INSTITUT POLYTECHNIQUE DE PARIS

NNT : 2022IPPAG003

Thèse de doctorat

# Anatomy of the Suburban Metropolis

*Labor Earnings, Large Scale Commuting, and Social Discontent*

Doctoral Thesis (PhD) from Institut Polytechnique de Paris
prepared at École nationale de la statistique et de l'administration économique
and École polytechnique

Doctoral School #626 – Institute Polytechnique de Paris (EDIPP)

Speciality : Economics

Thesis presented and defended at Palaiseau, the 5th of June, 2022, by

**Thomas Delemotte**

Composition of the Jury :

| | |
|---|---|
| Florence Goffette-Nagot<br>Directrice de recherche CNRS, GATE Lyon - St-Etienne | Examiner |
| Laurent Gobillon<br>Directeur de recherche CNRS, Paris School of Economics | Referee |
| Clément Bosquet<br>Professeur des Universités, Université Panthéon-Sorbonne | Referee |
| Yanos Zylberberg<br>Associate Professor, University of Bristol | Examiner |
| Clémence Tricaud<br>Assistant Professor, UCLA - Anderson School of Management | Examiner |
| Francis Kramarz<br>Professor, ENSAE | Director of thesis |
| Benoit Schmutz<br>Associate Professor, École polytechnique | Director of thesis |

*La forme d'une ville change plus vite,*
*on le sait, que le coeur d'un mortel.*

*The shape of a city changes more quickly,*
*as we know, than the heart of a mortal.*

in *La Forme d'une ville*, Julien Gracq

# Anatomy of the Suburban Metropolis

Thomas DELEMOTTE
July, 5, 2022

The location of jobs and populations has changed significantly over the last half century, with the rise of car use and the increase in commuting distances. It has led to emergence of the *suburban* metropolis. This model is now challenged by both increasing environmental concerns and social protests. This dissertation aims at contributing to this debate. The first chapter shows how labor market reforms have contributed to the reduction of inequality between rural and urban municipalities, while the second chapter describes the contribution of new transportation infrastructures in the making of the suburban metropolis. In France, the Yellow Vests (*Gilets jaunes*) movement has been the emblem of the call into question of this spatial organization with many blockades in metropolitan outskirts. Chapter three provides suggestive evidence that the protests are tied to mobility constraints, and chapter four describes the dynamics of the movement, which has also developed its own on digital territories.

Chapter 1: **Inequality and Earnings in France**: Labor Market Reforms, Gender and Places.

Chapter 2: **Railway, Highways, and the Suburban Metropolis**: Evidence from Paris.

Chapter 3: **The Origins of the Yellow Vests**.

Chapter 4: **Mobilization without Consolidation**: Social Media and the Yellow Vests.

# Anatomie de la métropôle périurbaine

Thomas DELEMOTTE
5 Juillet 2022

La localisation des emplois et des ménages a profondément changé depuis un demi-siècle, avec l'essor de la voiture et l'accroissement des distances domicile-travail. Ce qui a conduit à l'émergence de la métropole *périurbaine*. Ce modèle se confronte actuellement aux enjeux environnementaux et aux protestations sociales. Cette thèse vise à contribuer à ce débat. Le premier chapitre montre ainsi que les réformes du marché du travail ont contribué à réduire les inégalités entre communes rurales et urbaines, tandis que le deuxième chapitre décrit comment les infrastructures de transport ont contribué à la fabrique de cette métropole périurbaine. En France, les *Gilets jaunes* ont été l'emblème de la remise en cause de ce modèle, avec de nombreux point de blocages dans des zones périphériques. Le chapitre trois suggère que la protestation est bien liée à des enjeux de mobilité et le chapitre quatre détail la dynamique du mouvement, qui a aussi développé ses propres territoires numériques.

Chapitre 1 : **Inégalités et revenus en France**: réformes du marché du travail, genre et géographie.

Chapitre 2 : **Rails, autoroutes et métropole périurbaine**: l'exemple de Paris.

Chapitre 3 : **Les déterminants des Gilets jaunes**.

Chapitre 4 : **Mobilisation sans consolidation**: réseaux sociaux et Gilets jaunes.

# Contents

# Aknowledgment

I would like to begin this thesis by devoting few words to people without whom these four years of Ph.D. would not have been possible. Firstly, I am very thankful to my supervisors, Francis Kramarz and Benoît Schmutz for their guidance and support. Benoît's availability and kindness has been a precious element in the success of my PhD. I am grateful toward them for having shared with me their expertise and insight on labor, geography, and research more broadly. I really enjoy starting my PhD by the writing of the book on the employment and the territory with them two. Being supervised by Francis and Benoît was a great experience.

Then, I would like to thank Francis and Benoît again, with Pierre Boyer, Germain Gauthier, Vincent Rollet, Elio Nimier-David, and Corentin Trevien, who have co-authored with me the chapters of this thesis. Working in teams has been a worthwhile learning experience. I have gained a lot from rubbing shoulders with these talented researchers. Pierre and Benoît were strong pillars of the Gilet jaunes team, as well as Germain and Vincent, who had plenty of innovative idea. I enjoyed working on the interaction of urban and political economics with them. Francis and Elio shared with me their thorough expertise of the French matched Employer-Employees data. It is finally a true pleasure to work with Corentin, the "urban planner", with whom I always have extensive and interesting discussions on the making of the city.

I am also particularly thankful to Isabelle Méjean, who provided me substantial support and guidance throughout my PhD. Isabelle was part of my monitoring committee and she always took the time to provide insight that benefited me a lot, both on the scientific and human aspects. Rolland Rathelot, Patricia Crifo, and Grégory Corcos have also been helpful members of my monitoring committee. I thank them all for their very useful comments. Furthermore, I am indebted to faculty members at CREST, who contribute in many manners to improving the chapters of this thesis, among others Béatrice Cherrier, Etienne Ollion, Julien Combes, Pauline Rossi, Thibaud Vergé, Xavier D'Haultfoeuille or Bertrand Garbinti. I also acknowledge great support of the CASD team. I am thankful to Gabriel Ulyssea, who devoted time to discuss my research during my stay at Oxford, as well as Luigi Pistaferri, who organized the GRID workshop at Stanford, from which the first chapter of this thesis is an outcome.

I would like to thank my PhD jury, who provided me challenging comments during my pre-defense, which, I hope, have helped improving substantially this final manuscript. In particular, I am greatefully to Laurent Gobillon and Clément Bosquet, who accepted to be referee of my thesis and thus devote a bunched of time to review my research. I also thank Florence Gofette-Nagot for her thorough reading of my work and her comments, as well as Yanos Zylberberg for his insights that have allowed me to take a broader per se spective on my research. Finally, I am very grateful to Clémence Tricaud, who took time to help me structure my writting and was, during my PhD, an outstanding colleague.

Interactions with fellow students at CREST, but also during conferences or workshops, have also played a key role in my research. I am particularly grateful to Louis-Daniel Pape, my office partner. We have spent quite a good time gazing at the landscape of Saclay Plateau. Louis is a valuable friend, always ready to exchange about econometrics, baseball or the meaning of the zeros. Léa Bou Sleiman, my supervisory sister, has also been important in my well being at the lab. Our discussions on urban economics helped me improve my research. I thank Guidogiorgio Brodato for his unalterable joy of living, even after long RER journey. This thesis builds upon many discussions and heated arguments held with Ninon Moreau-Kastler, Raphael Lafrogne-Joussier, Antoine Bertheau, Alice Lapeyre, Félix Pasquier, Martin Mugnier, Alicia Bassière, Claire Leroy, Antoine Ferey, Bérengère Patault, Alexis Larousse, Pierre-Edouard Colignon, Germain and Vincent, who I would like to thank to help me improve this dissertation. I am also infinitely rewardful to all my PhD fellows, with many of whom we have overcome the disruption of the Covid period, Reda Aboutajdine, Anasuya Raj, Emilie Sartre, Felix Shleef, Pauline Carry, Heloise Cholery, Yannick Guyonvarch, Jérémy Hervelin, Gwen-Jiro Clochard, Yuanzhe Tang, Denys Médée, Morgane Guignard, Rémi Avignon, Elia Pérennes, Guillaume Bied, Elie Vidal-Naquet, Jeanne Ganault, Etienne Guigue, Morgane Hoffman, Margarita Kinerva, Inès Moutachaker, Bertille Picard, Elie Gerschel, Wael Bousselmi, Thomas van Casteren, Gautier Lenfant, and Elio.

Of course, I thank the administrative staff, especially Lyza Racon, Weronika Leduc, Eliane Madelaine, Fanda Traoré, Murielle Jules, Edith Verger and Teddy Arrif, as well as Arnak Dalalyan. They were always very helpful and diligent. They were always welcoming and have provided me with material and financial support to organize events, such as the *PhD Day* or the *PhD Breakfasts*. It was a true pleasure to work with them. I am also particularly grateful to Morgane Cure, Jérôme Trinh, Lucas Girard and Fabien Perez, as well as Jean-Baptiste Michau and Olivier Gosner for the organization, help and advice for the teachings at ENSAE and Ecole Polytechnique.

I have a spacial thought for Julien Sauvagnat, who gave me the willingness to pursue in research during my internship at Bocconi. I am also thankful to the reflection circle we build with Paris students to challenge teaching in economics and especially Quentin Louis, who has become a friend. I am also grateful to my colleagues at the French Treasury, who have hosted me warmly and have become true comrades, with a particular thought to Clémence Lenoir, who hired me.

To end, I am grateful to my family, my friends - including TeamBacki and La Délicieuse, and my neighbors for their understanding, support and affection throughout this time period. I am especially thankful to my beloved Ondine Berland, who always provides me wise advice and outstanding assistance. It was a joy to share my life with her during this challenging period.

Last but not least, I am indebted to all the researchers on whose work this thesis is grounded and I am grateful to the reader this thesis might inspire.

Au delà de la dimension scientifique et professionnelle que revet le doctorat, je suis reconnaissant envers ma famille et mes proches pour leur accompagnement et encouragement au cours des ces années de thèse.

Tout particulièrement, je souhaite remercier mes parents, Nathalie et Bruno, pour leur éducation et leur soutien, qui ont porté les germes de cette aventure. Ma soeur Lucie, pour son sens de la fête et des sentiments, mon frère, Jules, pour sa sagacité et ses intuitions. Vous avoir dans ma vie, me donne un socle précieux. Mes grand-parents, Jeanine et Claude, Françoise et Bernard, sont des piliers affectifs et moraux, qui m'ont guidé dans mes choix.

Ondine, qui partage ma vie, les moments de joies comme de détresse, qui jalonnent la thèse. Sans toi, cette aventure ne serait pas arrivée à son terme.

Je suis aussi reconnaissant envers mes amis de Reims pour leur affection. La TeamBacki, vous avez nourri mes questionnements et avait jalonné leur investigation, David, Clément, Gauthier, Amélie, Diane, Héléne, Arthur, Machu, Floriane, mais aussi Fayçal, Rémi et tant d'autres.

Mes pensées vont aussi vers mes amis, rencontrés durant mes études et ensuite. Les membres de la Délicieuse, sans qui ma vie serait moins *chill* et moins *kiffante*, dont mes colocs Pierre-Yves, Yahya, Sacha et Imad, mais aussi Sonia, Tam, Nathan, Louis, Sophie, Guillaume, Elodie et tous les autres délicieux.ses. Enfin, Adel, David, Raphael et Sylvain pour ne nommer que vous, qui avez été les cairns de mon parcours et m'avez donné un souffle précieux.

Que tout ceux que j'ai oublié de nommer me pardonnent. Je pense évidement à vous aussi. Tous les soutiens sont précieux pour avancer dans la thèse et la réussite individuelle est toujours tributaire de son environnement et du collectif, comme, je l'espère, vous aurez loisir de vous en rendre compte à la lecture de cet ouvrage.

# Foreword

The location of jobs and populations has changed significantly over the last half century, with the rise of car use and the increase in home-to-work distances. This development has led to emergence of the *suburban* metropolis, symbol of a new urban equilibrium. In France, the Yellow Vests (*Gilets jaunes*) movement has been the emblem of the call into question of this equilibrium. The importance of the mobilization has shed light on the reliance of individuals on car in their daily life and the place of the suburban metropolis in the reshaping of the organization of the production. A substantial share of the territory was blocked from the first Act of the movement's mobilization, underlying the need to better understand the spatial reshuffling of opportunities in recent decades.

This dissertation aims to shed light on the role of territories, both physical and digital, in the formation of individual choices. Its chapters contrast the effect of public policies on individuals depending on their place of living. They also provide insights for their efficient implementation accounting for the local contexts. Finally, through the study of social discontent, this thesis provides a better understanding on the dynamics of protests against public action.

The first two chapters focus on the evaluation of the public policies and their heterogeneous impact among places. The last two chapters study the Yellow Vests movement, which is rooted in the evolution of the metropolitan landscape. The first chapter examines the effect of labor market reforms, implemented in the turn of the 2000s, on individuals' labor earnings. Particularly it focuses on the disparities between men and women, and between places according to their distance the urban core. Its shows that these social policies have contributed to the reduction of inequality between rural and urban municipalities. The second chapter describes how transportation infrastructure programs contributed to the making of the suburban metropolis, through a case study of Paris urban planning in the second half of the XXst century. It shows that rail and road improvement were actually complement to support the geographic extension of the metropolis. These two chapters hence provide a background of the development of the French society. In this context the surge of the Yellow Vests movement was a surprise. The third chapter described this mobilization. It provides evidence that the protest is tied with individuals' mobility constraint and thus linked with the development of the suburban metropolis. The fourth, and last chapter, study the dynamics of the movement. It shed light on the interplay between the development of online territories and offline actions, which explains how the movement was able to coalesce protesters from remote locations on the first day of road blockades.

The suburban metropolis has contributed to the reshuffling of opportunities, notably in terms of jobs and wages. The dramatic extension of the catchment area of the urban centers has favored access to individual houses for workers. It has also been accompanied by a relocation of jobs at the edge of the built-up area. Therefor, we observe that the Paris urban planning, implementing both the RER and national highways, has favored the integration of remote workers into the metropolitan labor market, while network nodes benefited from the infrastructure improvement to attract more jobs. Meanwhile, social reforms of the labor market have led to a decline in inequality between rural and urban locations, with the earnings from labor of the low-paid workers catching-up. However, this long-term dynamics have been attenuated, if not reversed, since the beginning of the 2010s. Indeed, labor earnings inequality have tended to increase again in the wage of the 2008 financial crisis and the use of individual cars has no longer been favored by public authorities, in a context of increasing fuel prices and ecological concerns.

These phenomenon, inequality and car dependence, became in November 2018 the principal catchwords of the Yellow Vests. Indeed, we observe a correlation between the level of mobilization and the characteristics of the suburban metropolis. The territories most mobilized, either on Facebook or on the round-about, are also the ones with greater commuting distances and a higher density of roads slowed to 80km/h. As an emblem, Priscilla Ludosky, the initiator of the online petition that fueled the movement, is a motorist living a suburban municipalities in Paris region: Savigny-le-Temple, Seine-et-Marne. Her petition was about the cost of relying on car to commute in a context of increasing gasoline prices. The protest then extend to the rest of France, both online and in terms of roads blocked.

The remainder of this introduction outlines the scientific motivations for the study of the suburban metropolis and the case of the Gilets jaunes that it has witnessed. While the last section provides an overview of each chapter and details the contribution of this dissertation to the literature.

## Structure of the territory

The making of the city is a dynamic process, which often escaped planning, before becoming a topic of interest for policy maker. The concentration of jobs in metropolises and the increase of their catchment area has thus became a topical concern for public policies and an important topic of research for economics. Households are both consumers of space to live and seeking for jobs to earn a wage. Their competition for housing and jobs, within a limited space, constraint their choice of location and determines the level of housing prices and wages.[1]. Moreover, residential locations and places of work are not homogeneous between each others. Each location can be characterized by intrinsic features, called amenities[2], and by the spillovers it hosts, resulting in agglomeration

1 Pierre-Philippe Combes, Duranton Gilles Duranton et Laurent Gobillon: *Spatial wage disparities: Sorting matters!*, Journal of Urban Economics, 2008, & *The costs of agglomeration: House and land prices in French cities*, The Review of Economic Studies, 2019.

2 David Albouy. *What Are Cities Worth? Land Rents, Local Productivity, and the Total Value of Amenities.* The Review of Economics and Statistics. 2016.

economies[3], which makes them more or less attractive. The different components of the households' location choice have led them to reside away from their place of work to maximize the comparative advantage of places, while commuting costs were declining.[4]. This dynamics contributed to the emergence of urban sprawl and later to the relocation of jobs in the periphery. Urban planners have therefore sought to reduce travel time in order to bring workers closer to their jobs. Yet the opposite effect has occurred, with an increase in commuting distances that exceed the gain associated from time saving. In the present context of ecological transition and increasing gasoline prices, the current model, based on car commuting, is no longer sustainable. However, we lack studies on the modal shift and on how the ability of alternative infrastructure to substitute for car.

The canonical urban model relies on a monocentric approach of cities,[5] where jobs are located in the center and residents in its surrounding area. Cities, or Metrpolises, thereby define local labor market, which mesh the territory. Their size depends on the maximum travel cost compatible with commuting to the central business district. From an empirical point of view, it translates into a definition of the city based on observed commuting flows. In France, following international standard, two main nomenclatures have been developed by the Statistical Institute (INSEE): the *Urban Area* (Aire urbaine) and the *Employment Zone* (Zone d'emploi).

An urban area is defined as a group of municipalities, in one piece and without an enclave, consisting of a central Urban unit with more than 10,000 jobs, and rural municipalities or urban units in which at least 40% of the resident population with a job works in the central Urban unit or in municipalities attracted by it. Nearly eight hundred urban areas are defined accordingly and gather 85% of the French total population. The employment zones are also groupings of municipalities, with at least 5,000 workers, defined so that the number of workers living and working in the area is as high as possible. The aim of this definition is to cover the whole territory, without leaving aside a single municipality. It ends up with 321 employment zones that form a complete partition of France.

## Spatial fragmentation and revolts

The geography of the labor market is thus composed of interconnected metropolises that split the territory. However, the labor market is polarized by the divergence between metropolises, winners of the globalization, and declining territories, as described for the United-State by Enrico Moretti.[6] and for the French case[7] in our book with Francis Kramarz and Benoît Schmutz.[8] This fragmentation was first documented in the largest metropolises, where the gaps between globalized top-earners, who can afford substantial urban cost, and disadvantaged people, who are stuck on the margins of the urban opportunities, are the greatest.[9] This leads to a segregation of populations, with a

[3] Gilles Duranton et Diego Puga. *Micro-foundations of urban agglomeration economies*, in the Handbook of regional and urban economics. 2004

[4] Stephan Heblich, Stephen Redding et Daniel Sturm. *The making of the modern metropolis: evidence from London.* The Quarterly Journal of Economics, 2020.

[5] William Alonso. *Location and Land Use: Toward a General Theory of Land Rent*. 1964. Richard Muth. *Cities and Housing: The Spatial Pattern of Urban Residential Land Use.* Third Series: Studies in Business and Society, 1969. Edwin Mills. *An aggregative model of resource allocation in a metropolitan area.* The American Economic Review, 1967.

[6] Enrico Moretti. *The new geography of jobs.* Mariner Books, 2012.

[7] Donald Davis, Eric Mengus et Tomasz Michalski. *Labor market polarization and the great divergence: Theory and evidence*. National Bureau of Economic Research. 2020.

[8] Thomas Delemotte, Francis Kramarz et Benoît Schmutz. L'emploi et le territoire. Presses de Sciences Po, coll. "Sécurisation des Parcours professionnels", 2021.

[9] Edward W. Soja. *Postmetropolis: critical studies of cities and regions*, Wiley-Blackwell, 2000.

concentration of affluence in the city centers, close the most deprived neighborhood,[10] followed by a decrease in income level with distance to the city center.[11] This fragmentation of the territory can be stated at all geographic scales, between neighborhoods as well as between countries. If the need to address the juxtaposition of territories with different, if not contradictory, logics is at the core of the renewal of geographic analysis that has been underway for three decades,[12] it is clear that economics as a whole is lagging beyond on these issues.[13] Through the transfers it organizes, the welfare state has, nonetheless, greatly mitigated the effect of increased global competition, at the cost of high indebtedness.[14] Current budget reductions alter this implicit redistribution between places, which leads to tension among territories, as revealed with the emergence of the Yellow Vests.[15]

The Yellow Vests movement is part of the long history of the social movements in France.[16] In some ways, this mobilization reproduces well-known historical patterns. Indeed, fiscal reforms are often the roots of social protests.[17] Especially in France, tax revolts have marked the history of taxation, from medieval *jacqueries* to the rise of the gasoline tax at the pump.[18] At the same time, the Yellow Vests differ from traditional protests in several ways. First, the movement is characterized by numerous decentralized gathering points, often around traffic circles, symbols of the French "motorists". From the first Saturday, there were 788 blockades. Second, the demonstration seems to have been largely organized without the intervention of traditional intermediate bodies such as parties and unions, which were late in joining the protest. Finally, social media seems to have played a decisive role in the organization and the popularization of the movement. In mid-December, there were 1,548 Facebook groups with more than 100 members associated with the Yellow Vests movement. While massive protests have already been initiated and catalyzed by social media around the world, a mobilization of this magnitude was unprecedented in France. Furthermore, the role of social media in the dynamics of contemporary social movements is still poorly understood. However their importance seems to be growing in large-scale protest movements. In the context of the Arab Springs, for instance, Twitter activity was a strong predictor of effective protests[19]. Similar conclusions have been drawn from the Chinese,[20] American,[21] or Russian[22] context.

---

10 Edward Glaeser, Matthew Kahn et Jordan Rappaport. *Why do the poor live in cities? The role of public transportation*. Journal of Urban Economics.

11 Stephen LeRoy et Jon Sonstelie. *Paradise lost and regained: Transportation innovation, income, and residential location*. Journal of Urban Economics, 1983. Laurent Gobillon, Anne Lambert et Sandra Pellet. *The suburbanization of poverty: Home-ownership policies and spatial inequalities in France*. Document de travail INED 250, 2019.

