPR Discussion Papers 3433, C.E.P.R. Discussion Papers.
- Boeri, T. and Garibaldi, P. (2005). Shadow sorting. In *NBER International Seminar on Macroeconomics 2005*, NBER Chapters, pages 125–163. National Bureau of Economic Research, Inc.
- Borjas, G. J. (1994). The economics of immigration. *Journal of Economic Literature*, 32(4):1667–1717.
- Borjas, G. J. (1995). Ethnicity, neighborhoods, and human-capital externalities. *American Economic Review*, 85(3):365–90.
- Borjas, G. J. (2003). The labor demand curve is downward sloping: Reexamining the impact of immigration on the labor market. *Quarterly Journal of Economics*, 118(4):1335–1374.
- Bound, J. and Johnson, G. (1992). Changes in the structure of wages in the 1980's: An evaluation of alternative explanations. *American Economic Review*, 82(3):371–92.

- Bramoullé, Y., Currarini, S., Jackson, M. O., Pin, P., and Rogers, B. W. (2012). Homophily and long-run integration in social networks. *Journal of Economic Theory*, 147(5):1754–1786.
- Bruinsma, G. J. N. (2007). Urbanization and urban crime: Dutch geographical and environmental research. *Crime and Justice*, 35(1):453–502.
- Buonanno, P., Montolio, D., and Vanin, P. (2009). Does social capital reduce crime? *Journal of Law and Economics*, 52(1):145–170.
- Calvó-Armengol, A. and Jackson, M. O. (2004). The effects of social networks on employment and inequality. *American Economic Review*, 94(3):426–454.
- Calvó-Armengol, A., Verdier, T., and Zenou, Y. (2007). Strong and weak ties in employment and crime. *Journal of Public Economics*, 91(1-2):203–233.
- Card, D. (2001). Immigrant inflows, native outflows, and the local labor market impacts of higher immigration. *Journal of Labor Economics*, 19(1):22–64.
- Caselli, F. and Coleman, W. J. (2013). On the theory of ethnic conflict. *Journal of the European Economic Association*, 11:161–192.
- Chamboredon, J.-C. and Lemaire, M. (1970). Proximité spatiale et distance sociale. les grands ensembles et leur peuplement. *Revue Française de Sociologie*, 11(1):3–33.
- Chiswick, B. R. (1980). The earnings of white and coloured male immigrants in Britain. *Economica*, 47(185):81–87.
- Coleman, J. S. (1988). Social capital in the creation of human capital. *American Journal of Sociology*, 94(Supplement: Organizations and Institutions: Sociological and Economic Approaches to the Analysis of Social Structure):S95–S120.
- Combes, P.-P., Decreuse, B., Schmutz, B., and Trannoy, A. (2012). The neighbor is king: Customer discrimination in the housing market. AMSE Working Papers 1224, Aix-Marseille School of Economics, Marseille, France.
- Combes, P.-P., Mayer, T., and Thisse, J.-F. (2008). *Economic Geography: The Integration of Regions and Nations*. Princeton University Press.
- Corcoran, M., Datcher, L., and Duncan, G. (1980). Information and influence networks in labor markets. *Five Thousand American Families: Patterns of Economic Progress*, 7:1–37.
- Costa, D. L. and Kahn, M. E. (2003). Civic engagement and community heterogeneity: An economist's perspective. *Perspectives on Politics*, 1(1):103–111.
- Crane, J. (1991). The epidemic theory of ghettos and neighborhood effects on dropping out and teenage childbearing. *American Journal of Sociology*, 96(5):1226–1259.
- Currarini, S., Jackson, M. O., and Pin, P. (2009). An economic model of friendship: Homophily, minorities, and segregation. *Econometrica*, 77(4):1003–1045.

- Cutler, D. M. and Glaeser, E. L. (1997). Are ghettos good or bad? *The Quarterly Journal of Economics*, 112(3):827–72.
- Cutler, D. M., Glaeser, E. L., and Vigdor, J. L. (2008). When are ghettos bad? lessons from immigrant segregation in the united states. *Journal of Urban Economics*, 63(3):759–774.
- Dahlberg, M., Edmark, K., and Lundqvist, H. (2012). Ethnic diversity and preferences for redistribution. *Journal of Political Economy*, 120(1):41 – 76.
- Dancygier, R. M. (2010). *Immigration and Conflict in Europe*. Cambridge University Press.
- Debrand, T. and Taffin, C. (2005). Les facteurs structurels et conjoncturels de la mobilité résidentielle depuis 20 ans. *Économie et Statistique*, 381(1):125–146.
- Dunning, T. and Harrison, L. (2010). Cross-cutting cleavages and ethnic voting: An experimental study of cousinage in mali. *American Political Science Review*, 104(1):21–39.
- Easterly, W. and Levine, R. (1997). Africa’s growth tragedy: Policies and ethnic divisions. *The Quarterly Journal of Economics*, 112(4):1203–1250.
- Ehrlich, I. (1973). Participation in illegitimate activities: A theoretical and empirical investigation. *Journal of Political Economy*, 81(3):521–65.
- Enste, D. H. and Schneider, F. (2000). Shadow economies: Size, causes, and consequences. *Journal of Economic Literature*, 38(1):77–114.
- Esteban, J. and Ray, D. (1994). On the measurement of polarization. *Econometrica*, 62(4):819–51.
- Esteban, J. and Ray, D. (2011). A model of ethnic conflict. *Journal of the European Economic Association*, 9(3):496–521.
- Evans, W. N., Oates, W. E., and Schwab, R. M. (1992). Measuring peer group effects: A study of teenage behavior. *Journal of Political Economy*, 100(5):966–91.
- Fearon, J. D. and Laitin, D. D. (1996). Explaining interethnic cooperation. *American Political Science Review*, 90(4):715–735.
- Fearon, J. D. and Laitin, D. D. (2000). Violence and the social construction of ethnic identity. *International Organization*, 54(04):845–877.
- Fearon, J. D. and Laitin, D. D. (2003). Ethnicity, insurgency, and civil war. *American Political Science Review*, null:75–90.
- Fougère, D., Kramarz, F., and Pouget, J. (2009). Youth unemployment and crime in france. *Journal of the European Economic Association*, 7(5):909–938.
- Galenianos, M. (2010). Hiring through referrals. 2010 Meeting Papers 516, Society for Economic Dynamics.

- Glaeser, E. L., Sacerdote, B., and Scheinkman, J. A. (1996). Crime and social interactions. *The Quarterly Journal of Economics*, 111(2):507–48.
- Glennerster, R., Miguel, E., and Rothenberg, A. D. (2013). Collective action in diverse sierra leone communities. *The Economic Journal*, 123(568):285–316.
- Gould, E. D., Weinberg, B. A., and Mustard, D. B. (2002). Crime rates and local labor market opportunities in the united states: 1979-1997. *The Review of Economics and Statistics*, 84(1):45–61.
- Goux, D. and Maurin, E. (2007). Close neighbours matter: Neighbourhood effects on early performance at school. *Economic Journal*, 117(523):1193–1215.
- Granovetter, M. S. (1973). The strength of weak ties. *American Journal of Sociology*, 78(6):1360–1380.
- Granovetter, M. S. (1995). *Getting a Job: a Study of Contacts and Careers*. Cambridge, MA. Harvard University Press, 2nd edition.
- Greif, A. (1993). Contract enforceability and economic institutions in early trade: the maghribi traders' coalition. *American Economic Review*, 83(3):525–48.
- Gërkhani, K. (2004). The informal sector in developed and less developed countries: A literature survey. *Public Choice*, 120(3/4):267–300.
- Hanson, G. H. (2005). Market potential, increasing returns and geographic concentration. *Journal of International Economics*, 67(1):1 – 24.