12 Jacques Lévy et Michel Lussault. *Dictionnaire de la géographie et de l'espace des sociétés*, Belin, 2003.

13 David Autor. *Work of the past, work of the future*. AEA Papers and Proceedings, 109, pp. 1–32, 2019.

14 Laurent Davezies. *La crise qui vient. La nouvelle fracture territoriale*. Seuil, coll. « La République des Idées », 2012.

15 Jacques Lévy, Jean-Nicolas Faucille et Ana Póvoas. *Théorie de la justice spatiale. Géographies du juste et de l'injuste*. Odile Jacob, 2019.

16 Eric Neveu. *Sociologie des mouvements sociaux*. La Découverte. 1996.

17 Ponticelli, J. Voth, H.-J. *Austerity and anarchy : Budget cuts and social unrest in Europe, 1919-2008*. Journal of Comparative Economics, 2020.

18 Nicolas Delalande: *Les batailles de l'impôt. Consentement et résistances de 1789 à nos jours*, Média Diffusion, 2011, et *Le retour des révoltes fiscales?*, Pouvoirs, 2014.

19 Zachari Steinert-Threlkeld, Delia Mocanu, Alessandro Vespignani et James Fowler. *Online social networks and offline protest*. EPJ Data Science, 2015. Daron Acemoglu, Tarek Hassan et Ahmed Tahoun. *The power of the street: Evidence from Egypt's Arab Spring* The Review of Financial Studies, 2018.

20 Bei Qin, David Strömberg, et Yanhui Wu. *Why does china allow freer social media ? protests versus surveillance and propaganda*. Journal of Economic Perspectives, 2017.

21 Neal Caren et Sarah Gaby. *Occupy online : Facebook and the spread of occupy wall street*. Available at SSRN, 2011. Ion Vasi et Chan Suh. *Protest in the internet age : Public attention, social media, and the spread of 'occupy'protests in the united states*. In Politics and Protest workshop. 2013. Marcos Bastos, Dan Mercea et Arthur Charpentier. *Tents, tweets, and events : The interplay between ongoing protests and social media*. Journal of Communication, 2015.

22 Ruben Enikolopov, Alexey Makarin et Maria Petrova. *Social media and protest participation : Evidence from Russia*. Econometrica, 2020.

## This Dissertation

This dissertation builds upon these theoretical and empirical insights and provide investigations of the incidence of public policies on places and people's mobilization. Using a wide variety of data sources, from administrative to digital, my work aim at providing new evidence on how individuals and public policies shaped territories. First, we will see that the labor market reforms put in place at the national level actually translated into contrasted incidence on spatial inequality. We will then see how transportation infrastructure shaped the access of individuals to labor market opportunities. Finally, we discuss how the characteristics of the suburban metropolis can explain the rise of the Yellow Vests movement and how the movement then build its own territories of action.

## Chapter 1: Inequality and Labor Earnings in France

Since the beginning of the 21st century there has been a renewed interest among economists to study and understand the structure and evolution of income inequality. Focusing solely on cross-sectional inequality leaves out many aspects of the distribution of wealth and income: What are the forces that shape these trends? How do public policies affect earnings inequality? What is the local declination of national trends? These questions are crucial to evaluate policies that aim at addressing inequality in an efficient and inclusive way, such as labor market reforms. Since labor earnings constitute a substantial share of total income, it is important to know the impact of these reforms on individuals' total labor income and inequality among occupied workers. In addition, the recent social protests on the urban fringe have highlighted (such as the Yellow Vests in France) the importance of better accounting for the heterogeneous impacts of these policies according to where individuals live.

This chapter focuses on two major labor market reforms that have been implemented in France during the turn of the 2000s: the reduction of the workweek from 39 to 35 hours worked and the decrease of employer-paid payroll taxes around the minimum wage. To shed light on the incidence these reforms have had on individuals' labor earnings and inequality. We first provide evidence that hours worked have declined following the reform, while hourly wage increased to sustain workers' total earnings. However, the increase in hourly wage has been stronger for women over the period, whereas low-paid women do not reduced their number of hours worked, since they were not working full-time. We then turn to the study of the distribution of individuals' total labor earnings within a year and confirms the strong catching-up of female workers at the bottom of the earnings distribution associated at the national level. In addition, since the share of low-paid female workers is more important in low-density areas, we show that the catching-up benefited mostly to residents of rural areas. This trends induced a substantial decrease in the gender pay gap over the period, as well as a reduction of inequality between rural and urban places in term of median earning.

However, if this reduction in inequality holds for the core of the earnings distribution, we observe a divergence at its very top, with a dramatic increase of earnings for top percentiles. In addition, these top-earners are mostly located in Paris, resulting in a substantial disparity in mean earnings between Paris and the rest of France. This finding is in line with increasing gains from agglomeration economies with earnings level. It also relates to the recent social protests, as it calls for new public policies so to reduce earnings inequalities between the top and the rest of the earnings distribution and better account for the spatial differences.

**Contribution**

This chapter contributes to several strands of the literature, by studying income inequality, with a focus on total labor earning, gender differences and spatial disparities.

First, we add to the debate on income inequality by providing insights on how market reforms can affect inequality in the long run. In contrast to the studies on total income[23], we show that labor earnings inequality in France have rather declined in recent decades, following the improvement of working standards, which differ to results found in other European countries.[24] Nevertheless, we find a dramatic increase of the earnings of top-earners as in other countries. This result is in line with a "concentration of affluence" in the largest metropolises, similarly to what is observed in the US.[25]

In a second part, we contribute to the urban literature that looks at segregation and at the urban wage premium. In particular we extend previous approach by including all places, beyond the only urban unit. It allows us to observe all workers and to take into account the suburbanization of the metropolis, with most of rural residents commuting long distance to work in the urban unit (which includes inner suburbs). In contrast, previous studies have mainly emphasized both the gain from agglomeration economies at workplace, comparing metropolises between them or urban segregation, comparing neighborhood within the same urban unit. We add to these works by focusing on the heterogeneity in earnings changes over time according to individual access to labor market opportunities. We proxy this access by the distance of workers' place of residence from the metropolitan urban core, where the largest share of jobs locate.

Finally, we mitigate recent results on city center gentrification, at least for France, by showing first that earnings of (urban) suburbs are substantially higher than low-density residents and, second, that the urban core of metropolises still host a substantial share of low-paid workers. The main mechanism that seem able to explain these trends is the strong reduction of inequality between rural and urban residents fostered by major labor market reforms.

## Chapter 2: Railway, Highway, and the Suburban Metropolis

The making of the modern metropolis has been shaped by the separation of residential location (in the suburbs) and workplaces (in downtown areas). The radial design of most of transportation infrastructures has lead to many study of the core-periphery structure of the metropolis. However, the emergence of individual car commute has fostered the rise of the suburban metropolis, which attracts workers with residence outside the urban fringe, and increase dramatically suburb-to-suburb commuting, as we detail in this chapter. Therefore it is important to better understand how rail and road infrastructures interact in shaping large-scale commuting. Especially, are they complements or substitutes, and do they foster the extension of the metropolis in the same way?

23 Thomas Piketty. *Income inequality in France, 1901–1998*. Journal of Political Economy, 2003. Anthony Atkinson, Thomas Piketty et Emmanuel Saez. *Top incomes in the long run of history*. Journal of Economic Literature. 2011.

24 Brian Bell, Nicholas Bloom et Jack Blundell. *Income dynamics in the United Kingdom 1975-2020*. GRID Working Paper 2022. Moritz Drechsel-Grau, Andreas Peichl, Johannes F Schmieder, Kai D Schmid, Hannes Walz et Stefanie Wolter, *Inequality and income dynamics in Germany*. GRID Working Paper. 2022. Kevin McKinney, John Abowd et Hubert Janicki. *US long-term earnings outcomes by sex, race, ethnicity, and place of birth* GRID Working Paper, 2021.

25 Cécile Gaubert, Patrick Kline, Damián Vergara et Danny Yagan. *Trends in us spatial inequality: Concentrating affluence and a democratization of poverty*. AEA Papers and Proceedings. 2021

We focuses in this chapter on two major infrastructure programs in Paris region affecting commuting in the second half of the twentieth century: the development of mass-transit and the building of highways. It therefore provides a topical case study to disentangle the contribution of both rail and road in shaping commuting in the suburban metropolis. Using individual-level data on workers from 1968 to 2014, we link commuting flows with travel time over this period for the entire Paris region. This data allows us to describe the substantial evolution of commuting pattern over the period. We then micro-found our empirical framework to study the incidence of the reduction in travel time on commuting flows, accounting for potential confounding factors such as place-based policies or targeted investments. We find that the two modes are complements in the making of the suburban metropolis. One the one hand, rail enables the sustaining of significant commuting flows along radial lines. On the other hand, roads foster suburb-to-suburb connections with few detours. At the individual level, we show that the two modes are nevertheless substitutes: within residence-workplace pairs, longer travel time with rail cause additional commuting flows by car, while potential traffic on roads fosters the use of the public transport network to commute.

These findings suggest that reducing the use of cars for commuting would impact more suburban residents and jobs than the ones located in central areas. In addition, existing public transport infrastructures lack of flexibility to represent a substitute for all commuters. An opportunity to address these issues could reside in favoring multi-modal commuting, that have been underused -and studied- so far.

**Contribution** This chapter contributes to the quantification of the causal incidence of infrastructure improvement on effective commuting flows. Previous studies[26] have mostly emphasis the role of infrastructure in causing urban sprawl, relying on a core-periphery approach, rather than addressing effective flows.

We build our empirical strategy on quantitative spatial models.[27] It allows to derive a sufficient statistics that characterized the urban equilibrium and allow to control for confounding factors such as place-based policies or targeted investment. Our empirical approach amounts, ultimately, to a difference-in-difference with a fuzzy design, where our treatment is the infrastructure improvement that reduces travel time.

Finally the paper contributes to the modal choice literature initiated by McFadden. We show that in addition to different speed, different modes display different accessibility. In particular we show that rail and road can be complement in the making of the suburban metropolis, while, at the individual level, they are substituted one with each others.

## Chapter 3: The Origins of the *Gilets jaunes* movement

In this chapter, we approach the Yellow Vests movement through the prism of geography. We aim to answer the following question: What are the characteristics of the geographical areas with a high level of mobilization at the beginning of the movement? In this, our study differs from the two works mentioned above in several ways. First, we are interested in the mobilization itself (and

26 Nathaniel Baum-Snow. *Did highways cause suburbanization?*. The Quarterly Journal of Economics, 2007. Thierry Mayer et Corentin Trevien. *The impact of urban public transportation evidence from the Paris region.* Journal of Urban Economics, 2017.

27 Stephen Redding et Esteban Rossi-Hansberg. *Quantitative spatial economics*. Annual Review of Economics. 2017.

not its support in the wider population), on Facebook and in the physical territory (blockades). We document the Facebook groups associated with the Yellow Vests that we were able to locate. We thus highlight a strong correlation between online and offline activity, before mapping the mobilization in these two dimensions. Second, we built a geolocated database that combines administrative sources (jobs, income, voting history) with offline (blockades of roundabouts) and online(Facebook groups) mobilization indicators. Our econometric study highlights the strong link between mobilization and variables related to mobility: commuting distance and the reduction of the speed limit on secondary roads from 90 km/h to 80 km/h.

By focusing on the beginning of the Yellow Vests mobilization (the blockades of November 17 and Facebook activity from October to mid-December), this article examines the factors that triggered it. To this end, we propose an original approach, adapted to the spatial dimension of the movement. Our study shows that the reduction in the speed limit on secondary roads to 80 km/h played a role in the emergence of the mobilization, which could explain why its motivation was poorly understood by a part of the population and by political parties at the start. The results presented in this article are interesting in that they suggest a link between mobilization and mobility.

The Yellow Vests movement calls into question our relationship to physical and digital territories. These synergies between new phenomena of online mobilization and more traditional protests involving the occupation of public space constitute a new framework for interpreting social movements in France. Unlike other recent mobilizations, such as "*La Manif pour tous*" or "*Nuit debout*", the spatial dispersion of the action points of the Yellow Vests movement was remarkable, which gave it an unprecedented nationwide coverage from the first day of mobilization.

**Contribution** The media coverage received by the Yellow Vests was extensive. On television, radio, and in newspapers, numerous interpretations were put forward to explain the emergence of the movement. In the academic context, few studies have provided convincing evidence of the movement origins and subsequent dynamics. Indeed robust data to study the movement were missing. Our contribution is to gather an original collection of data from different sources to characterize the movement and provide a map of both the on-line (Facebook groups) and off-line (blockades) actions. We provide evidence on the correlation between these physical blockades and online activity. Finally, the spatial data on the movement allows us to provide evidence of socio-demographic variables that best fit the spatial heterogeneity of the movement. Our results support the role of social media in the organization of protest, as well as the role of the spatial context in prompting revolts.

## Chapter 4: Mobilization without Consolidation

The previous chapter provides insight on the link between internet mobilization and effective blockades of the roads by employment zones and department, with the case of the Yellow Vests movement. In this chapter, we extend this work by enriching our data with original sources, which includes the geolocated signatories of the petition that initiate the movement and the Facebook pages, that becomes online agora of the movement. This allows us to explore the interactions between the online and offline components of the movement at the municipality level, and to show the determinants of the movement growth and decline. We empirically characterized three phases of the movements: emergence, coalescence, and decline that support previous theoretical results[28].

First, the wide sharing of the Change.org petition against the rise in gasoline prices provided a social ferment for the Yellow Vests movement to *emerge*. As the petition became viral, a large social media infrastructure was put in place to organize road blockades. We show with at the municipality level that measures of local online activities of the Yellow Vests predict the occurrence of a blockade as precisely as a wide set of geographic, demographic, economic, and political controls.

Our analysis further explores the reverse direction of this link, and shows that offline demonstrations spur further online activity. We provide causal evidence of the impact of the first wave of blockades on subsequent online mobilization. This finding substantiates qualitative evidence that many protesters sought to *coalesce* by continuing their exchange with fellow protesters whom they had met on the blockades. In theory, this should have paved the way to a subsequent consolidation of the movement. However, the number of protesters quickly subsided after few weeks.

To understand the evolution of the movement during this *decline* phase, we extract a corpus of sentences posted by Yellow Vests on Facebook pages. This allows us to build a quantitative case study based on different Natural Language Processing (NLP) methods. We find that the share of messages related to organizational concerns and practical demands decreased over time, contrary to messages with more antagonistic content such as insults or mentions of violence. The study of (de-identified) individual users over time allows to distinguish between an extensive margin (changes in the composition of the population of discussants) and an intensive margin (an individual-level increased tendency to post antagonistic messages) in the radicalization process. We show that both margins contributed to the increase of antagonist posts over time, which might have fostered the way for the decline of the movement.

**Contribution** The importance of social media in the development of large protest movements has been investigated in theoretical and empirical studies. We provide new evidence to this body of research on the dynamics of social movements. Our methodological contribution with respect to this literature is twofold: first, we complement previous studies which showed that social media contributes to the emergence of offline protests by showing that offline demonstrations caused further online action. Together with the previous literature, our study shows how social media can play as a force keeping protest movements alive. We find that political mobilization in our dynamic setting has especially high returns from early efforts: initial blockades and early online mobilization created a persistent political engagement on Facebook.

Second, we are able to analyze the radicalization which accompanies the decline phase of many social movements. Using a panel of online protesters, we are able to disentangle which part of the radicalization comes from the extensive margin or from the intensive margin.

Turning to the literature studying the Yellow Vests specifically, this paper is the first one to bring together comprehensive data on signatories of the most important petition related to the movement, the dynamics on the movement on Facebook pages and groups, and administrative data at the local level.

[28] Herbert Blumer. *Principles of sociology*. Barnes Noble. 1969. Charles Tilly. *From mobilization to revolution.* CRSO Working Paper. 1977.

# Préambule

La localisation des emplois et des travailleurs a fortement évolué depuis un-demi siècle, avec l'essort de la voiture individuelle et l'accroissement des distances domicile-travail. Cette dynamique a fait émerger la métropole *périurbaine*. Symbole d'un nouvel équilibre urbain. Les *Gilets jaunes*, en France, ont été l'emblème de la remise en question de cet équilibre. L'ampleur de ce mouvement a mis l'accent sur le rôle de la voiture dans la vie des individus et la place de la métropole périurbaine dans l'organisation de la production. Une partie importante du territoire a été bloquée dès le premier acte de mobilisation du mouvement, soulignant la nécessité de mieux comprendre quels sont les fondements de la redistribution spatiale des opportunités et des richesses produites durant ces dernières décennies.

Cette thèse se propose d'éclairer le rôle des territoires, physiques et numériques, dans la formation des choix individuels. Les études présentées mettent en évidence l'effet contrasté des politiques publiques sur les individus en fonction de leur lieu de vie. Elles fournissent aussi des pistes de réflexion pour améliorer la mise en place de ces dispositifs en fonction des contextes locaux. Enfin, à travers l'étude de l'expression des mécontentements populaires, cette thèse fournit une meilleure compréhension de l'organisation des oppositions à l'action publique.

La métropole périurbaine a contribué à redéfinir la localisation des opportunités, en particulier en termes d'emploi et de salaire. L'extension de l'aire d'attraction des pôles urbains a favorisé l'accès à la maison individuelle pour les travailleurs de la métropole. Elle s'est aussi accompagnée d'une relocalisation de l'emploi à la limite de la zone urbanisée. Ainsi, on observe que le plan d'urbanisme parisien (construction du RER et des autoroutes) a favorisé l'intégration au marché de l'emploi des résidents des communes d'Ile-de-France les plus éloignées du centre urbain; tandis que les noeuds du réseau de transport, en proche banlieue, ont attiré les emplois. En parallèle, les politiques sociales du marché du travail ont permis une baisse des inégalités au niveau national, par un rattrapage des revenus des travailleurs les moins bien payés, jusqu'à la crise financière de 2008. Cette évolution a davantage bénéficié aux résidants des territoires ruraux. C'est en particulier le cas pour aux femmes vivant dans ces communes et dont la part à temps partiel et au salaire minimum ets plus forte qu'ailleurs. Ces dynamiques de long terme se sont cependant atténuées, si ce n'est inversées, depuis le début des années 2010. En effet, les inégalités salariales ont eu tendance à repartir à la hausse, à la suite de la crise financière de 2008, et le recours à la voiture individuelle a cessé d'être encouragé par les pouvoirs publics, dans un contexte de hausse des prix du carburant et de transition écologique.

Ces phénomènes, inégalités dans le travail et dépendance à la voiture, sont devenus en novembre 2018 les principaux mots d'ordre des Gilets jaunes. On observe en effet une corrélation entre le niveau de mobilisation et les caractéristiques de la métropole périurbaine. Les territoires les plus mobilisés, sur Facebook et les rond-points, sont ceux où les distances domicile-travail sont les plus importantes et où la densité de routes ralenties à 80km/h est la plus forte. De façon emblématique, Priscillia Ludosky, l'initiatrice de la pétition en-ligne à l'origine du mouvement, est une automobiliste habitant une commune périurbaine d'Ile-de-France: Savigny-le-Temple, en Seine-et-Marne. Sa pétition porte sur le coût de la dépendance à la voiture pour se rendre au travail. On observe ensuite une large extension géographique de la mobilisation à toute la France, tant sur internet qu'au niveau des blocages.

Les questions d'inégalité territoriales que le mouvement a soulevé ont incité à appréhender davantage les questions économiques et sociales à l'aune de leur incidence géographique. Cette thèse se propose de contribuer à ce débat.

Les deux premiers chapitres portent sur l'évaluation de politiques publiques et leur incidence au niveau local. Tandis que les deux derniers s'intéressent au mouvement des Gilets jaunes, qui prend ses racines dans l'évolution du paysage métropolitain et du marché du travail. Le premier chapitre porte ainsi sur l'effet des réformes du marché du travail misent en place au tournant des années 2000, en étudiant l'évolution de la distribution des revenus entre homme et femme, ainsi que sur les disparités entre territoires en terme de revenus du travail de leurs résidants. Le second chapitre investigue ensuite la fabrique de la métropole périurbaine à travers l'étude du cas Parisien et l'installation simultanée du RER et d'autoroutes nationales. Le troisième chapitre décrit la mobilisation des Gilets jaunes et s'intéresse aux liens entre les territoires mobilisés et les déplacements domicile-travail; tandis que le quatrième et dernier chapitre étudie la dynamique du mouvement, avec le renforcement de l'activité sur internet par le blocage des rond-points mais qui n'a pas permis de consolidation du mouvement sur le long terme.

La suite de cette introduction décrit les motivations scientifiques de l'étude de la métropole périurbaine et du mouvement des Gilets jaunes, qu'elle a vu naître, avant de présenter les différents chapitres et leur contribution à la recherche académique.

## Structuration du territoire

La construction de la ville est un processus dynamique, qui souvent échappe à la planification, avant d'être reprise en main par la puissance publique. La concentration des emplois dans les métropoles a fait de l'espace urbain et de sa capacité d'attraction un enjeux de politiques publiques et un sujet de recherche important en économie. Les ménages sont à la fois consommateurs d'espace pour vivre et en recherche d'emploi pour échanger leur force de travail contre un revenu. Leur mise en concurrence pour le foncier et les entreprises, sur un espace limité, contraint leurs choix de localisation et détermine le niveau des prix de l'immobilier et des salaires[1]. De

[1] Pierre-Philippe Combes, Duranton Gilles Duranton et Laurent Gobillon: *Spatial wage disparities: Sorting matters!*, Journal of Urban Economics, 2008, et *The costs of agglomeration: House and land prices in French cities*, The Review of Economic Studies, 2019.

plus, les lieux de résidence et de travail ne sont pas homogènes. Chaque lieu a en effet des caractéristiques intrinsèques, appelées aménités[2], mais aussi des effets d'entraînement, appelés économies d'agglomération[3] qui les rendent plus ou moins attrayants. Ces différentes composantes du problème de localisation rencontré par les ménages les ont conduit à s'éloigner de leur lieu de travail[4]. Ce phénomène participe à l'émergence de la métropole moderne (mono-centrique) puis périurbaine (avec la localisation de certains emplois en périphérie). La puissance publique a donc cherché à réduire le temps de ces trajets domicile-travail pour rapprocher les travailleurs des emplois. Mais c'est l'effet inverse qui s'est produit avec un accroissement plus important des distances parcourues que des gains de temps associés. Dans un contexte de transition écologique et de hausse des prix du carburant, le modèle actuel, basé sur le recours à la voiture individuelle, n'est plus soutenable. Cependant, les études portant sur les modes de transport alternatif à la voiture et la mobilité des ménages sont encore rares.

Le modèle d'économie urbaine classique décrit ainsi une ville mono-centrique[5] où les emplois se localisent au centre et les résidents autour. Les métropoles définissent ainsi les marchés locaux de l'emploi, qui maillent le territoire. Leur taille dépend alors du coût maximal de trajet compatible avec le fait d'aller travailler en centre-ville. D'un point de vue empirique, cette notion se traduit par la définition de villes basées sur des critères statistiques portant sur la direction et les flux de navetteurs. En France, les deux principales nomenclatures définies par l'INSEE sont les *aires urbaines* et les *zones d'emploi*. Les premières correspondent à des regroupements continus de communes, autour d'un centre ayant au moins 10 000 emplois, dont au moins 40% des résidents en emploi travaillent dans le centre de l'aire urbaine. Près de huit cents aires urbaines sont ainsi définies, qui regroupent 85% de la population française. Les secondes sont également des regroupements de communes, comportant au moins 5 000 actifs, définis de façon à ce que le nombre d'actifs résidant et travaillant dans la zone soit le plus élevé possible par rapport au nombre de personnes résidant dans la zone. Les 321 zones d'emploi forment une partition complète et précise du territoire français et la densité de population au niveau des zones d'emploi constitue une mesure alternative, continue, de l'espace urbain en France.

## Fragmentation spatiale et révoltes

Enfin, ces marchés locaux que représentent les métropoles et que capturent les notions d'aires urbaines ou de zones d'emploi définissent la géographie du marché du travail et elles sont intégrées au sein d'institutions et d'une économie nationale. Or, on observe une géographie du marché du travail qui se polarise avec une divergence entre des métropoles, gagnantes de la mondialisation, et des territoires en déclin, comme l'a décrit pour le cas américain Enrico

[2] David Albouy. *What Are Cities Worth? Land Rents, Local Productivity, and the Total Value of Amenities*. The Review of Economics and Statistics. 2016.

[3] Gilles Duranton et Diego Puga. *Micro-foundations of urban agglomeration economies*, in the Handbook of regional and urban economics. 2004

[4] Stephan Heblich, Stephen Redding et Daniel Sturm. *The making of the modern metropolis: evidence from London*. The Quarterly Journal of Economics, 2020.

[5] William Alonso. *Location and Land Use: Toward a General Theory of Land Rent*. 1964. Richard Muth. *Cities and Housing: The Spatial Pattern of Urban Residential Land Use*. Third Series: Studies in Business and Society, 1969. Edwin Mills. *An aggregative model of resource allocation in a metropolitan area*. The American Economic Review, 1967.

Moretti[6] et pour le cas français[7] dans notre ouvrage avec Francis Kramarz et Benoît Schmutz[8]. Ce phénomène de fragmentation a d'abord été documenté dans les plus grandes métropoles, où les écarts entre populations intégrées aux processus mondiaux et populations plus précaires, qui ne peuvent pas être pleinement intégrées à la ville du fait de la faiblesse de leurs ressources, sont les plus importantes[9]. Résultant en une ségrégation des populations, avec une concentration des populations les plus aisées et les précaires dans les centres urbains[10] suivi d'une décroissance des revenus avec l'éloignement du centre urbain[11]. Cette fragmentation du territoire s'observe à toutes les échelles géographiques, entre quartiers comme entre pays. Si la nécessité de penser la juxtaposition de territoires obéissant à des logiques singulières, voire contradictoires, est au coeur du renouvellement de l'analyse géographique opérée depuis trois décennies[12], force est de constater que la science économique est, dans son ensemble, en retard sur ces questions[13]. Par les transferts qu'elles organisent, les sociétés occidentales ont cependant fortement atténué les effets de l'accroissement de la compétition mondiale, au prix d'un endettement élevé[14]. Aujourd'hui, les réductions budgétaires affectent ce système de redistribution implicite et entraînent des tensions entre territoires, comme l'a montré l'émergence du mouvement des Gilets jaunes[15].

Le mouvement des Gilets jaunes s'inscrit dans la longue histoire des mouvements sociaux en France[16]. Par certains aspects, ce mouvement reproduit des schémas historiques familiers. Les politiques fiscales sont en effet souvent évoquées pour comprendre l'origine des soulèvements populaires[17]. En France, des jacqueries médiévales au prix de l'essence à la pompe, les révoltes fiscales ont émaillé l'histoire de la taxation[18]. Dans le même temps, les Gilets jaunes se distinguent des manifestations traditionnelles de plusieurs manières. Premièrement, le mouvement se démarque par de nombreux points de rassemblement décentralisés, souvent autour de rond-points, symboles de “l'automobilisme” à la française. Dès le premier samedi, on dénombre 788 points de blocages. Deuxièmement, la manifestation semble s'être largement organisée sans l'intervention des corps intermédiaires traditionnels que sont les partis et les syndicats : ceux-ci qui ont tardé à rejoindre le mouvement de contestation. Enfin, les réseaux sociaux semblent avoir joué un rôle déterminant dans l'organisation et la médiatisation du mouvement. Mi-décembre, on dénombrait 1 548 groupes Facebook de plus de 100 membres associés au mouvement des Gilets jaunes. Si des manifestations massives ont déjà été lancées et catalysées par des réseaux sociaux dans le monde,

---

6 Enrico Moretti. *The new geography of jobs*. Mariner Books, 2012.

7 Donald Davis, Eric Mengus et Tomasz Michalski. *Labor market polarization and the great divergence: Theory and evidence*. National Bureau of Economic Research. 2020.

8 Thomas Delemotte, Francis Kramarz et Benoît Schmutz. L'emploi et le territoire. Presses de Sciences Po, coll. "Sécurisation des Parcours professionnels", 2021.

9 Edward W. Soja. *Postmetropolis: critical studies of cities and regions*, Wiley-Blackwell, 2000.

10 Edward Glaeser, Matthew Kahn et Jordan Rappaport. *Why do the poor live in cities? The role of public transportation*. Journal of Urban Economics.

11 Stephen LeRoy et Jon Sonstelie. *Paradise lost and regained: Transportation innovation, income, and residential location*. Journal of Urban Economics, 1983. Laurent Gobillon, Anne Lambert et Sandra Pellet. *The suburbanization of poverty: Homeownership policies and spatial inequalities in France*. Document de travail INED 250, 2019.

12 Jacques Lévy et Michel Lussault. *Dictionnaire de la géographie et de l'espace des sociétés*, Belin, 2003.

13 David Autor. *Work of the past, work of the future*. AEA Papers and Proceedings, 109, pp. 1–32, 2019.

14 Laurent Davezies. *La crise qui vient. La nouvelle fracture territoriale*. Seuil, coll. « La République des Idées », 2012.

15 Jacques Lévy, Jean-Nicolas Faucille et Ana Póvoas. *Théorie de la justice spatiale. Géographies du juste et de l'injuste*. Odile Jacob, 2019.

16 Eric Neveu. *Sociologie des mouvements sociaux*. La Découvert. 1996.

17 Ponticelli, J. Voth, H.-J. *Austerity and anarchy : Budget cuts and social unrest in europe, 1919-2008*. Journal of Comparative Economics, 2020.

18 Nicolas Delalande: *Les batailles de l'impôt. Consentement et résistances de 1789 à nos jours*, Média Diffusion, 2011, et *Le retour des révoltes fiscales?*, Pouvoirs, 2014.

une mobilisation d'une telle ampleur est une nouveauté en France. Le rôle des réseaux sociaux dans l'organisation et la dynamique des mouvements sociaux contemporains est encore méconnu. Cependant leur importance semble croissante dans l'émergence de mouvements de protestation de grande ampleur. Dans le contexte des Printemps Arabes, par exemple, l'activite sur Twitter etait un fort predicteur des manifestations effectives[19]. Des resultats similaires ont ete derives dans les contextes chinois[20], americain[21] et russe[22].

[19] Zachari Steinert-Threlkeld, Delia Mocanu, Alessandro Vespignani et James Fowler. *Online social networks and offline protest*. EPJ Data Science, 2015. Daron Acemoglu, Tarek Hassan et Ahmed Tahoun. *The power of the street: Evidence from Egypt's Arab Spring*. The Review of Financial Studies, 2018.

[20] Bei Qin, David Strömberg, et Yanhui Wu. *Why does china allow freer social media ? protests versus surveillance and propaganda*. Journal of Economic Perspectives, 2017.

[21] Neal Caren et Sarah Gaby. *Occupy online : Facebook and the spread of occupy wall street*. Available at SSRN, 2011. Ion Vasi et Chan Suh. *Protest in the internet age : Public attention, social media, and the spread of 'occupy'protests in the united states*. In Politics and Protest workshop. 2013. Marcos Bastos, Dan Mercea et Arthur Charpentier. *Tents, tweets, and events : The interplay between ongoing protests and social media*. Journal of Communication, 2015.

[22] Ruben Enikolopov, Alexey Makarin et Maria Petrova. *Social media and protest participation : Evidence from russia*. Econometrica, 2020.

# Outreach – Vulgarisation

## L'emploi et le territoire

Presses de Sciences Po[1]

Dans les pays industrialisés, la divergence de territoires pourtant voisins en matière d'accès à l'emploi et de niveau de vie constitue l'un des phénomènes les plus marquants des dernières décennies. Dans un contexte d'amélioration constante des moyens de communication et de transport des biens et des personnes, comment expliquer une telle fragmentation spatiale aux multiples conséquences sociales et politiques ? Dans cet ouvrage, nous apportons une réponse à cette question en dressant un panorama des principaux apports de l'analyse économique d'un marché du travail confronté aux contraintes géographiques.

Le premier chapitre pose les jalons à l'aide de la notion de marché local du travail qui correspond à un échelon intermédiaire entre le niveau microéconomique des travailleurs/producteurs et le niveau macroéconomique des pays et des flux liés au commerce international. Il explicite les principaux arbitrages à l'œuvre entre prix du foncier et coûts de la mobilité quotidienne d'un côté, et qualité de l'environnement et opportunités économiques de l'autre. Alors que les coûts associés aux plus grandes agglomérations affectent indifféremment toutes les entreprises et tous les ménages qui s'y trouvent, les opportunités qu'elles offrent en termes de gains de productivité, de carrière ou d'expansion, sont très inégalement réparties. Il en résulte une concentration spatiale des travailleurs les plus qualifiés et des entreprises les plus productives au cœur des agglomérations les plus dynamiques. L'étalement urbain, rendu possible par le progrès technologique, atténue ces différenciations régionales, mais bien souvent au prix d'une forte dépendance à la voiture individuelle. Combinés à la tertiarisation de l'économie, ces mécanismes conduisent à la métropolisation du marché du travail : concentration spatiale des emplois, doublée d'un éloignement les lieux de vie de la plupart des travailleurs.

[1] Thomas Delemotte, Francis Kramarz et Benoît Schmutz: *L'Emploi et le territoire*. Presses de Sciences Po, *Coll.* Sécuriser l'Emploi. Mai 2021. `http://www.pressesdesciencespo.fr/fr/book/?gcoi=27246100751580`

Simultanément, près d'un quart des Français vivent en milieu rural et les régions les plus touchées par des décennies de déclin industriel abritent, encore aujourd'hui, une population importante. Le second chapitre analyse les interactions régionales et les divergences entre villes, à partir de l'étude comparative de la mobilité des travailleurs et de la mobilité de l'emploi. Aux différentes étapes du cycle de vie des ménages et des entreprises correspondent des besoins géographiques spécifiques, mais dans un contexte où les infrastructures changent lentement, déménager est une décision coûteuse, souvent accomplie à contrecœur. Pour les ménages, contraints financièrement et qui disposent d'informations parcellaires sur les perspectives d'emploi et le mode de vie offerts par les autres régions, un déménagement s'accompagne d'un risque élevé de perte en capital tant financier qu'humain et social. Ce constat vaut d'autant plus pour les travailleurs les moins qualifiés. Du côté des entreprises, les choix de localisation s'accompagnent d'investissements souvent irréversibles. C'est ainsi la création de nouveaux établissements qui accompagne majoritairement la (re)localisation des emplois.

La puissance publique a un rôle à jouer pour rapprocher travailleurs et emplois. Cependant, comme le détaille le troisième chapitre, les évaluations disponibles pointent presque toujours la faible efficacité des politiques insuffisamment ciblées et qui pâtissent du manque de coordination des différentes strates de gouvernance qu'elles impliquent. En revanche, lesprogrammes ciblés sur la mobilité des résidents des territoires marginalisés semblent plus efficaces. Toutefois, les aides au déménagement doivent être massives pour compenser les risques encourus et le fort attachement des ménages à leur lieu de résidence. Quant aux aides à la mobilité quotidienne, elles se heurtent aux contraintes liées à la transition écologique, ce qui limite la faisabilité politique des programmes d'accession à la voiture individuelle. Elles se heurtent aussi aux contraintes budgétaires qui bornent l'ambition des grands travaux d'infrastructure. Enfin, il est très difficile de prévoir le succès ou l'échec des aménagements destinés à améliorer l'attractivité des territoires : les aménités locales jouent un rôle de premier ordre pour attirer la main-d'œuvre qualifiée, mais leur pouvoir d'attraction obéit à des logiques qui échappent largement aux leviers de l'action publique. Le même constat s'applique aux politiques volontaristes visant à créer, souvent ex nihilo, des clusters industriels.

Chercher à modifier directement les comportements de mobilité des demandeurs d'emploi ou des entreprises revient à jouer davantage sur les symptômes que sur les racines du problème : les territoires les plus dynamiques sont devenus inabordables pour une majorité d'acteurs. Dans cette optique, intervenir sur l'offre immobilière afin que ces territoires ne soient plus inabordables pour de nombreux ménages et entreprises s'avère être une piste à privilégier. Elle se justifie d'autant plus que les acteurs privés n'ont pas la force de frappe suffisante pour pallier la pénurie de logements au niveau régional, comme en témoignent les difficultés rencontrées par les géants de la tech en Californie qui, malgré les moyens immenses dont ils disposent, peinent à assurer un logement décent à tous leurs employés. En France, s'il n'y a pas de pénurie globale de logement, de nombreux facteurs limitent l'offre immobilière au centre des agglomérations les plus attractives. Les raisons sont d'ordre patrimonial, institutionnel et politique, à l'image des mouvements d'opposition aux grands projets immobiliers qui ponctuent régulièrement l'actualité et fédèrent riverains et responsables politiques locaux autour de préoccupations esthétiques et environnementales. Pour contourner ces obstacles, contrairement à la France, l'Allemagne a confié aux préfets, plutôt qu'aux maires, la responsabilité du permis de construire.

L'heure est à l'innovation. La surélévation des bâtiments, qui revient à ajouter un ou plusieurs étages tout en préservant le tissu urbain existant, est une solution encore trop peu exploitée alors qu'elle est monnaie courante dans d'autres pays. A plus court terme, il faudrait pouvoir diminuer la proportion de logements vacants dans les zones à forte demande. Dans la mesure où l'introduction, en 1999, de la taxe sur les logements vacants a conduit à la remise sur le marché d'un nombre important de logements, une nouvelle hausse de cette taxe ne devrait pas être exclue. Par ailleurs, si l'essor des plateformes numériques comme Airbnb semble avoir eu un effet négatif sur la quantité de logements mis en location pour une longue durée, cette tendance n'est pas une fatalité. La souplesse offerte par de tels outils pourrait être mise à profit, dans un cadre réglementaire adapté, pour réduire les problèmes d'appariement entre l'offre et la demande immobilières, aussi bien dans le secteur résidentiel que commercial.

Sur le plan institutionnel, on peut espérer qu'un renforcement de la dynamique actuelle des regroupements de communes permette une meilleure prise en compte des tensions immobilières à l'échelle de l'agglomération et non plus seulement à l'échelle du quartier. Encore faut-il que ces regroupements atteignent une taille suffisante pour offrir des logements mais aussi des services publics et des moyens de transports adéquats. A cet égard, il ne serait pas inutile de s'inspirer de l'exemple suédois où cet effort de regroupement a été conduit de façon très volontariste.

A plus long terme, augmenter l'offre de logement impose évidemment de prendre en compte les coûts monétaires, sociaux et environnementaux associés à la croissance métropolitaine encore trop souvent synonyme d'étalement urbain et de congestion. Une réflexion sur les transports urbains est donc nécessaire. Les projets de trains rapides et à grande capacité de voyageurs, type RER (Grand Paris Express ou les RER métropolitains préconisés par la Loi d'Orientation des Mobilités), sont des pistes intéressantes pour réduire l'usage de la voiture dans les métropoles. Le covoiturage et l'intermodalité sont d'autres pistes prometteuses pour accroître les capacités du trafic routier (nombre de personnes par voiture, vélos, bus) et mettre en avant les complémentarités entre les différents modes de transport.

## Entre facebook et le rond-point: « La double originalité du mouvement des 'Gilets jaunes' »

Le Monde[2]

Alors que le mouvement des « *Gilets jaunes* » fête son premier anniversaire et que la quatrième « assemblée des assemblées » a récemment réuni à Montpellier quelques 450 représentants de 200 ronds-points, les origines de cet objet politique non identifié restent très discutées. En quelques mois d'été et d'automne 2018, son irruption dans le débat public a pourtant obligé le gouvernement

[2] Pierre Boyer, Thomas Delemotte, Germain Gauthier, Vincent Rollet et Benoît Schmutz: *Entre facebook et le rond-point: « La double originalité du mouvement des 'Gilets jaunes' »*. Le Monde, Tribune. 15 novembre 2019. https://www.lemonde.fr/idees/article/2019/11/15/entre-facebook-et-le-rond-point-la-double-originalite-du-mouvement-des-gilets-jaunes_6019218_3232.html#xtor=AL-32280270

à infléchir de façon drastique la mise en place de son programme de fiscalité environnementale et remis sur le devant de la scène la dimension territoriale des inégalités qui traversent aujourd'hui la société française.

Rappelons la chronologie : annoncée dès le début de l'année 2018, alors qu'elle ne faisait pas partie du programme du candidat Macron, la limitation de la vitesse à 80km/h sur les routes secondaires entre en vigueur le 1er juillet 2018. Différents rassemblements, d'abord de motards puis d'automobilistes, ont lieu pendant l'été pour s'opposer à cette mesure et à la hausse des prix du carburant, sans pour autant parvenir à s'agréger. En octobre, tout s'accélère : le 10, un appel en faveur d'un « blocage national » est lancé sur le réseau social Facebook ; le 17, une vidéo est publiée sur YouTube contre la hausse des prix du carburant, les péages et les radars ; enfin, le 24 octobre, le gilet de haute visibilité de couleur jaune est proposé comme signe de ralliement, à mettre en évidence sur les pare-brise. Parallèlement, une carte nationale des différents projets de blocage est mise en ligne, afin de coordonner la mobilisation du 17 novembre. Celle-ci constituera l'acte I du mouvement, réunissant, selon les estimations, entre 300 000 et plus de 1 million de manifestants à travers tout le pays.

Manifestations physiques et ses manifestations virtuelles Ce bref déroulé des événements souligne la double originalité du mouvement des « *Gilets jaunes* », formé de nombreux points de rassemblement décentralisés et organisé sans l'intervention des corps intermédiaires traditionnels, mais avec un recours massif aux réseaux sociaux numériques. Pour comprendre le mouvement, il faut donc à la fois documenter ses manifestations physiques et ses manifestations virtuelles. Nous avons identifié près de 800 points de blocage sur le territoire métropolitain, dont un bon nombre ont été occupés pendant plusieurs mois. Quant à la mobilisation numérique, notre recension de l'activité sur Facebook à la mi-décembre, à la fin de la phase d'expansion du mouvement, relève plus de 1 500 « groupes Facebook » de plus de 100 membres directement associés aux « gilets jaunes ».

Facebook a joué un double rôle dans le mouvement : d'une part, la plate-forme a permis la diffusion d'informations et d'opinions liées au mouvement. D'autre part, elle a permis de coordonner les protestataires. Notre étude de l'évolution du nombre de groupes associés aux « *Gilets jaunes* » (représentée par le graphique [cf. Chapitre 3]) met en évidence différentes étapes de la montée en puissance du mouvement. En nombre négligeable à la mi-octobre, ces groupes sont déjà plus de 500 à la veille du 17 novembre. Ils rassemblent notamment des automobilistes mécontents, et s'appuient en partie sur des groupes hostiles aux politiques gouvernementales créés plus tôt dans l'année.

Ces nouveaux espaces de mobilisation ont grandement facilité l'organisation de l'acte I. On observe ensuite une seconde phase de croissance, encore plus importante, entre le 17 novembre et la mi-décembre. Lancée sur Facebook, la mobilisation du 17 novembre a donc elle-même, en retour, contribué à l'expansion de la mobilisation numérique. A l'issue de cette seconde phase, plus de 1 500 groupes de plus de 100 membres sont constitués, soit plus de 4,2 millions de membres (en intégrant les groupes créés aux échelles régionale et nationale), qui ont déjà partagé près d'un million et demi de messages. La rapidité de ce processus témoigne de la convergence des contestations sur Facebook et de l'efficacité de ce réseau social pour diffuser et coordonner un mouvement social de grande ampleur. Si des manifestations massives avaient déjà été lancées et catalysées par des réseaux sociaux dans le monde, à commencer par les « printemps arabes » de 2011, une mobilisation d'une telle ampleur est inédite en France.