- Harris, C. D. (1954). The market as a factor in the localization of industry in the united states. *Annals of the Association of American Geographers*, 44(4):315–348.
- Holzer, H. J. (1988). Search method use by unemployed youth. *Journal of Labor Economics*, 6(1):1–20.
- Hong, L. and Page, S. E. (2001). Problem solving by heterogeneous agents. *Journal of Economic Theory*, 97(1):123–163.
- Huntington, S. P. (1993). The clash of civilizations? *Foreign Affairs*, 72(3):22–49.
- Ioannides, Y. M. and Datcher Loury, L. (2004). Job information networks, neighborhood effects, and inequality. *Journal of Economic Literature*, 42(4):1056–1093.
- Ioannides, Y. M. and Topa, G. (2010). Neighborhood effects: Accomplishments and looking beyond them. *Journal of Regional Science*, 50(1):343–362.
- Jacquot, A. (2007). L'occupation du parc hlm : éclairage à partir des enquêtes logement de l'insee. *Direction des Statistiques Démographiques et Sociales, Document de Travail N° F0708*.
- Katz, L. F., Kling, J. R., and Liebman, J. B. (2001). Moving to opportunity in boston: Early results of a randomized mobility experiment. *The Quarterly Journal of Economics*, 116(2):607–654.

- Kelly, M. (2000). Inequality and crime. *The Review of Economics and Statistics*, 82(4):530–539.
- Kleven, H. J., Knudsen, M. B., Kreiner, C. T., Pedersen, S., and Saez, E. (2011). Unwilling or unable to cheat? evidence from a tax audit experiment in denmark. *Econometrica*, 79(3):651–692.
- Kling, J. R., Liebman, J. B., and Katz, L. F. (2007). Experimental analysis of neighborhood effects. *Econometrica*, 75(1):83–119.
- Kling, J. R., Ludwig, J., and Katz, L. F. (2005). Neighborhood effects on crime for female and male youth: Evidence from a randomized housing voucher experiment. *The Quarterly Journal of Economics*, 120(1):87–130.
- Lazear, E. P. (1999a). Culture and language. *Journal of Political Economy*, 107(S6):S95–S126.
- Lazear, E. P. (1999b). Globalisation and the market for team-mates. *Economic Journal*, 109(454):C15–40.
- Lemieux, T., Fortin, B., and Frechette, P. (1994). The effect of taxes on labor supply in the underground economy. *American Economic Review*, 84(1):231–54.
- Light, I. (1979). Disadvantaged minorities in self-employment. *International Journal of Comparative Sociology*, 20(1-2):31–45.
- Light, I. (2005). *The Handbook of Economic Sociology*, chapter 28. The Ethnic Economy. Princeton University Press, 2nd edition.
- Light, I. and Gold, S. J. (2000). *Ethnic Economies*. Academic Press Inc.
- Light, I. and Rosenstein, C. (1995). *Race, Ethnicity, and Entrepreneurship in Urban America*. Walter deGruyter Inc.
- Lubell, H. (1991). *The Informal Sector in the 1980's and 1990's*. OECD Publishing.
- Ludwig, J., Duncan, G. J., Gennetian, L. A., Katz, L. F., Kessler, R. C., Kling, J. R., and Sanbonmatsu, L. (2013). Long-term neighborhood effects on low-income families: Evidence from moving to opportunity. *American Economic Review*, 103(3):226–31.
- Ludwig, J., Duncan, G. J., and Hirschfield, P. (2001). Urban poverty and juvenile crime: Evidence from a randomized housing-mobility experiment. *The Quarterly Journal of Economics*, 116(2):655–679.
- Luttmer, E. F. P. (2001). Group loyalty and the taste for redistribution. *Journal of Political Economy*, 109(3):500–528.
- Machin, S. and Meghir, C. (2004). Crime and economic incentives. *Journal of Human Resources*, 39(4).
- Manacorda, M., Manning, A., and Wadsworth, J. (2012). The impact of immigration on the structure of wages: Theory and evidence from britain. *Journal of the European Economic Association*, 10(1):120–151.

- Manning, A. and Roy, S. (2010). Culture clash or culture club? national identity in Britain. *Economic Journal*, 120(542):F72–F100.
- Martin-Houssart, G. and Tabard, N. (2002). Représentation socio-économique du territoire, typologie des quartiers et communes de France selon la profession et l'activité socio-économique de leurs habitants. Working Papers F0208, INSEE.
- McPherson, M., Smith-Lovin, L., and Cook, J. (2001). Birds of a feather: Homophily in social networks. *Annual Review of Sociology*, 27:415–444.
- Michelis, A. D., Estevão, M., and Wilson, B. A. (2013). Productivity or employment: Is it a choice? Working Paper 13/97, IMF.
- Miguel, E. (2004). Tribe or nation? nation building and public goods in Kenya versus Tanzania. *World Politics*, 56:328–362.
- Miguel, E. and Gugerty, M. K. (2005). Ethnic diversity, social sanctions, and public goods in Kenya. *Journal of Public Economics*, 89(11-12):2325–2368.
- Montalvo, J. G. and Reynal-Querol, M. (2005). Ethnic polarization, potential conflict, and civil wars. *American Economic Review*, 95(3):796–816.
- Montgomery, J. D. (1991). Social networks and labor-market outcomes: Toward an economic analysis. *American Economic Review*, 81(5):1407–18.
- Montgomery, J. D. (1994). Weak ties, employment, and inequality: An equilibrium analysis. *American Journal of Sociology*, 99(5):1212–1236.
- Munshi, K. (2003). Networks in the modern economy: Mexican migrants in the U.S. labor market. *The Quarterly Journal of Economics*, 118(2):549–599.
- Nordhaus, W. (2005). The sources of the productivity rebound and the manufacturing employment puzzle. Working Paper Series 11354, NBER.
- Observatoire (2011). Observatoire du logement et de l'habitat de Paris : "l'accès au logement social à Paris: Analyse de la demande de logement social et bilan des propositions et des attributions aux logements sociaux à Paris en 2010".
- Oreopoulos, P. (2003). The long-run consequences of living in a poor neighborhood. *The Quarterly Journal of Economics*, 118(4):1533–1575.
- Ottaviano, G. I. and Peri, G. (2006). The economic value of cultural diversity: evidence from US cities. *Journal of Economic Geography*, 6(1):9–44.
- Ottaviano, G. I. P. and Peri, G. (2012). Rethinking the effect of immigration on wages. *Journal of the European Economic Association*, 10(1):152–197.
- Page, S. (2007). *The Difference: How the Power of Diversity Creates Better Groups, Firms, Schools, and Societies*. Princeton University Press.
- Patacchini, E. and Zenou, Y. (2008). The strength of weak ties in crime. *European Economic Review*, 52(2):209–236.



- Patacchini, E. and Zenou, Y. (2012). Ethnic networks and employment outcomes. *Regional Science and Urban Economics*, 42(6):938–949.
- Pendakur, K. and Pendakur, R. (2011). Color by numbers: Minority earnings in canada 1995-2005. *Journal of International Migration and Integration*, 12(3):305–329.
- Portes, A. and Haller, W. (2005). *The Handbook of Economic Sociology*, chapter 18. The Informal Economy, pages 403–428. Neil J. Smelser and Richard Swedberg, Princeton University Press, 2nd edition.
- Posner, D. N. (2004). The political salience of cultural difference: Why chewas and tumbukas are allies in zambia and adversaries in malawi. *American Political Science Review*, pages 529–545.
- Prarolo, G., Bellini, E., Ottaviano, G. I., and Pinelli, D. (2009). Cultural diversity and economic performance: Evidence from european regions. Working Papers 2009.63, Fondazione Eni Enrico Mattei.
- Putnam, R. D. (2007). E pluribus unum: Diversity and community in the twenty-first century the 2006 johan skytte prize lecture. *Scandinavian Political Studies*, 30(2).