Des départements peu denses, très mobilisés Autre particularité, ces groupes Facebook sont très souvent, dans leur appellation, associés à des territoires précis : les trois quarts d'entre eux font référence à un département ou à un ensemble de communes, regroupant sur ces deux échelles plus de 1,5 million de membres. L'emprise spatiale de la mobilisation sur Facebook est illustrée ici par deux cartes [cf. Chapitre 3]. Elles montrent le nombre de personnes membres d'un groupe Facebook dans chaque département, en valeur absolue, puis rapportées au nombre d'habitants du département. Si, en valeur absolue, la mobilisation en ligne est liée à la densité de population (à l'exception notable de la région parisienne), la proportion d'habitants appartenant à un groupe Facebook est particulièrement élevée dans les territoires périphériques : façade atlantique, arc méditerranéen, Nord et Alsace. Certains départements peu denses sont en fait très mobilisés si l'on rapporte la mesure au nombre d'habitants, à l'instar du Lot, de la Charente ou des Hautes-Alpes.

La comparaison des deux modes de mobilisation, sur les ronds-points et sur Facebook, met globalement en évidence le fait que les territoires les plus fortement touchés par des blocages sont également ceux où la mobilisation sur Facebook est la plus importante : les deux formes de mobilisation sont bien complémentaires l'une de l'autre. Enfin, si l'on change de focale et que l'on observe le mouvement à l'intérieur même de chaque département, on relève parfois une forte hétérogénéité entre des territoires voisins, comme dans le Cher ou la Marne, par exemple.

Contraintes géographiques et difficultés économiques locales Les déterminants de la mobilisation sont à la croisée de nombreux enjeux, qu'il est possible de regrouper en quelques dimensions principales : d'une part, les contraintes géographiques auxquelles font face les populations, notamment au travers de la distance moyenne entre le lieu de travail et le lieu de résidence des salariés ; ces contraintes géographiques viennent s'ajouter aux difficultés économiques locales. D'autre part, les décisions politiques prises ou programmées en 2018, à savoir l'augmentation des taxes sur le diesel et la réduction de la vitesse autorisée, qui viennent s'imposer à des territoires plus ou moins favorablement disposés vis-à-vis du pouvoir en place.

Ces différentes dimensions sont bien sûr liées, selon des modalités diverses qui expliquent l'hétérogénéité spatiale du mouvement. De leur analyse jointe découlent plusieurs enseignements importants. Si, en première instance, la mobilisation semble associée à certaines préférences politiques, frappant par exemple des zones à fort taux d'abstention ou ayant voté de façon plus importante qu'ailleurs pour Jean-Luc Mélenchon ou Marine Le Pen, ces corrélations disparaissent dès que l'on prend en compte le contexte socio-économique local, au travers du taux de chômage ou des inégalités salariales : le mouvement des « *Gilets jaunes* » ne peut donc pas s'interpréter comme un « troisième tour » de la dernière élection présidentielle.

A contrario, l'isolement géographique de certains territoires semble avoir joué un rôle déterminant dans la constitution du mouvement. En particulier, les territoires plus dépendants de l'automobile (notamment en raison de plus grandes distances domicile-travail) ont davantage contribué aux premières mobilisations. L'abaissement de la vitesse maximale autorisée sur un grand nombre de routes à 80km/h avait déjà suscité la colère d'un grand nombre de français. Le mouvement des « *Gilets jaunes* » a puisé dans ce potentiel latent de mobilisation, et la perspective d'une nouvelle hausse des prix du carburant, nouvelle contrainte à la mobilité, a transformé ce potentiel en l'une des mobilisations les plus fortes de cette décennie.

# Faut-il doter les métropoles françaises de RER?

Le Monde[3]

Les métropoles françaises seront-elles dotées un jour de RER comme Paris ? C'est du moins ce que propose SNCF Réseau dans son schéma directeur présenté au gouvernement[4]. En parallèle, de nombreux candidats aux municipales réclamés la mise en place d'un réseau express métropolitain dans leur ville. Si l'objectif affiché de ces propositions est de réduire les émissions polluantes et les embouteillages en ville, les solutions retenues impacteront le quotidien de chacun et nécessitent un éclairage des répercussions sur l'économie locale.

L'organisation des villes a considérablement changé ces dernières décennies avec l'affirmation de l'automobile dans les déplacements du quotidien, rapide et individuelle, la voiture a considérablement agrandi l'espace des possibles (maisons individuelles, grandes surfaces, … ). A l'inverse, la part des transports en commun a décliné. Organisés par un réseau de stations, les habitants y ont significativement moins recours quand ils habitent à plus de quelques centaines de mètres d'une station. Ainsi, relancer les mobilités collectives implique une réflexion sur la densité urbaine.

Les équilibres économiques sont centraux pour comprendre la localisation des ménages. Le plus grand accès d'un quartier au reste de la métropole, permis par un arrêt de RER, provoque un accroissement de la demande de logement à proximité des stations. Néanmoins, le faible ajustement du nombre de logement disponible se traduit souvent par une hausse des prix des loyers. L'excès de demande est absorbé par les communes voisines, plus éloignées, conduisant à un étalement urbain et finalement un plus faible recourt au transport en commun. Une analyse de la mise en place récente d'un réseau express à Los Angeles montre une densification de l'emploi autour des nouvelles stations, mais pas d'augmentation du nombre de résidents dans ces quartiers[5]. Une autre étude sur le RER en Ile-de-France présente des résultats similaires, avec une évolution de la sociologie des résidents, suggérant une gentrification des communes du nouveau réseau[6].

Les effets sur l'emploi sont tributaires des interactions entre : les gains générés par la densité d'emploi (mutualisation des ressources, co-apprentissage, meilleurs appariements) et les caractéristiques spécifiques de chaque quartier (patrimoine, géographie, … ). Un projet de RER métropolitain doit prendre en compte ces interactions économiques pour cibler les quartiers les plus prometteurs : où une densification de l'emploi permettrait de démultiplier les atouts locaux. Une étude à Bogota, en Colombie, montre que les zones où l'emploi a le plus augmenté étaient aussi … les plus centrales[7].

Initialement plus productives, elles bénéficient davantage de l'accroissement de leur marché de l'emploi. Ces remarques éclairent les enjeux des RER métropolitains. Par exemple, la géographie

---

3 Thomas Delemotte: *Faut-il doter les métropoles françaises de RER?*. Le Monde, Tribune. 4 novembre 2020. https://www.lemonde.fr/idees/article/2020/11/04/faut-il-doter-les-metropoles-francaises-de-rer-comme-paris_6058426_3232.html

4 https://www.ecologie.gouv.fr/sites/default/files/EF%26SEM-SD%20VF%2006%2004%202020.pdf

5 C. Severen. Commuting, Labor, and Housing Market Effects of Mass Transportation. FED Philadelphia, 2019.

6 T. Mayer et C. Trevien. The impact of urban public transportation evidence from the Paris region. Journal of Urban Economics, 2017.

7 N. Tsivanidis. Evaluating the Impact of Urban Transit Infrastructure: Evidence from Bogota's TransMilenao. UC Berkley, 2020

particulièrement accidentée de Marseille y contraint fortement l'extension de la surface foncière. Cela doit inciter à identifier les zones qui pourront accueillir les navetteurs attirés par l'extension du réseau actuel. La gare Saint-Charles est déjà saturée, et l'offre de logement n'ouvre plus de débouché à moins de renforcer le projet « Euroméditerranée » en développant une skyline à la New-Yorkaise pour contourner les contraintes physiques ! A Lyon ou Toulouse, la géographie est plus clémente, en résulte la cohabitation avec d'autres agglomérations. L'offre de transport doit y être clarifiée avec : des TER reliant le pôle des agglomérations de la région et un RER centré sur les mobilités du quotidien. Le premier réseau existe déjà et doit être concentré sur un minimum d'arrêts. Le second, le nouveau RER, ne doit servir qu'une aire urbaine à la fois, pour favoriser son insertion dans le tissu urbain et fournir un service adapté à des usages fréquents (horaires, ponctualité, places debout, ...).

La France a besoin de métropoles dynamiques. Cela implique de mettre en valeur leurs atouts (centres de décision, universités, ...) et d'y faciliter les mobilités quotidiennes, afin de stimuler l'économie locale et de permettre l'accueil des nouveaux résidents. Ces objectifs doivent s'accompagner d'une réduction massive des émissions carbone. La prise en compte de l'intermodalité doit encourager l'usage des vélos en libre-partage pour réduire le temps d'accès aux stations du RER.

De telles ambitions nécessitent une gouvernance claire, potentiellement indépendante de la gestion du reste des affaires locales. En s'inspirant par exemple des Waterschap hollandais : conseils élus par la population pour gérer (uniquement) l'administration des polders (terres endiguées), dans un principe de « démocratie fonctionnelle ». Par ailleurs, une limitation des projets de RER sur quelques métropoles, au début, doit permettre de concentrer les moyens de l'Etat et d'identifier les bonnes pratiques. Enfin les objectifs à atteindre en terme d'aménagement urbain doivent être clairement définis, et le RER ne doit être qu'un moyen de les atteindre.

# CHAPTER 1

## Inequality and Earnings in France
### *Labor Market Reforms, Gender, and Places*

with Francis Kramarz and Elio Nimier-David

**Abstract**

Since the beginning of the 21st century there has been a renewed interest for income inequality. This chapter focuses on labor earnings, to shed light on the impact of labor market reforms on the distribution of individuals' earnings. We first provide evidence of a strong catching-up of female workers at the bottom of the earnings distribution over the period, especially within rural municipalities. We then show that it has induced a substantial decrease in the gender pay gap, as well as a reduction in inequality between rural and urban places in term of median earning. Finally, we show that, if the reduction in inequality holds for the core of the distribution, a strong divergence has occurred with very top earners, who have seen a dramatic increase of their earnings over the whole period. These top-earners are mostly located in the urban unit of Paris, it has thus induced a divergence in mean earnings between Paris and the rest of France.

**Keywords**: Income Dynamics, Inequality, Labor Market, Public Policies.
**JEL Classification**: D63, J38, J31, R12.

Freely adapted from: Kramarz, Francis, Nimier-David, Elio, and Delemotte, Thomas. "*Inequality and Earnings Dynamics in France: National Policies and Local Consequences*". Quantitative Economics (*conditionally accepted*).
We would like to thank Fatih Guvenen, Luigi Pistaferri, and Gianluca Violante for initiating and coordinating the ambitious *Global Income Dynamics* database project. We also thank Serdar Ozkan and Sergio Salgado for there tremendous work on harmonizing the core statistics, as well as Ondine Berland, Pierre Boyer, Pauline Carry, Bertrand Garbinti and seminar participants at the Global Income Dynamics Conferences at Stanford (SITE, 2019) and held virtually from the University of Minnesota (SED, 2020) for helpful comments and suggestions.
This paper has been conducted in collaboration with the CASD with special thanks to Kamel Gadouche, Marie Vidal and Raphaelle Fleureux for their help and support. The access to the French administrative data has been made possible within a secure environment offered by the CASD (Ref. 10.34724/CASD). This paper has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme: Grant Agreement: 741467 FIRMNET.
The results presented in this chapter are the sole responsibility of its authors.

# Introduction

Since the beginning of the 21st century there has been a renewed interest among economists to study and understand the structure and evolution of income inequality. Focusing solely on cross-sectional inequality leaves out many aspects of the distribution of wealth and income: What are the forces that shape these trends? How do public policies affect earnings inequality? What is the local declination of national trends? These questions are crucial to evaluate policies that aim at addressing inequality in an efficient and inclusive way, such as labor market reforms. Since labor earnings constitute a substantial share of total income,[1] it is important to know the impact of these reforms on individuals' total labor income and inequality among occupied workers. In addition, the recent social protests on the urban fringe have highlighted (such as the Yellow Vests in France) the importance of better accounting for the heterogeneous impacts of these policies according to where individuals live.

Public policies do not always explicitly account for their incidence on price setting and individual earnings. For instance, some labor market reforms, such as the reduction of the workweek, aim first at enhancing working conditions rather than modifying wage setting. However, working less for the same wage mechanically increase employers' labor costs. Therefore, tax level adjustments are often used to mitigate this undesired effect on market wage setting. Which in turns can impact other type of workers, like part-time workers. The effects of these reforms on individuals' total earnings and inequality are ambiguous, although their impacts might be substantial.

This chapter focuses on two major labor market reforms that have been implemented in France during the turn of the 2000s: the reduction of the workweek from 39 to 35 hours worked and the decrease of employer-paid payroll taxes around the minimum wage. It shed light on the incidence these reforms have had on individuals' labor earnings and inequality. We first provide evidence that hours worked have declined following the reform, except for part time workers, who are mostly women. Hourly wage have conversely increased to sustain workers' total earnings. In addition, we show that the increase in hourly wage has been stronger for women over the period. We then turn to the study of the distribution of individuals' total labor earnings by year and show a strong catching-up of female workers at the bottom of the earnings distribution. In addition, since the share of low-paid female workers is more important in low-density areas, we find that the effect of the reforms have corresponded in a catching-up of rural areas residents. This trends induced a substantial decrease in the gender pay gap over the period, as well as a reduction of inequality between rural and urban places in term of median earning. Finally, we show that, if the reduction in inequality holds for the core of the distribution, a strong divergence has occurred with very top earners, who have seen a dramatic increase of their earnings over the whole period. These top-earners are mostly located in the urban unit of Paris, it thus induced a substantial disparity in mean earnings between Paris and the rest of France.

France is a particularly interesting setting with major reforms of the labor market institutions that had a strong impact on inequality.These national labor market reforms affect places and individuals with very heterogeneous economic conditions, making the effect of these policies difficult to assess. This framework is therefore particularly relevant for studying the disparities

[1] For instance, Drechsel-Grau et al. (2022) show that total labor earnings account for 83.6% of men total income and 89.4% for women, in Germany.

that national reforms induce between individuals, and between places.

We compute individual labor earnings from Social Security data (DADS), which provide exhaustive information for each employment spell, to study earnings inequality and individual dynamics between 1991 and 2016. The first part of the chapter highlight the effect of the labor market reforms, we decompose earnings into hours worked and wage, the 35-hour working week implied a mechanical reduction in hours for all, except for women at the bottom of the wage distribution (below the median), who mostly work part-time. For female workers, wage growth has been substantial but particularly so at the bottom and higher at all percentiles compared to men, where it only compensates the reduction in hours worked.

We then provide a systematic investigation of the variation of the earnings over time. We find that the reduction of the workweek and the suppression of employer-paid payroll taxes around the minimum wage, translates into a marked increase for the bottom percentiles of the earnings distribution in the 2000s, especially among female workers. Earnings flatten after the 2008 financial crisis, the more so at the bottom of the distribution. By contrast, we observe a very strong increase in the labor earnings concentrated at the very top of the distribution, which can not be related to the reforms we focus on and is not captured by standard measure of inequality.[2] We then define five categories of territories that follow the gradient of urbanization, including rural areas to account for all workers, to distinguish individuals in their relative access to the local labor market, since most of jobs concentrated in the build-up area of suburban metropolises. We show that the strong increase in earnings at the bottom of the earnings distribution arise almost only in low-density areas. It is consistent with these territories hosting the largest share of female workers at the minimum wage that the labor reforms targeted. In contrasts, suggestive evidence shows that alternative mechanisms, such as workers' cross-territory mobility or job-to-job transitions, are not able to explain these trends. By contrast, for Paris urban unit residents, the highest earnings growth occurred for top-earners. The increase of Paris highest earnings percentiles reflects the concentration of France top-earners in Paris, who are also the one, at the national level, who experienced the largest earnings growth over our study period.

The second part of the paper study the effect of these changes in the earnings distribution on inequality. The strong earnings growth at the bottom of the women earnings distribution induced a decrease in overall inequality and especially in the gender pay gap, living aside very top earnings. The opposing trends in earnings growth between male and female workers resulted in a 37% decrease in the (unconditional) gender pay gap. However the gap remains substantial. The study of the earnings dynamics suggests that maternity remains a substantial source of earnings risks, since female workers between ages 25 to 34 face higher uncertainty. Between territory of residence, the median earnings are converging between rural and urban municipalities right after the reforms. Conversely, the level of mean earnings inequality between territories exhibits strong diverging trends at the beginning of our period of study, in the 1990s, especially between Paris urban unit and the rest of France. It suggests that labor market reforms foster a democratization of standard labor conditions between places with higher earnings for low-paid workers and a catching up of low-density residential locations. While, the behavior of mean earnings is heavily influenced by top earnings, who concentrate in Paris urban unit and experience a sharp growth of their earnings

[2] e.g. Standard deviation or 10th to 90th percentiles ratio).

in the late 1990s. It highlights the segregation of labor market opportunities by metropolises. We thus pursue by providing insight of the evolution of the urban wage premium over time by earnings percentile and show that agglomeration economies benefit more to workers with higher earnings.

### 1.0.1 Related literature

This chapter contributes to several strands of the literature, by studying income inequality, with a focus on total labor earning, gender differences and spatial disparities.

First, we add to the debate on income inequality by providing insights on how market reforms can affect inequality in the long run. To this end, we focus on major labor market reforms and labor earning. Previous studies provide insight of the incidence of exogenous source of variation like wars (Piketty, 2003), or tax reforms (Bozio et al., 2020b) or Bozio et al. (2020a) and Garbinti et al. (2018) that show that the long-term (1900-2018) decline in inequality in France is explained in part by the fall in pretax inequality. Many other sources of inequality have been emphasised like Godechot (2012), that looks at the role of finance in the rising wage inequality, or Autor et al. (2020) that looks at superstar firms. We show in complement to these studies that labor market reforms are also able to affect inequality. In contrast to the studies on total income, we show that labor earnings inequality in France have rather declined in recent decades, following the improvement of working standards, which differ to results found in other European countries like the United-Kingdom (Bell et al., 2022), Germany (Drechsel-Grau et al., 2022) or Italy (Hoffmann et al., 2022). However, in line with other studies, we find a dramatic increase of the earnings of top-earners, a category of workers that is not targeted by the reforms. Our results are also in line with a "concentration of affluence" in the largest metropolises, similarly to what is observed in the US (Gaubert et al., 2021a).

In addition, new debates on income risks have been open by the seminal contribution of Guvenen et al. (2021), which highlight the non-linearity of income growth, as well as its substantial deviation from normality. The Global Repository on Income Dynamics (GRID), to which this chapter belong, confirm this pattern in other countries. Pora and Wilner (2020) already provided insight on the non-linearity of earnings risk for France, emphasizing the role of working time in the deviation from normality, especially at the extensive margin (transition to or out of employment). We complement this work by comparing the distribution of earnings risks between men and women. We show that the difference in the earnings dynamics between the two groups are substantial and need to be taken into account. Especially our results suggest a major incidence of maternity in women earnings growth process, compared to men, with an important volatility of earnings risks, associated with larger negative shocks, below age 35.

It also relates to the specific literature at the country level for France. We therefore complement Verdugo (2014) with a systematic analysis of labor earnings evolution for the whole distribution, that we extend until 2016 and to women. Our approach also contrast with previous work, like Charnoz et al. (2014), since we look at total labor earnings rather than market-level wage setting. It allows us to better account for labor earnings risk that individuals faces in their career of work, which proved important when we look at gender pay gap or spatial inequality. More broadly, we also link earnings dynamics with labor market reforms. We thus complement works focusing on the incidence of the minimum wage at the extensive margins (unemployment) of labor market reforms in France (e.g. Kramarz and Philippon, 2001).

Our comprehensive approach of employment, accounting for both male and female workers, allow us to provide new insights on the evolution of the gender pay gap over the earnings distribution and over time (see e.g. Goldin 2014 or Blau and Kahn, 2017 for a review). We describe how the labor market reforms, targeting low-wage workers, contributes to reduce gender earnings gap in the bottom half of the earnings distribution, while disparities remain large at its top. Contrasting earnings dynamics between men and women provide also strong support for a substantial incidence of the "child penalties" (see e.g. Kleven et al. 2019a; Kleven et al. 2019b) with substantially higher earnings volatility and negative skewness (higher probability of large negative shocks) for women aged 25 to 34 compared to men at the same age and older men or women.

In a second step, we contribute to the urban literature that looks at segregation and at the urban wage premium. In particular we extend previous approach by including all places, beyond the only urban unit. It allows us to observe all workers and to take into account the suburbanization of the metropolis, with most of rural residents commuting long distance to work in the urban unit (which includes inner suburbs). In contrast, previous studies have mainly emphasized both the gain from agglomeration economies at workplace, comparing metropolises between them (e.g. Duranton and Puga, 2004; Combes et al. (2012a); Combes et al., 2012b) or urban segregation, comparing neighborhood within the same urban unit (e.g. Glaeser et al., 2008a; Couture et al., 2020). We add to these works by focusing on the heterogeneity in earnings changes over time according to individual access to labor market opportunities, proxy by the distance of their place of living from the urban core, where the largest share of jobs locate. Focusing on job search, LeBarbanchon et al. (2021) have shown that commuting could affect employment decision and reservation wage, which matter to explain the gender wage gap. In the present chapter we look at individuals' total earnings and distance of residents to city centers. We thus provide a new systematic investigation of the changes of the yearly earnings distribution by residential places between men and women.

Finally, we mitigate recent results on city center gentrification (Glaeser et al., 2001; ; Couture et al., 2020), at least for France, by showing first that earnings of (urban) suburbs are substantially higher than low-density residents and, second, that the urban core of metropolises still host a substantial share of low-paid workers (in line with Glaeser et al., 2008a). The main mechanism that seem able to explain these trends is the strong reduction of inequality between rural and urban residents foster by major labor market reforms. We observe a strong catching up of earninf for low-paid female residents of low-density areas. In addition, we show for France that, the "concentration of affluence", described by Autor (2019) for the US, is almost only a matter of Paris versus the rest of the country. In addition, for labor market insiders, we cannot speak about a "democratization of poverty" like Gaubert et al. (2021b) observe in the US, but rather a democratization of labor standard (at least for labor earnings) that labor market reforms have promoted. Our results provide thus insights for the renewing debate on place-based redistribution (Gaubert et al., 2021c) since we show how national reforms can contribute to the reduction of inequality between places.

The next section presents the institutional background of the French labor market, while section 1.2 presents the data we used in this chapter. Section 1.4 presents our main results at the national level, with a focus on gender disparities. Section **??** refines the analysis accounting for spatial heterogeneity. Finally, the last section discuss these results and conclude.