- Renooy, P. H. (1990). *The informal economy: Meaning, Measurement and Social Significance*. Netherlands Geographical Studies.
- Reynal-Querol, M. (2002). Ethnicity, political systems, and civil wars. *Journal of Conflict Resolution*, 46(1):29–54.
- Rouquette, R. and Lipietz, H. (1991). *Droit du logement social organismes d'HLM, SEM, ANAH, PACT, ARIM*. Collection L'actualité juridique. Éditions du Moniteur, Paris, première édition à jour des textes et de la jurisprudence au 15 février 1991 edition.
- Saiz, A. (2007). Immigration and housing rents in american cities. *Journal of Urban Economics*, 61(2):345–371.
- Selod, H. and Zenou, Y. (2006). City structure, job search and labour discrimination: Theory and policy implications. *Economic Journal*, 116(514):1057–1087.
- Senik, C. and Verdier, T. (2011). Segregation, entrepreneurship and work values: the case of france. *Journal of Population Economics*, 24(4):1207–1234.
- Shaw, C. and McKay, H. (1942). *Juvenile Delinquency and Urban Areas*. University of Chicago Press.
- Simon, P. (2003). Le logement social en france et la gestion des "populations à risques". *Hommes et migrations*, 1246:76–91.
- Spenkuch, J. L. (2010). Understanding the impact of immigration on crime. MPRA Paper 22864, University Library of Munich, Germany.
- Staiger, D. and Stock, J. H. (1997). Instrumental variables regression with weak instruments. *Econometrica*, 65(3):557–586.

- Tabard, N. (2002). *Données sociales, La société française*, chapter Inégalités et disparités entre les quartiers en 1999. INSEE.
- Tajfel, H., Billig, M., Bundy, R. P., and Flament, C. (1971). Social categorization and intergroup behavior. *European Journal of Social Psychology*, 1:149–179.
- Tiebout, C. M. (1956). A pure theory of local expenditures. *Journal of Political Economy*, 64(5):416–424.
- Tienda, M. and Raijman, R. (2000). Immigrants' income packaging and invisible labor force activity. *Social Science Quarterly*.
- Topa, G. (2001). Social interactions, local spillovers and unemployment. *Review of Economic Studies*, 68(2):261–95.
- Topa, G. (2011). *Handbook of Social Economics*, volume 1, chapter Labor Markets and Referrals. North Holland, 1st edition.
- Verdugo, G. (2009). Les immigrés et le marché du travail en France : quelques éléments descriptifs. *Immigration, Qualifications et Marché du Travail, Conseil d'analyse économique, complément A*, pages 115–130.
- Vigdor, J. L. (2002). Interpreting ethnic fragmentation effects. *Economics Letters*, 75(2):271–276.
- Wahba, J. and Zenou, Y. (2005). Density, social networks and job search methods: Theory and application to Egypt. *Journal of Development Economics*, 78:443–473.
- Weinberg, B. A., Reagan, P. B., and Yankow, J. J. (2004). Do neighborhoods affect hours worked? evidence from longitudinal data. *Journal of Labor Economics*, 22(4):891–924.
- Wellman, B. (1996). Are personal communities local? a dumptarian reconsideration. *Social Networks*, 18:347–354.
- Wilson, W. (1987). *The Truly Disadvantaged: The Inner City, the Underclass, and Public Policy*. Sociology, urban studies, black studies. University Press.
- Winters, P., de Janvry, A., and Sadoulet, E. (2001). Family and community networks in Mexico-U.S. migration. *Journal of Human Resources*, 36(1):159–184.
- work (2007). Undeclared work in the European Union. Report 67.3, European Commission. Special Eurobarometer 284.
- Zenou, Y. (2005). Crime, location and the housing market. CEPR Discussion Paper 5389, Center for Economic Policy Research.
- Zenou, Y. (2013). Spatial versus social mismatch. *Journal of Urban Economics*, 74(C):113–132.