## 1.1 Institutional Background

### 1.1.1 Participation rate and Unemployment in the French Labor Market

The French labor market underwent structural changes during the period covered by our data, from 1991 to 2016. The first important trend, shared with many other developed countries, is the increase of the participation rate driven, one the one hand by a massive entry of women in the labor market, since the 1980s, and on the other hand by the increasing participation of the oldest workers since 2000. Figure 1.1 display participation rate over time for different workers groups. This increase has occurred jointly with a convergence between men and women participation rate, started in the 1980s. In trends, it has translated into a decline of labor force participation among men, while the participation has expanded among women. For the men, most of the decline arisen at the beginning of the 1980s, with a participation rate that already declined in the 1980s. This strong decline is due to the reduction of the participation rate among older age groups. In particular, pre-retirement plans in the 80s pushed down participation of people above 55, especially for men. Recent pension reforms in the 2000s and 2010s, implying the extension of the retirement age, have had huge positive impact on the participation rate, and have partially inverted previous tendency. Among women the participation rate has grown from 59% in 1980 to 75 in 2016. It has been driven by social changes and a structural transformation of the labor market toward services.[3]

In addition, over this period, the unemployment rate in France remained fairly stable (between 7.5% and 10%) and was negatively correlated with the business cycle with two recessions in 1993 and 2008 (see Figure 1.11 in appendix). Expansions thus appear unable to decrease unemployment below the 7.5% threshold, and recessions do not seem to have the same effect as, for instance, in the United States where the unemployment rate varies widely between a trough and a peak. Finally the share of part-time jobs has increased for both women and men over the period, despite with a persistent gap between genders. [4] Over our period of study the French labor market seems thus characterized by a steady state with high unemployment and moderate growth. The large share of the GDP dedicated to "social shock absorbers" tends to smooth both ups and downs.

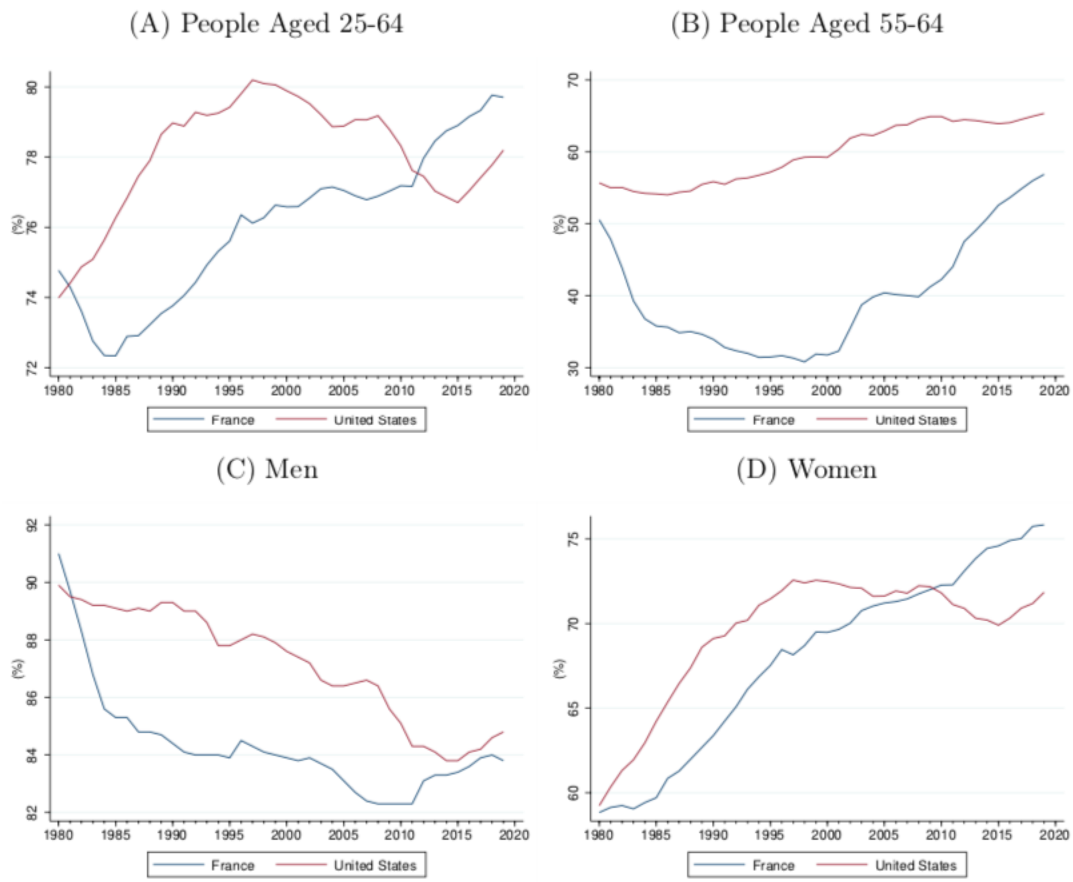
### 1.1.2 Labor market reforms

In this context, the socialist government, in place from 1997 to 2002, seeking to improve employees working conditions, has implemented two major labor market structural reforms: regulation of the workweek and reduction of labor cost at the minimum wage. These two policies were gradually implemented in France at the national level for most workers and all territories. They have had substantial impacts on labor market outcomes in the long-run. Conditional on being employed, these two reforms have affected individuals' labor earnings from both hours worked and hourly

[3] It was not until the law of July 13, 1965 that married women were allowed to work without needing their husband's approval. The participation of women in 1965 was around 45%. In addition higher productivity gains in agriculture and industry have reduced labor demand in these sectors, while the demand for services was increasing, without substantial productivity gains in this sector (see e.g. Husson (2018)). Section **??** studies further the occupational and industry difference composition between male and female workers.

[4] In 2016, 30.1% of women and 8.2% of men had a part-time job. It was respectively 23.4% and 4% in 1990. See "*Emploi, chomage, revenus du travail*", INSEE, 2020.

**Figure 1.1:** Participation Rate




*Note:* Figure 1.1 plots against time the participation rate for: (a) people aged between 25 and 64, (b) people aged between 55 and 64, (c) men aged 25-64, (d) women aged 25-64. Source: OECD.

wage.[5] Due to data constraint, our analysis period will restraint to the period: 1991-2016. We start by describing the changes on the workweek. Then, we explain those in the minimum wage policy. Additional information are provided in appendix 1.A.

**From 39 to 35 hours** First, we describe the reduction of the workweek. The path from the 39 to 35 hours workweek took place mostly between 1998 and 2001. Importantly, wage compensation schemes and wage moderation agreements were implemented at the same time so that monthly wages stayed constant in the short-term and did not increase too rapidly in the longer-run. Labor costs for low-wage workers did not increase too strongly thanks to the payroll tax exemptions that were expanded in those years. At the end of the 1990s, the socialist government decided to fulfill an electoral promise and implement the 35-hours workweek. Negotiations started within various industries and firms. In the end, the government enacted a law essentially forcing firms above 20 employees to come up with some agreement with their workers' unions or delegates.[6] Hence,

5 See for instance Kramarz and Philippon (2001) for a study of the impact of minimum wage introduction on the rise of unemployment in France. See Gautié (2020) for a literature review.

6 In addition, various incentives and subsidies were proposed at different moments in time. For instance, in June 1998, the so-called "Aubry I" laws gave establishments incentives to reduce their workweek and preserve employment in exchange for subsidies. The 2000 law, "Aubry II", offered payroll tax subsidies for all firms that decided to go to 35 hours per week.

among firms with more than 20 employees, at the beginning of the 21st century, various agreements prevailed.[7] Until 2005, when all minimum wages were unified, various levels of minimum wages coexisted depending on when the firm implemented the reduction of the workweek. As depicted in Figure 1.14 in appendix, the reduction of the workweek was followed by a strong increase in the minimum wage between 2002 and 2005, the highest over our period of interest. As can be seen on Figure 1.15 in appendix, hours worked in France were decreasing at a brisk rate in the years preceding the implementation of the 35 hours workweek. Then, they stabilized from 2002 on. In the US though, hours decreased too, at a lower rate and annual working time today is much higher than in France as it was in the 1980s. We will come back to this question of hours and to its impact on earnings inequality later in our analysis.

**Labor costs reduction** Second, we describe the minimum wage policy. The debates on minimum wages that took place in the 1990s in the United States tend to obscure why the French case was one of a dramatically high minimum wage. For employers at least, the minimum wage by itself is only one part of the story. What really matters is the total labor cost, i.e. the wage plus the payroll taxes. These payroll taxes have two components: one paid by the worker and the other paid by the firm. Until the beginning of the 1990s, France was characterized by both a high minimum wage and high labor costs at the bottom of the wage distribution.[8] As Figure 1.12 in appendix shows, the nominal and the real value of the minimum wage increased dramatically, by respectively 70% and 30%, between 1991 and 2016. However, as evidenced by Figure 1.13 in appendix, the total labor cost barely budged thanks to the very strong decrease in employers' contributions in the 1990s and the 2000s. These contributions are now virtually equal to zero. We will see in Section 1.4 that this strong increase in the minimum wage, especially since the middle of the 2000s, had a substantial impact on labor earnings inequality.

If the reduction of employers' labor cost at minimum wage should foster wage increase for low-paid workers, it could simultaneously promote less productive and more unstable jobs, implying more unemployment spell that would correspond to zero labor earnings over this spell (whenever the worker might received unemployment benefit). The reduction of the workweek could have the opposite effect, by favoring stable jobs, implying less unemployment spell. Nevertheless it could also mitigate wage growth and, by reducing productivity with higher hourly wage, it could slow down the climbing of the job ladders. Therefore, the incidence of the reforms on individuals' labor earnings is not straight forward.

### 1.1.3 Spatial Heterogeneity

At the end of 2018, France was shaken by a large-scale protest movement known as the "*Gilets jaunes*" (Yellow Vests). The movement initially brought together motorists angry about rising fuel prices and purchasing power. But it quickly turned into a general protest against government policy

[7] Some firms were still at 39 hours and had to pay overtime, others went to 35 hours between June 1998 and January 2000 and received incentives and subsidies. Firms also went to 35 hours after January 2000, receiving only the "structural" subsidies. Finally, remaining firms went to 35 hours and decided to receive no subsidies.

[8] It is still characterized by a very high minimum wage since the ratio of the net minimum wage to the median wage is equal to .63 in 2015, one of the highest in the OECD.

and inequality.[9] This movement stood out due to its local character and its nationwide coverage. The inequalities of opportunities between territories were put forward by the demonstrators, often coming from low-density areas. This uprising of remote territories with demands on inequality challenges the previous results of a convergence of labor earnings at the national level. Recent results in the Presidential elections also display sharp contrast in the vote shares for the different candidates between urban cores (that voted more Macron and Mélenchon) and low-density areas (that voted more for Le Pen), superposed with regional variations.

The making of the suburban metropolis find its roots in the rapid surge of car commute that allow workers to reside in different places than their workplace with potentially long distance to drive. It has resulted in the integration of low-density areas to the metropolitan local labor market, that previously excluded non urban residents with too long travel time.[10] On the other hand, jobs have clustered in urban business districts to benefit from agglomeration externalities, whether in city centers or their inner suburbs (like La Defense in Paris). Chapter 2 describes this making of the suburban metropolis, with the case of Paris in the second half of the XXst century.

We address the question of spatial heterogeneity in earnings changes in this chapter. In section 1.3, we describe the evolution and a share of workers at minimum wage by gender and territories. It reveals that women who reside in low-density area are much more likely to work at minimum wage. It allows to contextualize changes of the distribution of earnings across places in light of the labor market reforms. More precisely, we distinguish workers by their relative access to labor market opportunity with the distance of their residence to the urban core, where most of jobs locate. We describe our methodology with the data, section 4.1.

## 1.2 Data and Descriptive Statistics

### 1.2.1 Social Security Data

To study the dynamics of labor earnings at the individual level we use in what follows the Social Security data, so-called DADS ("*Declaration Annuelle des Donnees Sociales*"), that links Employer-Employee Data over the period 1991-2016. These administrative data are based on mandatory employer reports of the earnings and employment statue of each employee subject to French payroll taxes. It comprises all employer's and their (declared) employees. In each year $t$, the data comprise information on year $t-1$ and year $t$. Because of legal constraints the full panel version does not include all workers. The so-called panel combines a random sample (individuals born in October of an even year) from the DADS.[11] In addition, the panel can be matched to the *Echantillon Demographique Permanent* (EDP) a sample (individuals born the first 4 days of October) of the various censuses that allows to recover information on the level of education. Around 13% of the workers from the DADS can be matched with census data.

The sample we use covers private sector and public sector workers, excluding civil servants

[9] One of the main demands was the return of the wealth tax on top-earners ("*Impôt sur la fortune*"), in addition to more democracy and a lower cost of living.

[10] Therefore, in standard nomenclature of local labor market: urban areas account both the urban unit and the residential location it attracts, beyond the build-up continuity. We will use this nomenclature in what follows (see the data section, 4.1).

[11] The sample size was multiplied by two in 2002 by including individuals born in October of an odd year.

working for the central State.[12] In addition, the self-employed are not included in the sample when unemployment benefits are not available. According to the French Statistical Office (INSEE), wage employment represented 89.25% of total employment in 2019.[13] Finally, data availability precludes access to employees working outside of metropolitan France, employees working in the agricultural sector, or for private individuals. We also exclude apprentices, interns, and people working for the clergy.

The DADS data is aggregated at the job spell level (in an establishment in a given year for a given individual). Hence, our earnings measure is at this employment (job) spell level. For each individual, we define total earnings in year $t$ as the sum of earnings across all employment spells in that year. We measure earnings using their gross definition (net labor earnings inclusive of workers' mandatory social contributions at the exclusion of employers payroll taxes). This measure includes the sum of wages, over-time hours, paid leaves, bonuses, in-kind benefits, and several kinds of compensations (sickness, short-time work, severance payments, etc.). It does not include stock options. Earnings are expressed in 2018 euros deflated using the Consumer Price Index (CPI) computed by INSEE.

In line with other countries requirements for the Global Repository of Income Dynamics (GRID) project, we impose a minimum level of annual earnings for an individual to be included in the data. More precisely, an observation (i.e. an individual) must have earnings above the equivalent of 260 hours paid at the French minimum wage.[14] Appendix Table 1.2 depicts the annual minimum earnings threshold for the period of interest and Appendix Figure 1.18 plots the share of individuals with earnings below this minimum threshold. Every year, we exclude between 6 and 7% of the observations of the sample. Interestingly, this share is slightly decreasing over time while the income threshold is increasing due to the rising minimum wage.[15] This suggests that the decrease in inequality observed in Section 1.4 is not mechanically driven by the tightening of the income threshold.

We observe a strong increase in the share of excluded individuals in 1994. For unknown reasons, the share of jobs that cannot be matched with their individual identifier is higher in 1994, resulting in more individuals with earnings below the threshold. It is likely to explain the peculiar patterns observed at the bottom of the distribution for this year.[16] Therefore, we will not comment results for this year in what follows.

[12] The civil servants working for the State are not available in the comprehensive Social Security data before 2009 so we exclude them from the analysis. We always keep observations of civil servants working in hospitals and local governments.

[13] The share of wage employment stayed high over the period of interest. Between 1990 and 2002, it slightly increased from 87.6% to 91.2%. We then observe a small decrease since the beginning of the century.

[14] It corresponds approximately to a part-time job for one quarter.

[15] In addition, we computed the share of excluded observations for alternative minimum earnings thresholds. Using either 200 hours or 300 hours paid at the French minimum wage (baseline: 260 hours) would lead to exclude respectively 5.5% and 7.2% of the sample. These statistics suggest that the share of excluded observations does not vary much around the threshold.

[16] We also excluded around 3,000 observations with abnormally high gross earnings when compared to both earnings observed over the preceding years and current net earnings. Including these observations leads to erratic changes in the P99.9 and P99.99 between 2005 and 2007, which are inconsistent with the variations observed in the comprehensive DADS.

### 1.2.2 Measures of Earnings

The construction of the key statistics on earnings in this chapter follow the guidelines Global Repository of Income Dynamics. It accounts for the different data constraint of the project members to enhance cross country comparison over time.

In the remaining of the chapter, we will use the *raw real log earnings* as our main measure of individual labor earning. It is computed using the log of total annual worker compensation (provided in the DADS describe previously) deflated by the French national price index (CPI) with earnings in Euros.[17] Therefore, changes over time can be interpreted in percentage points. The main statistics are computed using two samples: the cross-sectional (CS) and the longitudinal (LX) samples. The CS sample includes all workers between 25 and 55 years old with raw real annual earnings above the minimum earnings threshold in the current year. The LX sample includes all workers of the CS sample who have a permanent earnings measure that is a measure of earnings for at least two years in a three years period. We denote the full time period available for the analysis (i.e. 1991-2016) by Tmax.

As for the other set of statistics used in this chapter, we define *residual earning* that takes into account the evolution of the age structure and educational attainment. To do so, we regress the raw real log earnings on a full set of age dummies (respectively age dummies and four education groups)[18] separately for men and women. We thus compute *one-year* (residual log) *earnings changes* for workers with earnings above the minimum threshold in years $t$ and $t+1$. We also compute five-year earnings changes to check the robustness of previous choice. Permanent earnings in year $t$ is then defined as the average earnings over a three years period, only for workers with earnings above the minimum threshold in year $t$ and at least one additional year.

### 1.2.3 Sample Selection

As highlighted in Gronau (1974) and Heckman (1974), a sample selection challenge occurs when studying wage. It corresponds to the fact that only observed wages are accounted for when studying wage inequality. Indeed, in the data only employed individuals are observed, so conventional measures of wage inequality may be biased. Thus wage inequality for those at work may provide a distorted picture of market-level wage. Nevertheless, the present chapter does not aim at studying hypothetical market-level wage but rather the sum of total labor earnings at the individual level within a year. We thus account for unemployment, not a source of bias but as a source of earnings risks (Pora and Wilner, 2020). Therefore, latent wage level is of no information about individual earnings when workers do not earn it effectively, event though it would represent opportunity costs.[19]

Arellano and Bonhomme (2017) provides an empirical exercise with the study of wage inequality in the UK in the last quarter of the twentieth century based on quantile selection models

[17] Conversions, when needed, are done according to the exchange rate at the introduction of the euro 1EUR =6,55957FRA

[18] We divide education into four groups: no diploma, less than high-school, high school, and some college. Education is only available for workers in the Panel DADS merged with the EDP. Appendix Table 1.3 describe the distribution of workers in the CS sample by age and education.

[19] Note however that we are not able to account for unemployment benefit, since they are not provided by the Unemployment benefit over our period of study. Therefore, our study will only labor earnings during employment spell.

**Table 1.1:** Earnings Distribution in France

(a) Men

| Year | P1 | P5 | P10 | P25 | P50 | P75 | P90 | P95 | P99 | P99.9 |
|---|---|---|---|---|---|---|---|---|---|---|
| 1995 | 2,511 | 5,171 | 9,098 | 18,373 | 25,225 | 34,873 | 51,158 | 67,447 | 115,424 | 227,547 |
| 2005 | 3,031 | 5,939 | 10,029 | 19,831 | 26,325 | 36,077 | 52,741 | 69,308 | 123,877 | 295,169 |
| 2015 | 3,208 | 6,000 | 9,757 | 20,079 | 27,745 | 38,345 | 55,990 | 72,719 | 132,730 | 332,115 |

(b) Women

| Year | P1 | P5 | P10 | P25 | P50 | P75 | P90 | P95 | P99 | P99.9 |
|---|---|---|---|---|---|---|---|---|---|---|
| 1995 | 2,271 | 3,668 | 5,841 | 12,335 | 20,247 | 28,025 | 37,057 | 43,846 | 67,395 | 119,759 |
| 2005 | 2,761 | 4,425 | 6,923 | 14,064 | 21,516 | 29,157 | 39,465 | 48,055 | 77,505 | 150,145 |
| 2015 | 3,055 | 5,170 | 8,012 | 15,714 | 23,556 | 31,525 | 42,884 | 53,644 | 89,075 | 186,786 |

*Note:* Table shows summary statistics for CS sample separately for (a) men and (b) women. *Dataset:* Panel DADS.

they develop. They find that correcting for selection into employment affects men wage at the bottom of the distribution (below the 30th percentile, for instance over 25 years the 10th percentile of men wage increased by 10% conditional on employment, while latent wages remained broadly flat), which is consistent with low-skilled men being progressively driven out of the labor market. Sample selection has smaller effect for women. A tentative explanation could be that for women non-economic factors play a bigger role in participation decision (suggesting also the strong reliance of the quantile selection model on specified individual characteristics that are not available in the French data). As a result, correcting for sample selection accentuates the decrease in the gender wage gap but almost only for low wage workers.

We focus on earnings rather than wages. We thus allow workers to combine several employment spell for different employers at different wages during a year, or to leave the labor market and come back. A particularly important gain with this approach is that it allows to account for unemployment risk (unemployment spell within the year). In addition, because we study the incidence of labor market reforms on workers earnings we are less interested in the latent-wage distribution than in the effective yearly earnings of workers.

### 1.2.4 Descriptive Statistics

Now that we have described the data we will used in this chapter, as well as its limitation in term of sample selection bias, we provide descriptive statistics of the French labor earnings distribution over our period of study.

Table 1.1 display the absolute value of the distribution of individuals' total labor earnings in France by gender, with panel (a) that corresponds to men and (b) to women. We first observe a huge difference between the bottom (first percentile, P1) and the top (last percentile, P99) of the distribution, with value 41 times higher for men and 29 times for women in 2015. earnings increase substantially all along the distribution. It is multiply by two between P1 and P5, as well as between

P95 and P99, at least for men in 2015. Nevertheless the specter of annual value is very large, going from less than 3,500€ at the P1 to more than 100,000€ at the P99. It corresponds to very different life condition and it suggests, at the bottom that additional source of income might compensate low labor earning, either unemployment and/or redistribution benefit (which we do not capture in our data). While it is not suggested with labor earnings values at the bottom, it is shown in the literature that top-earners also get a substantial share of income from other sources, like corporate ownership or financial benefits (Drechsel-Grau et al., 2022).

In this chapter we will stress on the difference between men and women inequality and earnings dynamics. The share of women in our sample is of 43.2% in 1995 and 47.8% in 2015, in line with both the increase share of female participation rate and the decrease share of male participation rate. Table 1.1 allows to provide first evidence on the difference between the distribution of earnings for male compared to women. At the bottom of the distribution, until P10, the two distribution are pretty similar for all years. The two start diverging at the first quartile (P25), with a difference of 4,365 between men and women. An explanation for this divergence is the higher share of women working part-time, implying lower earnings. In addition, women are also more likely to work at minimum wage level. Table 1.2 in appendix gives the value of the annualized minimum wage over years. It was 10,065€ in 1995, 14,833€ in 2005, and 17,490 in 2015, which is below than P25 for men at all years and around the P25 for women. The difference at this first quartile corresponds to a 28% higher wage for men in 2015, 18% at the median value (P50), 22% for the last quartile (P75) and it rises to 49% for the last percentile (P99). However, the differences were larger in 1995, with 49% at P25, 24% at P75, and 71% at P99, which suggest a catching-up of female workers over the two periods, we will see in what follow how labor market reforms have contributed to this trend, like the minimum wage that rise similarly to the women P25 and the workweek that induced higher hourly wage.

### 1.2.5 Characterization of Residential Locations

To address the spatial dimension of inequality, we propose to distinguish between places according to their distance to the urban core. That is to account for their level of urbanization, following the negative gradient of residents density from the city center, where a substantial share of jobs locate, to rural areas, where jobs have flowed away. Following international standard, the French Statistical Office (INSEE) define the *urban unit* as a municipality, or a group of municipalities, that has within its territory a built-up area of at least 2,000 inhabitants where no dwelling is separated from the nearest by more than 200 meters.[20] If the urban unit extends over only one single municipality and do not attract any other municipality, it is referred to as a *remote* city in what follows. When an urban unit is made up of several municipalities, it is referred to as a multi-municipal agglomeration. The municipalities that make up the agglomeration are labeled either as the *central* city or as the *suburbs*. If a municipality represents more than 50 percent of the population of the multi-municipal agglomeration, it is the only central city and the rest corresponds to its suburbs. Otherwise, all municipalities with a population greater than 50 percent of that of the most populous municipality, as well as the most populous municipality, are central municipalities. Urban municipalities that are

[20] Note that each municipality concerned should have more than half of its population in this built-up area. The urban units considered here have been established in 2010 with reference to the population known at the 2007 census.

not central then constitute the suburbs. The rest, municipalities that do not belong to an urban unit, even though they might be linked through commuting, are classified as *rural*. They corresponds to low-density areas without building continuity, where individuals are more likely to commute long distance to access business districts and jobs. These municipalities are similar to remote cities, except that they host less than 2,000 residents. Due to its prominent size and role in the French economy, we make a specific category for the Paris urban unit (hereafter Paris). Therefore, we end up with 5 categories of territories: Paris, Central, Suburbs, Remote, and Rural.

Social Security data (DADS) provides the municipalities of residence and work for all workers in each of their jobs during the year since 1993. Residential location and workplace do not always coincide (actually they are more likely not to coincide as a result of the specialization of places as residential or business district). Note that workers may have different jobs in different locations, the residential locations appears to be more stable than their workplace.[21] In addition, because we are interested in workers earnings changes, we choose to focus our study on workers residential location for two reasons. First, due to agglomeration economies, jobs locate mostly in central or, to a lesser extend, suburban municipalities, leaving aside a substantial portion of territories, especially rural areas, where people lives and access jobs by car commute. Therefore, it is important to assess how this divergence of places in terms of specialization between workplace and residence could be associated with diverging earnings trends. Second, the wage premium and the sorting of more productive firms and workers across locations has already been extensively discuss, showing increasing gains for larger urban unit over time.[22] To sum up, we choose to take an urban economics approach (disparities within urban unit according to the distance to the urban core) with a geographic scale (all municipalities) rather than the economic geography approach (disparities across urban unit according to their size). This methodological choice allows us to better describe spatial inequality (in term of labor earnings) in a context of the affirmation of the suburban metropolis, which labor market attraction goes far beyond its urban limits (see next chapter, 2, for details).

## 1.3 Labor Market Reforms, Hours Worked and Wages

### 1.3.1 Incidence of the labor market reforms between hours worked and wage

Before studying the evolution of the whole distribution of earning, we focus in this section on the evolution of hours worked and hourly wage by percentile of earning. We expect a rapid reduction in hours worked during workweek after the 35 hours workweek reform. We also expect an increase in hourly wage for earnings at the bottom of the distribution, as a result of the decrease of labor costs at the minimum wage.

Hours worked are available since 1993, we are thus able to contrast wages, hourly wages, and hours on the sub-period 1993-2016. Hours and hourly wages are computed for people with non-missing hours for all their jobs in the given year. About hours worked, we observe notable differences between men and women at the bottom of the earnings distribution (see 1.16 in appendix). Men with earnings close to the 25th percentile work full time (45% of a full time for

[21] If people change residential location within a year, we define residency as the one with the highest paying job.
[22] See e.g. Combes et al. (2008), Combes et al. (2012b), Gaubert (2018) and Delemotte et al. (2021) for a literature review.

**Figure 1.2:** Evolution of hours worked and hourly wages by earnings percentiles


*Note:* Using real raw log earnings and the CS+TMax sample, the figures plot against time the following variables: [**WARNING: panel identifier**] Men and Women median number of hours worked for the P10, P25, P50, P75, P90 of the earnings distribution and [**WARNING: panel identifier**] Men and Women median number of hourly wage for the P10, P25, P50, P75, P90 of the earnings distribution. All percentiles are normalized to 0 in the first available year. Shaded areas are recessions. *Dataset:* Panel DADS.

the 10th percentile) while it is only two third of a full time (respectively one third) for women at the 25th percentile and one third at the 10th percentile. Looking at the trends, men with low earnings tend to work less over time while hours worked are increasing for women at the 10th and 25th percentiles. Note that this variations of hours worked might either results from variation in per week and/or variations in the duration of employment spell. This decomposition between the intensive-extensive margins and the extensive-extensive margin goes beyond the scope of this chapter. The overall decrease in hours between 1999 and 2002 is extensively discussed below.

Turning to hourly wages, we observe that workers with earnings below, or at, the median earnings tend to earn similar hourly wages. This is especially the case for women where hourly wages are very close to the minimum wage (see Figure 1.17 in appendix). It results that women's differences in earnings at the bottom of the distribution are mainly driven by the number of hours worked, not by their hourly wage. The dispersion of hourly wage by income percentiles is higher for men, especially at the top of the earnings distribution, in line with what can be observed with annual earning.

We now discuss further the potential impact of the reduction of the working week. Figure 1.2 shows, on the left, the evolution of the median number of hours worked for various percentiles of the earnings distribution and, on the right, the evolution of the median hourly wage for the same percentiles. The median number of hours worked clearly display the impact of the reduction of the workweek for those percentiles, above the 25th percentile for men, only at and above median for women, that worked 39 hours in 1993 and moved to 35 hours from 2002 on. Contrasting men and women clearly shows that men with income below the median tend to work less hours in the aftermath of the workweek reduction, whereas it is the opposite for women with income below the median. This trend is in line with more men working part-time over years and the increasing participation of women into the labor market. The hourly wages increase very clearly and consistently across the earnings distribution over the period, corresponding to a mechanic increase in hourly wage to compensate for the reduction of hours worked a week. For men, the increase is smaller than for women and starts because of the workweek reduction whereas women increase their hourly wages before, at, and after the workweek reduction. This results in a 20% increase in hourly wages for men but in a 30% increase for women between 1993 and 2016. Interestingly, the 10th percentile for women closely follows the evolution of the minimum wage. We thus expect a moderate increase in labor earnings, especially for men. We also expect substantial differences between men and women due to both a higher increase in hourly wage and a smaller decrease in hours worked for women, especially at the bottom of the earnings distribution.

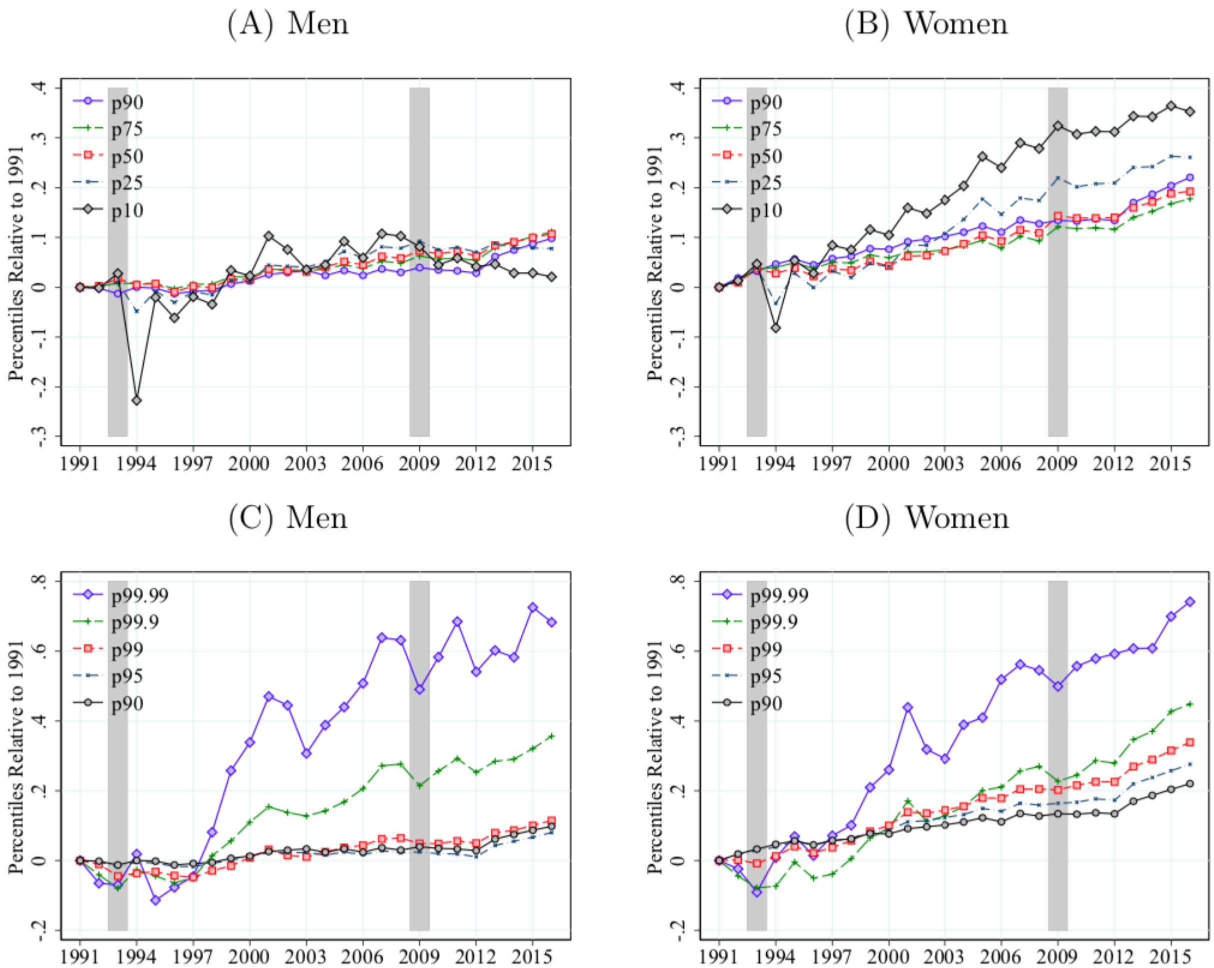
## 1.4 Evolution of the earnings Distribution

### 1.4.1 Evolution between men and women

We begin our analysis by studying the main features of the earnings distribution in France and its evolution over the 1991 to 2016 period. Figure 1.3 presents the evolution of the main percentiles that describe the distribution of earnings: P10, P25, P50, P75, P90. We also plot the evolution of percentiles and permiles at the top of the distribution: P95, P99, P99.9, P99.99. Values are normalized to 1991 level (see 1.1 for absolute values in 1995, 2005 and 2015). Vertical lines corresponds to recession periods. In line with the observation done in previous section, we expect almost flat earnings trends for men, while low-paid women should experienced a substantial increase in their total earning as the product of slightly more hours worked times higher hourly wage.

We see little wage growth for men in the core of the distribution, while there is clearly much higher growth for women at all percentiles, and most striking at the very bottom of the distribution. Real raw earnings grown by 36% for women at the 10th percentile of earnings and by 25% at the 25th percentile, compared to 2 and 9%, respectively, for men. This rapid increase at the bottom of the distribution coincides with the strong increase of the minimum wage that took place at the beginning of the 2000s. Indeed, women at the 10th and 25th percentiles of the earnings distribution have hourly wages close to the minimum wage and face a higher increase in their earnings during this period, while the rest of the distribution has grown much more slowly. Men at the bottom of the earnings distribution, who are more likely to work in production and to be paid more than the minimum wage, have faced a slight increase (P10) that doesn't last after the 2008 Financial crisis, which might have foster structural transformation in the production sector. In addition, the share of part-time workers among women is much more important than among men. Therefore, the increase in the hourly wage that resulted from the reduction of hours in the workweek contributed to this

**Figure 1.3:** Change of Percentiles of the Log Real Earnings Distribution




*Note:* Using real raw log earnings and the CS+TMax sample, the figure plots against time the following variables: (a) Men: P10, P25, P50, P75, P90 (b) Women: P10, P25, P50, P75, P90, (c) Men: P90, P95, P99, P99.9, P99.99, (d) Women: P90, P95, P99, P99.9, P99.99. All percentiles are normalized to 0 in the first available year. Shaded areas are recessions. *Dataset:* Panel DADS.

increase of earnings for low-paid women, as described in previous section.

Turning to the highest percentiles (above the 90th percentile, Figure 1.3 panel C and D). We observe a dramatic increase at the very top (above P99) and substantial year-to-year variations. These growth of very top percentiles is slightly higher for women than for men. Between 1991 and 2016, the top 1 permiles (0.1%) increased by 57% for women, 43% for men.This result can not directly be linked with the labor market reforms we focuses on in this chapter. It rather echoes the macro-labor literature on strong increasing wages for top-skill workers which can be linked with the rise of superstar-firms (e.g. Autor et al. 2020) or labor market polarization (e.g. Autor et al. 2006, Goos et al., 2009). In addition, since top-earners are also more mobile internationally, it is not surprising that this pattern is similar to many of the other countries part of the GRID project.

To sum-up the findings at the national level, variations in labor earnings inequality seem to be only mildly related to the business cycle and substantially driven by changes in labor market institutions, such as the strong increase in minimum wage over the period. Women, who are more likely have part time jobs and be paid at the minimum wage, have particularly benefited from these reforms, with a strong growth of hourly wage and a small increase in hours worked.

However, those findings does not account for heterogeneous access to the labor market opportunities that might vary with individual residential location and thus implies contrasted benefit from the national reforms. Indeed, LeBarbanchon et al. 2021 shows that the gender wage gap results in part in the different between gender of the willingness to commute long distance. The following sub-section address the incidence of the distance of residence to the urban core on previous results at the national level.

### 1.4.2 Evolution within Territories

In what follow we study how previous changes at the national level have translated in different places? Especially in a context of metropolization highlighted by the Gilets jaunes, with enlarging metropolis that concentrate jobs: does the reduction of inequalities at the national level hide an increase in disparities between territories?

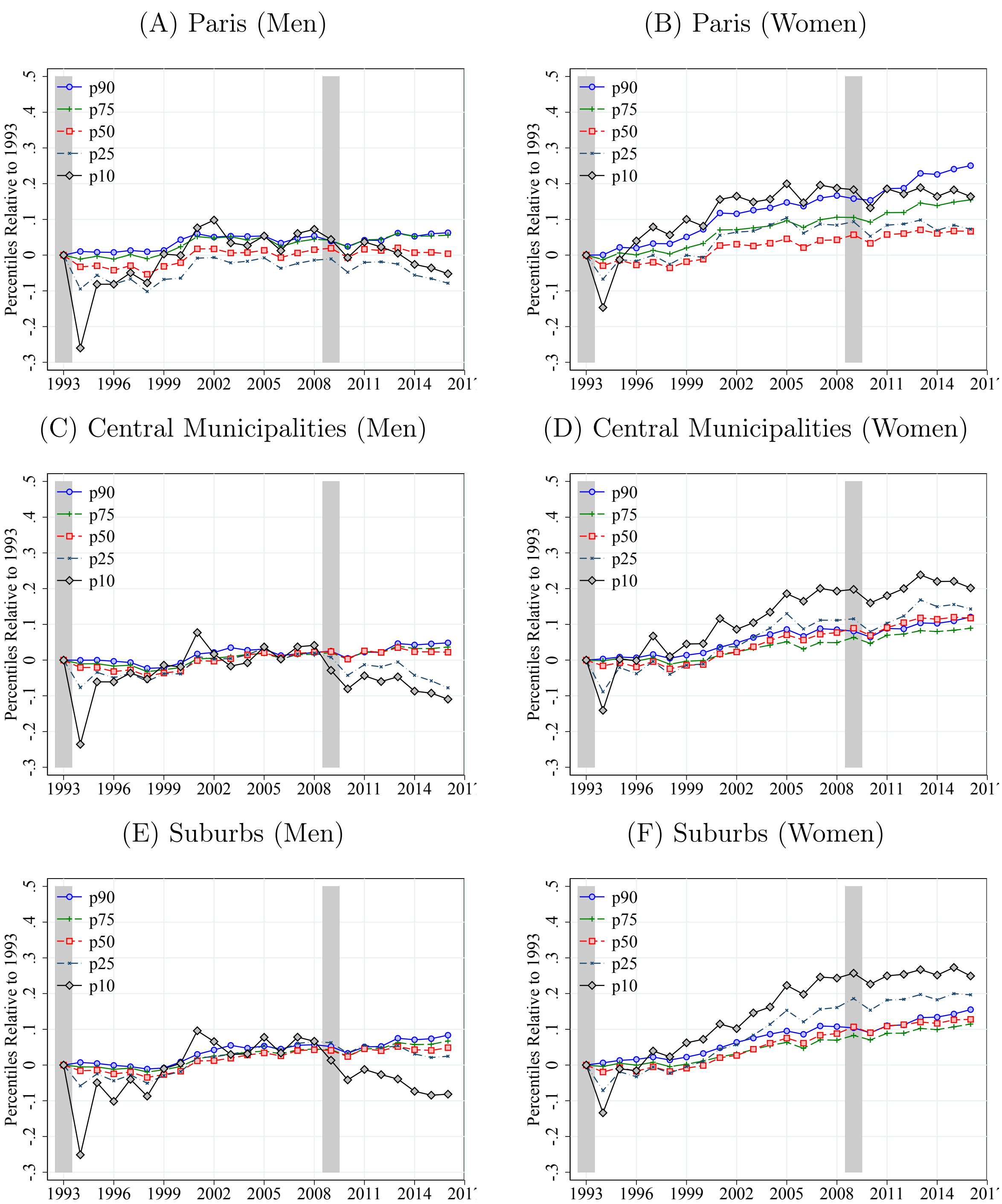
Figures 1.4 and 1.5 replicate the approach presented in the previously (Figure 1.3) by territories. It thus corresponds to the evolution of the earnings distribution *within* territories by gender. In all territories, we observe a pattern that is similar to the national level one: a stable earnings distribution for men and increasing earnings for women, especially at the bottom. However, we see large differences in magnitudes between territories, especially for women.

As for men, the earnings distribution is rather flat over our period of study, with low difference between territories. The 10th bottom percentile is the most volatile, since there is more margin of variation, especially in terms of hours worked, as discussed in the previous section. It display a phase of growth, which ended at different year across territories, around the turn of 2000 in urban municipalities (Paris, central and suburbs) while it last almost until the 2008 financial crisis in low density municipality (rural and remote). In Paris, we observe but a larger increase of the top earner (75th and 90th percentiles) in comparison to median earnings all over the period.

We now turn to the earnings distribution for women. For the bottom percentiles (especially the 10th lowest one), we observe an important wage growth for both rural and remote territories during the period of the major labor reforms (almost 40% earnings growth of the 10th percentile between 2000 and 2008). Strikingly, this increase is smaller for both suburban and central territories (respectively 30 and 25). In Paris, the earnings growth at the bottom is modest, this is likely due to the lower share of minimum wage women workers in the main city. This potential link between the share of minimum wage workers and earnings growth suggests that the policy of minimum wage increase during the 1990s and the 2000s, as well as previous results on the incidence of the workweek reduction, could explain a substantial share of earnings dynamics observed at the national level. At the other end of the earnings distribution, in rural and remote locations, highest percentiles (above the median) have had a smaller growth than the rest of the distribution (below median), while the trend between the 75th, the 90th, and the median are similar in central and suburban municipalities. At the opposite, the picture is much more different in Paris, where top female earners (90th percentile) have had the largest earnings growth (around 25% growth between 1991 and 2016). The 75th percentile has also had a higher growth than the median percentile. This suggest concentration of top-skilled jobs in Paris, associated with higher wage growth over the period. A direct consequence of these results is the decrease in inequality for women in all territories, except for Paris, at least until 2009. As seen above, the decrease is much larger in rural and remote territories.

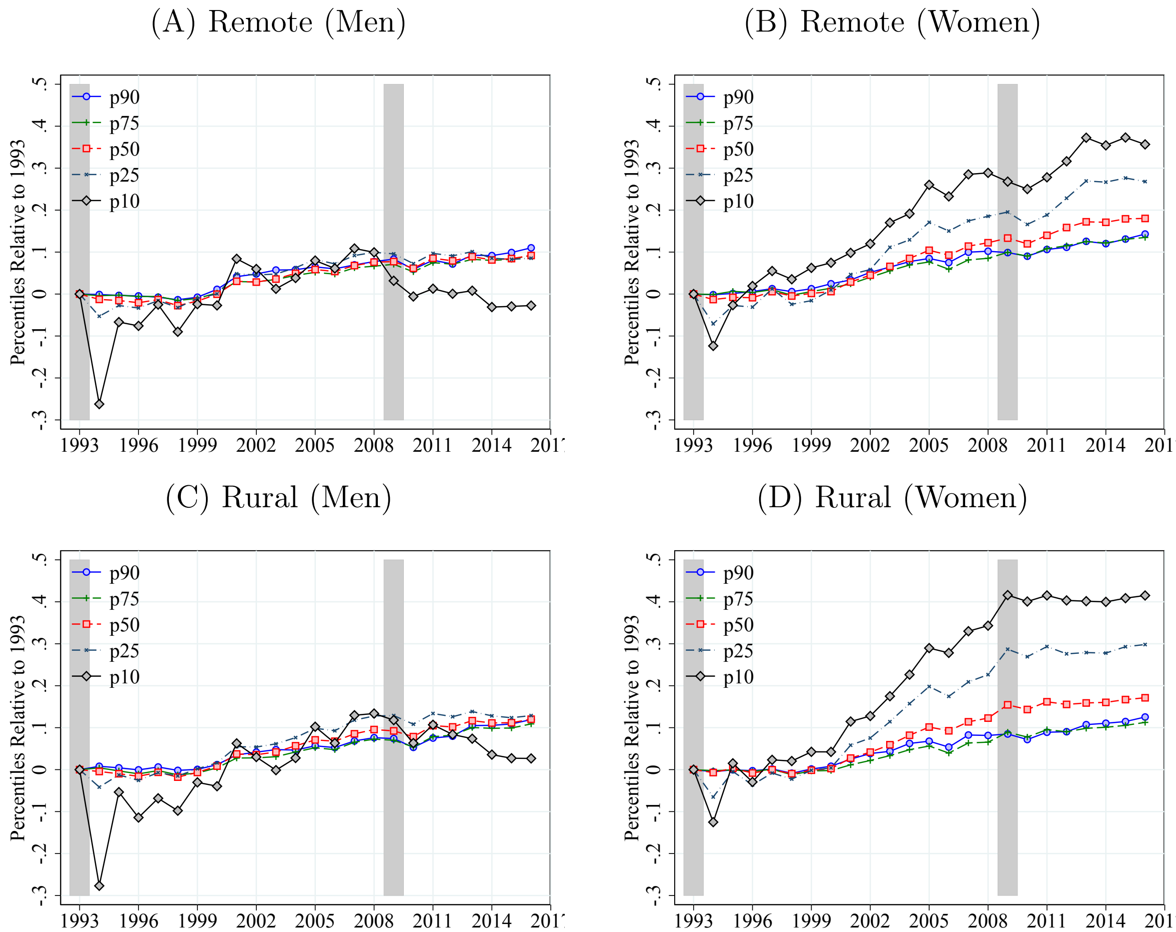
Turning to inequality levels, Figures 1.27 and 1.28 display the difference between the 90th and 10th percentiles. It is much higher in Paris than in other territories. It decreases following the urban gradient at similar speed for both men and female. However, the level of inequality was much lower for men in rural territories in the 1990s than it was for women. Interestingly, it suggests that the catching up of earnings of low-paid women is a convergence to standard labor conditions and thus inequality among women is similar to men after the implementation of the major labor market reforms. Conversely, inequality among women in Paris are lower than men over the whole period.

**Figure 1.4:** Change of Percentiles of the Log Real Earnings Distributions by Territory (1/2)




*Note:* Using real raw log earnings separately for men and women, Figure plots against time the P10, P25, P50, P75, P90 for: (a-b) the urban unit of Paris, (c-d) central municipalities and (e-f) the suburbs. All statistics are normalized to 0 in the first available year. Territories and Paris are defined using urban units. Shaded areas are recessions. *Dataset:* Panel DADS.

**Figure 1.5:** Change of Percentiles of the Log Real Earnings Distributions by Territory (2/2)




*Note:* Using real raw log earnings separately for men and women, Figure plots against time the P10, P25, P50, P75, P90 for: (a-b) remote municipalities, (c-d) rural municipalities. All statistics are normalized to 0 in the first available year. Territories and Paris are defined using urban units. Shaded areas are recessions. *Dataset:* Panel DADS.

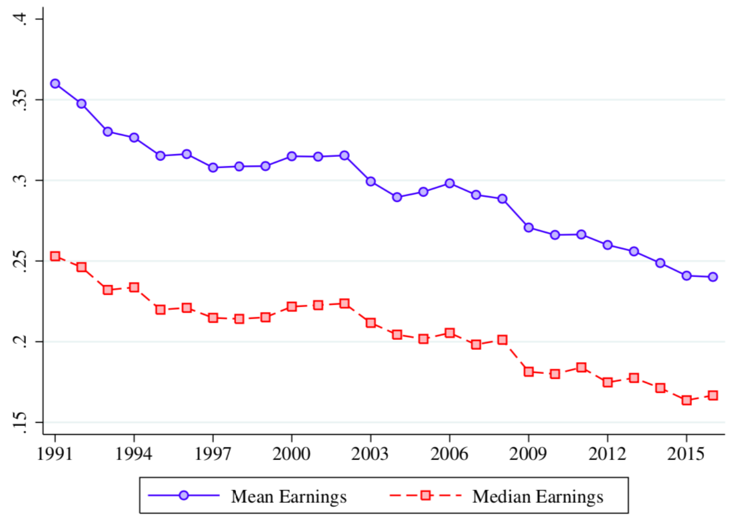
## 1.5 Evolution of Inequality

### 1.5.1 Inequality at the national level

The various dynamics of earnings percentiles described in the previous section highlight the contribution of the labor market reforms in increasing the earning of low-paid workers. It results in an overall decline in earnings inequality, except with the very-top of the distribution, and at least until the 2008 financial crisis. Most striking, in particular when compared to many other developed countries, the higher growth of the bottom percentiles entails a decrease in inequality for women over the period. This results in comparable levels of inequality for men and women at the end of the period. However, inequality tends to increase since the financial crisis, most particularly for men, again driven by the bottom of the earnings distribution -but with a negative growth in this case.

Interestingly, the level of labor income inequality is relatively low in France compared to other developed countries. The P90-10 differential of log labor earnings is on average 171 log points for men and 177 log points for women over our sample period, a level much lower than what is found

**Figure 1.6:** Unconditional Gender Pay Gap




*Note:* Using raw log earnings, Figure plots against years the difference between men and women mean (respectively median) earnings.
*Dataset:* Panel DADS.

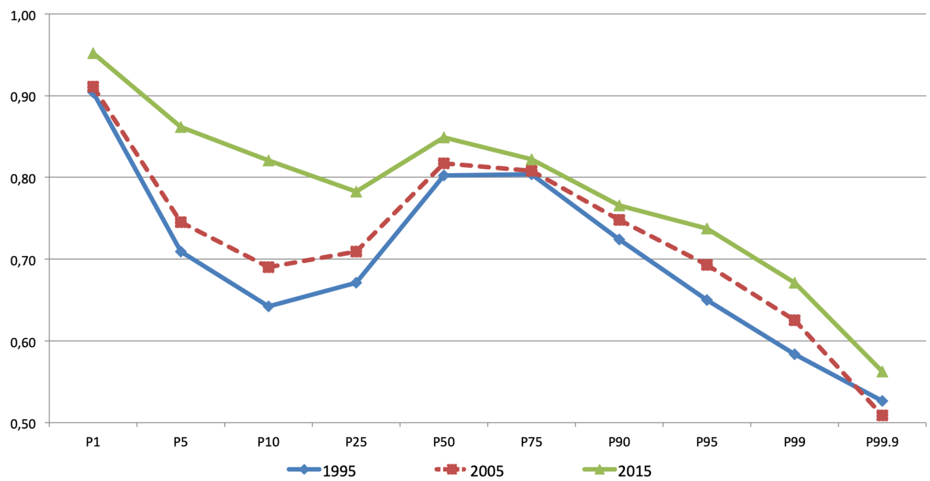
in the United States and comparable to that of Norway (see other papers in this issue McKinney et al., 2021; Halvorsen et al., 2021). It also appears to be close to that prevailing in a Gaussian distribution, especially for women (the P90-10 is close to 2.56 standard deviation). However, and in stark contrast with the Gaussian distribution, the distribution of earnings is not symmetric. It is negatively (left) skewed, in particular for women (i.e. P90-50 is much lower than P50-10), implying higher probability of large values at the top of the distribution, while the its bottom is truncated by our selection threshold.

### 1.5.2 Gender pay gap

While previous measures focused on *within*-gender inequality, we conclude this section by investigating the impact of heterogeneous distributional changes on inequality *between* genders. To do so, we compute the log differential between men and women earnings, using either mean or median earnings. Figures 1.6 plots the unconditional gender pay gap defined as the difference between men and women yearly labor earnings without taking into account neither workers' nor jobs' characteristics. First, we observe large differences in earnings between men and women. Mean and median earnings are respectively 43% and 29% higher for men than for women in 1991. We then observe a decrease of more than a third in the gender pay gap over the period of interest due to the higher growth for women at all percentiles described in Figure 1.3. Using residual earnings yields essentially identical trends in earnings inequality, when controlling either for age or age and education.

We provide further details of the earnings dynamics in the appendix, section 1.D. It essentially shows that the distribution of individuals' earning growth between two years does not follow a Gaussian distribution, as shown in previous work by Guvenen et al. (2021) or Pora and Wilner (2020). It implies that large negative shocks are more frequent than positive one (left-skewed) and than what the normal distribution would predict (excess kurtosis). In addition, one age-group seem to

**Figure 1.7:** Gender Distribution Gap

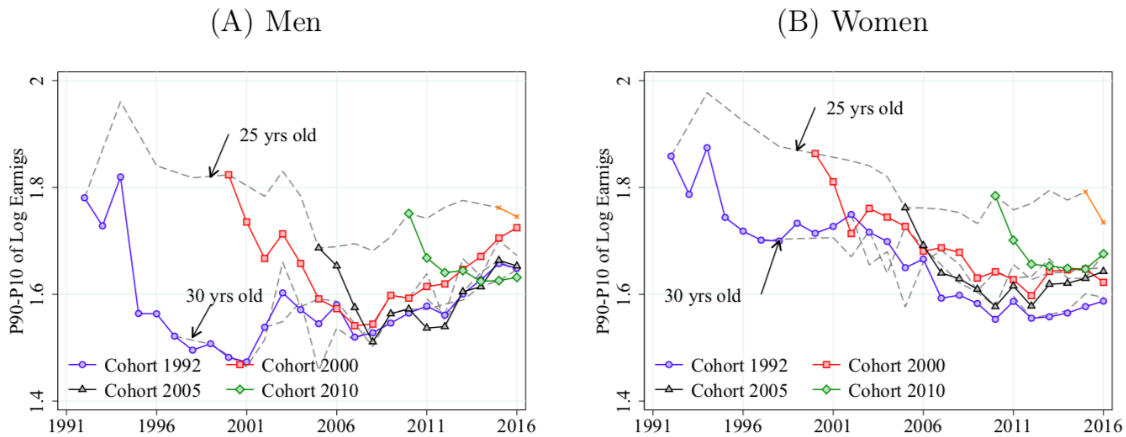



*Note:* Figure plots the difference between percentiles of the men and women earnings distribution in 1995, 20005, and 2015. P1 = P1-women / P1-men with absolute value describe in 1.1. *Dataset:* Panel DADS.

face higher earnings uncertainty: female workers aged between 25 and 34 years old. Indeed, the dispersion of earnings growth between one year and the other is much larger than the rest of the working population and for all earnings level. Furthermore, this uncertainty is related with negative shocks, since the distribution of one-year earning growth is left-skewed. Such pattern might be linked with maternity that constrain women, in France, to substantially reduced their number of hours worked.

Figure 1.7 display the difference between men and women earnings distribution in 1995, 20005, and 2015. That is the difference ration of absolute values percentile by percentile along the men and women earnings distribution (absolute values are described in Table 1.1). In 1995, the shape clearly describe a horizontal "S". Implying that workers at the very-bottom of the distribution have similar (and very low) earnings. In the next part of the earnings distribution, men earnings rise faster than for women due to the latter working more part-time. Working conditions tend to be closer around the two median of the men and women earnings distribution, with a ratio of women median-earners corresponding to 80% of male median-earners. Then the two distributions start diverging again and more so with earning. The female top-earners (P99.9) get hence half the return from work than male top-earners. Over time, it is stinking to observe that the bottom half of the two distribution get closer. The difference ratio between the women and men P25 went from 67.1% to 78.3% between 1995 and 2015. While the divergence at the top of distribution remains pretty unchanged. The lying "S" thus has became a straight line over time, implying that earnings inequality are higher between high-skilled workers than low-skilled one, in line with previous results in the literature. Blau and Kahn (2017) highlight in their literature review that "*although human-capital factors are now relatively unimportant in the aggregate, women's work force interruptions and shorter hours remain significant in high-skilled occupations, possibly due to compensating differentials*".

**Figure 1.8:** Life-Cycle Inequality Within Cohorts




*Note:* Using real raw log earnings and the CS+TMax sample, the Figure plots against time the following variables: (a) Men: P90-10 over the life cycle for all cohorts available, (b) Women: P90-10 over the life cycle for all cohorts available. The dash gray lines plot for each year the inequality between p90 and p10 for people aged 25, 30 and 35 this year. The four remaining lines plot the evolution of the P90-10 over time within cohorts by tracking them over time. *Dataset:* Panel DADS.

### 1.5.3 Inequality between cohorts

Previous results were cross-sectional (as evolution of earnings along repeated cross-section comparisons), they hence do not track individual earnings progressions over time. In particular cross-sectional results might hide substantial variations between cohorts, that is workers' age groups. We thus contrast previous results on earnings changes by cohorts. Indeed, yearly statistics over the whole income distribution can hide disparities due to age or cohort effect. Hence, trends described in the previous sub-sections can be due to variations in initial conditions at labor market entry and/or variations in earnings dispersion over the life cycle. We study now how inequality compares between cohorts and how it evolves over the life-cycle of various cohorts.

We study how the P90-10 differential evolved over the life cycle for four cohorts of workers, the ones who turn 25 at 1992, 2000, 2005, and 2010.[23] Figures 1.8 plot the P90-P10 inequality ratio when workers turn 25, 30 or 35 (dashed lines).[24] We observe that initial earnings inequality, at age 25, are very similar between men and women, contrary to what we describe in terms of inequality trend. Hence, this means that the distribution for women becomes asymmetric as they age. We observe a decrease in inequality over time for workers entering the labor market.

Figures 1.8 also plot the evolution of inequality within these cohorts, over time and by gender (solid lines). Inequality was much higher, within cohorts, at age 25 than at age 30 in the 1990s. This gap has been decreasing over time and has become small in the late 2000s, mainly because of the reduction in inequality at age 25 and a rapid convergence (around 5 years) of new cohort toward the average dispersion ratio, which suggest a reduction in the age premium over years. Similarly with other results of the chapter, within cohorts inequality has decreased until the financial crisis and increased since then for male workers, respectively stagnate for female workers. These results suggest that yearly statistics over the whole population hide substantial differences for younger

[23] It implies that people from the first cohort, aged 25 in 1992, are born in 1967. The others cohorts corresponds, respectively, to year of birth 1975, 1980, and 1985.

[24] Note that, in the figures, we start plotting the 30 years old inequality line when the first cohort (1992) reach this age, which corresponds to year 1998.

workers, which fades after age 30 for the French labor market.

Furthermore, the life-cycle inequality is decreasing for women, all over their working life, and U-shaped for men. It suggests that the inequality faced by women a the entry in the labor market fades over years, while they tend to increase for men after age 30. In addition, inequality at the entry in the labor market (age 25) has declined over time, in part attributable to the catching up of women yearly earnings at the bottom of the distribution. Inequality within cohorts has declined faster around the 2000s, when the 35hours workweek has been implemented (especially when looking at the cohorts 1992 the trend before this date and after 30 y.o. was flat). Conversely, for men, initial entry inequality seems to fall in the 2000s. Within cohorts, earnings inequality decreases after age 25, until most individuals of the cohorts gets a stable position, after 5 years. Then, after age 30, labor earnings inequality start increasing. With a steeper increase after the financial crisis.

### 1.5.4 Inequality between Territories

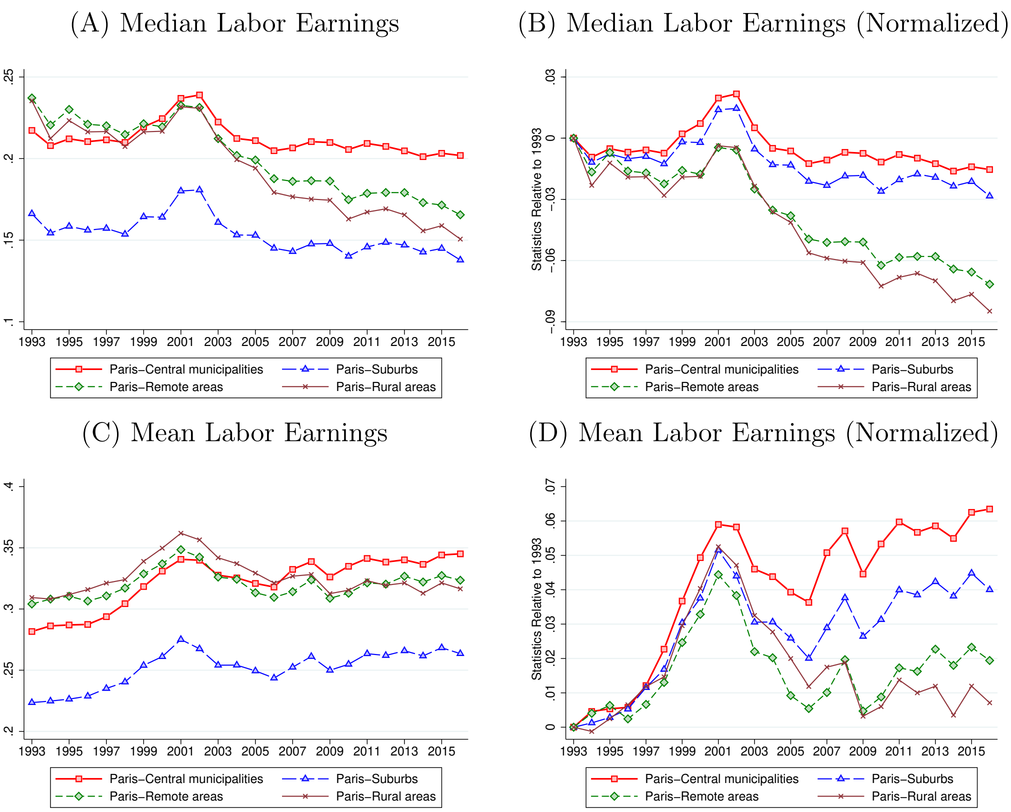
We now turn to the differences between territories. To this end, we compare labor earnings in the different territories in comparison with a benchmark, the Paris urban unit, the largest urban unit in France. Figure 1.9 displays the evolution of the difference between the median and the mean income by territories compared to Paris. First, we observe a huge gap in earnings between Paris and other territories, ranging between 22 to 36 higher mean labor earnings in Paris compared to the other territories, between 14 to 24% for the median. This gap is much lower between Paris and the suburban municipalities compared to difference between Paris and central, remote or rural municipalities.

Second, over the years, we observe a catching up in terms of median earnings of low-density municipalities compared to Paris, starting in the turn of 2000. With a reduction by more than 35% of the median earnings gap. By contrast the difference between Paris and other urban units (central or suburban metropolis) have slightly increased until 2002, before a small catching up, lasting at most 5 years. Interestingly, while the median earnings gap of low-density area was high and close to the one of central municipalities at the beginning of the period, at the end, in 2016, the gap of low-density with Paris is much smaller and closer to the one of suburban metropolis compared to central municipalities. These dynamics might be explained by residential mobility toward non-urban outer suburbs, favored by the reliance on individual vehicles (we discuss mobility later).

For instance, urban models with two transportation modes can also explain why low-paid workers reside in the city center, despite higher housing costs, because low-wage workers cannot afford the fixed-costs of buying a car but can choose to reside in smaller surface in the urban core (see LeRoy and Sonstelie, 1983, empirical evidence are provided in Glaeser et al., 2008b). This higher earnings inequality in central municipalities is also revealed in our statistics, see Figures 1.27 and 1.28 in appendix (discussed above).

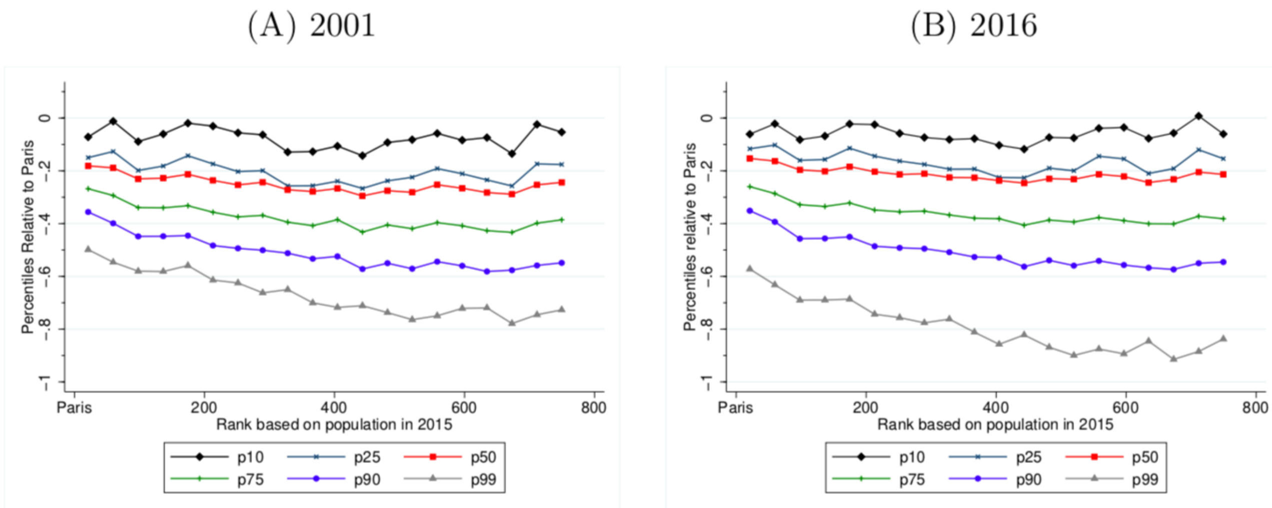
Finally, it is worth denoting that, for all territories, the mean earnings gap with Paris grows. This reflects high growth rates at the very top of the earnings distribution in Paris (the 99th percentile) relatively to other territories. This trend echoes the one observed in the U.S: a "great divergence" between metropolises (see e.g. Moretti 2012; Diamond 2016; Gaubert et al. 2021a), highlighted in the case of France by Davis et al. (2020). This trend is driven by top-skilled jobs concentrating in Paris, the largest French metropolis. To dig further this trend, we discuss in the

**Figure 1.9:** Mean and Median Earnings Gap with Paris




*Note:* Using the real raw log labor earnings of both men and women, Figure 17 plots against time the differential between: (a) median labor earnings in Paris and in other territories, (b) median labor earnings in Paris and in other territories normalized to 0 in 1993, (c) mean labor earnings in Paris and in other territories, (d) mean labor earnings in Paris and in other territories normalized to 0 in 1993. Territories and Paris are defined using urban units. *Dataset:* Panel DADS.

**Figure 1.10:** Between-Urban Areas Inequality




*Note:* Using raw log earnings for men and women, the Figure plots against city rank the 10th, 25th, 50th, 75th, 90th, 99th percentiles in: (a) 2001, (b) 2016. City rank is based on 2015 census population. Observations are the 759 French urban areas. Data are displayed using bin-scatter. *Dataset:* Comprehensive DADS.

following the evolution of the Urban wage premium over time and by earnings percentiles.

### 1.5.5 Distribution of the urban wage premium

The first part of this section has focus on a segregational approach of spatial disparities, with clustering of people according to their place of residence and their heterogeneous access to job opportunities. This choice is motivated by the rise of social protests and far-right voting at the periphery of the metropolis, where people need to commute large scale distance to access jobs. We will come back on this social protests later in the thesis (Chapter 3 and 4),. Indeed, increasing urban sprawl seems an important factor of heterogeneity among workers in their access to job opportunities, as we discuss in the Chapter 2.

However, an other well documented fact on spatial inequality is the difference of labor earnings between cities (or urban areas) according to their size. The urban wage premium is the strong link between job density and higher productivity (Combes et al., 2012b, Gaubert, 2018) and wages (Combes et al., 2008). We provide some new evidence in the following on the distribution of labor earnings across metropolis by plotting earnings percentiles in function of city size.

In this part, the scope of the study changes substantially and we now focus on *Urban Areas* that enlarges the perimeters of the urban unites to their catchment areas, that is including attracted municipality without build-up continuity with the urban unite. We are indeed interested here on city size rather than residential distance to the city center. An urban area is then a group of municipalities, in one piece and without an enclave, consisting of a central Urban unit with more than 10,000 jobs, and rural municipalities or urban units in which at least 40% of the resident population with a job works in the central Urban unit or in municipalities attracted by it.

Figure 1.10 ranks the 759 urban areas by size (population in 2015), Paris being the first (starting from the left hand-side).[25] Then, various percentiles of the real raw log-earnings distribution are represented (bin-scattered) along rank. All percentiles are displayed in difference from the Paris urban area. Figure on the left show the distribution for 2001 and, on the right, that for 2016. We use the comprehensive DADS (rather than the panel) in order to compute statistics for all urban areas.

In both 2001 and 2016, the bottom of the distribution (the 10th percentile) is almost indistinguishable from that of Paris, across all cities (ie urban areas); a reflection of the prevalence of the minimum wage for low-earnings workers. It echoes the leveling effect of minimum wage on spatial inequality, as seen in previous results. However, this result masks some "idiosyncratic" disparities: even though Lyon (2nd in size) or Bordeaux (5th) low-earners are not very different from the Paris ones, those in Marseilles (3rd) or Dunkirk (40th) are 17% less than in Paris at the 1st decile. This difference persist also at the first quartile. Above the median percentile, the gap is clearly increasing when city size decreases (moving to the right of the x-axis of each Figure). For instance, the worker at the 99th percentile in Lyon (2nd) makes already 31% less than the corresponding worker in Paris, and the gap increases to 44% in Marseilles or Bordeaux and to 61% in Dunkirk. Thus, on the one hand, the gap is clearly largest for the top of the distribution, and is larger in 2016 than in 2001. On the other hand, the gap at the bottom of the distribution is closing over the period (in Dunkirk, for instance, it decreases to 6%).[26]

In Appendix Figure 1.30, we examine the same question but controlling for demographic characteristics (gender, age, education), job characteristics (hours and occupation) and firms characteristics (4-digit industry code and firm size). We plot the decomposition by quartile, plus the 90th percentile. Adding controls strongly decreases the gaps between percentiles as well as the slopes across urban areas, making these lines almost flat. The gap becomes very small at the bottom of the distribution (less than 5%) while it stays significant for top percentiles. Controls matter more for top percentiles that for those at the bottom, a reflection of a larger homogeneity of observable characteristics of individuals in the latter. The characteristics that matter the most to explain disparities across urban areas are hours worked and occupations. Age, education, and gender explain a bunch of the top-earners urban wage premium, as well as it flatten it. However, a substantial contrast with Paris urban area persist (more than 20% at the 75th percentile and around 30% at the 90th percentile once we control for age, education and gender). Finally, industry and firm size only explain a small share of the remaining disparities between urban areas. Interestingly, around 20% of earnings inequality remains after all controls between Paris and other areas suggesting substantial productivity gains induced by agglomeration externalities and sorting (Gaubert, 2018) that do not benefit low-paid jobs. In addition labor earnings magnify productivity gains and hide a bunch of the negative externality from agglomeration, since the consumer price index is computed at the national level: workers have to pay higher housing costs and longer commuting time.

Moretti (2013) discusses the difference between nominal and real wage inequality. Using a city-specific CPI, it finds that real wage differences between high and low-skilled groups have grown significantly less than nominal differences (as in our approach). Changes in the geographical location of different skill groups are to a significant degree driven by city-specific shifts in relative

[25] Note that Paris urban area is larger than Paris urban units, since it contains additional urban unit and rural municipalities.

[26] The small differences in earnings at the bottom of the distribution may mask large differences in employment rates and job loss rates between areas (see e.g. Bilal, 2021).

demand (see also Diamond 2016 on this shift). Albouy (2008) discuss for instance the difference of cities' quality of life (especially determined by mild seasons, sunshine, hills, and coastal proximity in the US) and their impact on cities' cost of living and wage level. These findings imply that utility differences between skill groups might be smaller than what nominal wage difference predict. Unfortunately, we are not able to compute convincing CPI index either for territories or at the city level in France. However, Combes et al. (2019) shows that housing prices follow a decreasing gradient from the urban core to the countryside, suggesting that our results indicating lesser inequality in rural areas could be, in part, driven by lesser disparities in cost of living in low-density areas compared to urban core. In addition, Handbury (2021) focuses on consumption price and variety. It shows large differences in how wealthy and poor households perceive the choice sets available in wealthy and poor cities, suggesting that it might become tricky to properly account for the whole set of location-specific dimensions that affect individual welfare. Our focus in this chapter was much more modest and aim at providing new insight on how labor market reforms could affect individuals' labor earnings in different way by gender and place of residence. How it translates in term of utility and welfare is let for further research.[27]

## 1.6 Conclusion

Variations in labor earnings inequality over the last three decades seem to be only mildly related to the business cycle and substantially driven by changes in labor market institutions, such as the strong increase in minimum wage over the period. Women, who are more likely to have part-time jobs and to be paid at the minimum wage, have particularly benefited from these reforms which translated into a strong reduction of the gender pay gap, that hold true within cohorts. It also translate into a strong catching up of low-density areas in term of median earnings. As compared to other developed countries, the major labor market reforms that took place in France at the turn of the year 2000 could explain why the country is one of the only among developed countries with a clear decrease in labor earnings inequality over the last decades. It also reflects that national policies can help reduce spatial inequality.

Conversely it does not explain why France is politically divided, as expressed in the conflicted reactions to the rise of the Yellow Vests movement, described later in this thesis. The reduction in inequality holds for the core of the distribution, while at its very top we observe a dramatic increase in earnings that contrasts to the modest growth we see for other workers. In addition, these top-earners are mostly located in Paris, resulting in a substantial disparity in mean earnings between Paris and the rest of France. This finding is in line with increasing gains from agglomeration economies with earnings level. It also relates to the recent social protests, as it calls for new public policies so to reduce earnings inequalities between the top and the rest of the earnings distribution and better account for the spatial differences.

[27] For instance Berland and Etilé (2022) look at the heterogeneous incidence of income shocks on household budget and food consumption according to their income and place of residence.

# Bibliography


**Albouy, David**, “Are big cities bad places to live? Estimating quality of life across metropolitan areas,” Technical Report, National Bureau of Economic Research 2008. ↑64

**Arellano, Manuel and Stéphane Bonhomme**, “Quantile selection models with an application to understanding changes in wage inequality,” *Econometrica*, 2017, *85* (1), 1–28. ↑46

**Autor, David**, “Work of the Past, Work of the Future,” Technical Report, National Bureau of Economic Research 2019. ↑40

**_ , David Dorn, Lawrence F Katz, Christina Patterson, and John Van Reenen**, “The fall of the labor share and the rise of superstar firms,” *The Quarterly Journal of Economics*, 2020, *135* (2), 645–709. ↑39, ↑52

**Autor, David H, Lawrence F Katz, and Melissa S Kearney**, “The polarization of the US labor market,” *American economic review*, 2006, *96* (2), 189–194. ↑52

**Bell, Brian, Nicholas Bloom, and Jack Blundell**, “Income dynamics in the United Kingdom 1975-2020,” Technical Report, GRID Working Paper 2022. ↑39

**Berland, Ondine and Fabrice Etilé**, “The Effects of Gasoline Prices on Food Purchases: Quantity, Quality and Heterogeneity,” Technical Report, Mimeograph - AgroParisTech 2022. ↑64

**Bilal, Adrien**, “The geography of unemployment,” Technical Report 29269, National Bureau of Economic Research 2021. ↑63

**Blau, Francine D. and Lawrence M. Kahn**, “The Gender Wage Gap: Extent, Trends, and Explanations,” *Journal of Economic Literature*, 2017, *55* (3), 789–865. ↑40, ↑58

**Boyer, Pierre C, Thomas Delemotte, Germain Gauthier, Vincent Rollet, and Benoît Schmutz**, “Mobilization without Consolidation: Social Media and the Yellow Vests Protests,” *Mimeograph - CREST at Ecole polytechnique and Ensae*, 2022.

**Bozio, Antoine, Bertrand Garbinti, Jonathan Goupille-Lebret, Malka Guillot, and Thomas Piketty**, “Predistribution vs. Redistribution: Evidence from France and the US,” 2020. ↑39

**_ , Thomas Breda, and Malka Guillot**, “The Contribution of Payroll Taxation to Wage Inequality in France,” 2020. ↑39

**Charnoz, Pauline, Elise Coudin, Mathilde Gaini et al.**, “Forty Years of Decreasing Wage Inequality in France: The Role of Supply and Hidden Skill-Biased Technical Change,” Technical Report 2014. ↑39

**Combes, Pierre-Philippe, Gilles Duranton, and Laurent Gobillon**, “Spatial wage disparities: Sorting matters!,” *Journal of urban economics*, 2008, *63* (2), 723–742. ↑49, ↑62

**_ , _ , and _** , “The costs of agglomeration: House and land prices in French cities,” *The Review of Economic Studies*, 2019, *86* (4), 1556–1589. ↑64

— , — , — , **and Sébastien Roux**, "Sorting and local wage and skill distributions in France," *Regional Science and Urban Economics*, 2012, *42* (6), 913–930. ↑40

— , — , — , **Diego Puga, and Sébastien Roux**, "The productivity advantages of large cities: Distinguishing agglomeration from firm selection," *Econometrica*, 2012, *80* (6), 2543–2594. ↑40, ↑49, ↑62

**Couture, Victor, Cecile Gaubert, Jessie Handbury, and Erik Hurst**, "Income growth and the distributional effects of urban spatial sorting," Technical Report, National Bureau of Economic Research 2020. ↑40

**Davis, Donald R, Eric Mengus, and Tomasz K Michalski**, "Labor market polarization and the great divergence: Theory and evidence," Technical Report, National Bureau of Economic Research 2020. ↑60

**Delemotte, Thomas and Corentin Trevien**, "The Making of the Suburban Metropolis: Evidence from Paris (1960-2020).," *Mimeograph*, 2021.

— , **Francis Kramarz, and Benoît Schmutz**, *L'emploi et le territoire*, Presses de Sciences Po, 2021. ↑49

**Diamond, Rebecca**, "The determinants and welfare implications of US workers' diverging location choices by skill: 1980-2000," *American Economic Review*, 2016, *106* (3), 479–524. ↑60, ↑64

**Drechsel-Grau, Moritz, Andreas Peichl, Johannes F Schmieder, Kai D Schmid, Hannes Walz, and Stefanie Wolter**, "Inequality and income dynamics in Germany," Technical Report, Global Repository of Income Dynamics (GRID) Working Paper 2022. ↑37, ↑39, ↑48

**Duranton, Gilles and Diego Puga**, "Micro-foundations of urban agglomeration economies," in "Handbook of regional and urban economics," Vol. 4, Elsevier, 2004, pp. 2063–2117. ↑40

**Garbinti, Bertrand, Jonathan Goupille-Lebret, and Thomas Piketty**, "Income inequality in France, 1900–2014: evidence from distributional national accounts (DINA)," *Journal of Public Economics*, 2018, *162*, 63–77. ↑39

**Gaubert, Cecile**, "Firm sorting and agglomeration," *American Economic Review*, 2018, *108* (11), 3117–53. ↑49, ↑62, ↑63

— , **Patrick Kline, Damián Vergara, and Danny Yagan**, "Trends in us spatial inequality: Concentrating affluence and a democratization of poverty," in "AEA Papers and Proceedings," Vol. 111 2021, pp. 520–25. ↑39, ↑60

— , — , **Damian Vergara, and Danny Yagan**, "Trends in us spatial inequality: Concentrating affluence and a democratization of poverty," in "AEA Papers and Proceedings," Vol. 111 2021, pp. 520–25. ↑40

— , **Patrick M Kline, and Danny Yagan**, "Place-based redistribution," Technical Report, National Bureau of Economic Research 2021. ↑40

**Gautié, Jérôme**, *Le salaire minimum et l'emploi*, Presses de Sciences Po, Sécuriser l'emploi, 2020. ↑42

**Glaeser, Edward L, Jed Kolko, and Albert Saiz**, "Consumer city," *Journal of economic geography*, 2001, *1* (1), 27–50. ↑40

— **, Matthew E Kahn, and Jordan Rappaport**, "Why do the poor live in cities? The role of public transportation," *Journal of urban Economics*, 2008, *63* (1), 1–24. ↑40

— **,** — **, and** — , "Why do the poor live in cities? The role of public transportation," *Journal of urban Economics*, 2008, *63* (1), 1–24. ↑60

**Godechot, Olivier**, "Is finance responsible for the rise in wage inequality in France?," *Socio-Economic Review*, 2012, *10* (3), 447–470. ↑39

**Goldin, Claudia**, "A grand gender convergence: Its last chapter," *American Economic Review*, 2014, *104* (4), 1091–1119. ↑40

**Goos, Maarten, Alan Manning, and Anna Salomons**, "Job polarization in Europe," *American economic review*, 2009, *99* (2), 58–63. ↑52

**Gronau, Reuben**, "Wage comparisons–A selectivity bias," *Journal of political Economy*, 1974, *82* (6), 1119–1143. ↑46

**Guvenen, Fatih, Fatih Karahan, Serdar Ozkan, and Jae Song**, "What do data on millions of US workers reveal about life-cycle earnings risk?," 2021, *89*, 2303–2339. ↑39, ↑57

**Halvorsen, Elin, Serdar Ozkan, and Sergio Salgado**, "Earnings dynamics and its intergenerational transmission: Evidence from Norway," *GRID Working Paper*, 2021. ↑57

**Handbury, Jessie**, "Are Poor Cities Cheap for Everyone? Non-Homotheticity and the Cost of Living Across US Cities," *Econometrica*, 2021, *89* (6), 2679–2715. ↑64

**Heckman, James**, "Shadow prices, market wages, and labor supply," *Econometrica: journal of the econometric society*, 1974, pp. 679–694. ↑46

**Hoffmann, Eran B, Davide Malacrino, and Luigi Pistaferri**, "Labor Reforms and Earnings Dynamics: The Italian Case," Technical Report, GRID Working Paper 2022. ↑39

**Husson, Michel**, "L'emploi des femmes en France depuis 1960," *Document de travail, Institut de Recherches Économiques et Sociales*, 2018, *03.* ↑41

**Kleven, Henrik, Camille Landais, and Jakob Egholt Søgaard**, "Children and Gender Inequality: Evidence from Denmark," *American Economic Journal: Applied Economics*, 2019, *11* (4), 181–209. ↑40

— **,** — **, Johanna Posch, Andreas Steinhauer, and Josef Zweimuller**, "Child penalties across countries: Evidence and explanations," in "AEA Papers and Proceedings," Vol. 109 2019, pp. 122–26. ↑40

**Kramarz, Francis and Thomas Philippon**, "The impact of differential payroll tax subsidies on minimum wage employment," *Journal of Public Economics*, 2001, *82* (1), 115–146. ↑39, ↑42

**LeBarbanchon, Thomas, Roland Rathelot, and Alexandra Roulet**, "Gender differences in job search: Trading off commute against wage," *The Quarterly Journal of Economics*, 2021, *136* (1), 381–426. ↑40, ↑53

**LeRoy, Stephen F and Jon Sonstelie**, "Paradise lost and regained: Transportation innovation, income, and residential location," *Journal of Urban Economics*, 1983, *13* (1), 67–89. ↑60

**McKinney, Kevin L, John M Abowd, and Hubert P Janicki**, "US long-term earnings outcomes by sex, race, ethnicity, and place of birth," *GRID Working Paper*, 2021. ↑57

**Moretti, Enrico**, *The new geography of jobs*, Houghton Mifflin Harcourt, 2012. ↑60

— , "Real wage inequality," *American Economic Journal: Applied Economics*, 2013, *5* (1), 65–103. ↑63

**Piketty, Thomas**, "Income inequality in France, 1901–1998," *Journal of political economy*, 2003, *111* (5), 1004–1042. ↑39

**Pora, Pierre and Lionel Wilner**, "A decomposition of labor earnings growth: Recovering Gaussianity?," *Labour Economics*, 2020, *63*, 101807. ↑39, ↑46, ↑57

**Verdugo, Gregory**, "The great compression of the French wage structure, 1969–2008," *Labour Economics*, 2014, *28*, 131–144. ↑39

## 1.A French context

### Labor contracts

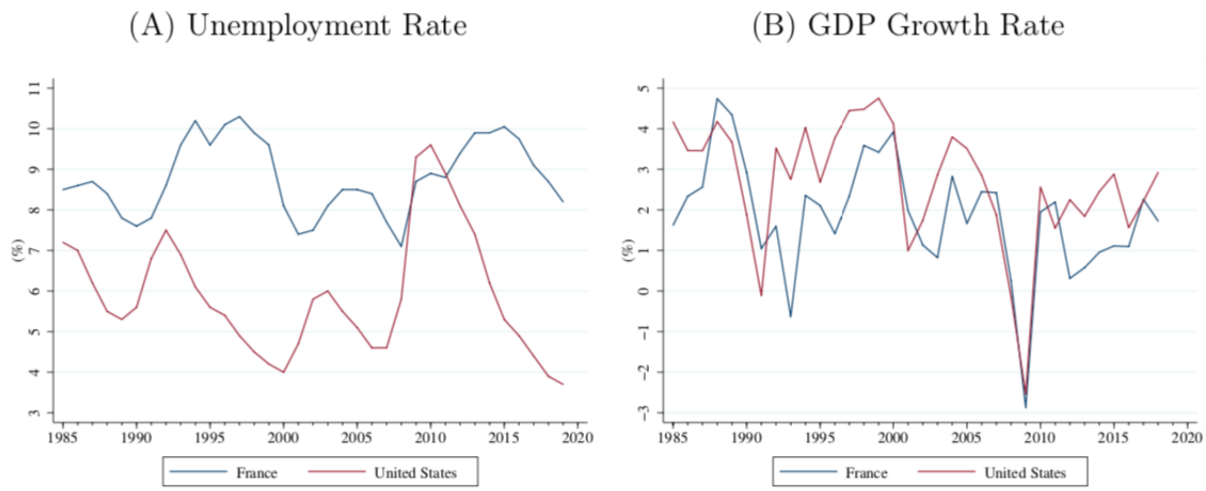
As in multiple other European countries, there are two main types of labor contracts : permanent and temporary. Most hires are made under the latter when most positions are held under the former. In 2017, 88% of the wage workers have a permanent contract while 87% of the new hires are made under temporary contracts. This rate has steadily increased since 1993 when it represented 76% of the new hires. The duration of these contracts tend to be short and the use of extremely short term contracts has strongly increased over the past decades. In 2017, almost one third of the temporary contracts last only one day.

### Unemployment Rate and Business Cycle

Figure 1.11 presents, on the left (A), the unemployment rate in France and in the United States over the period 1985-2019 and, on the right (B), the GDP growth rate for the same countries. As can be seen from the Figure, the unemployment rate in France has been staying consistently between 7.5% and 10% over the last 30 years (with one exception, 7%, in 2000). Expansions appear unable to decrease unemployment below this point, and recessions do not seem to have the same effect as, for instance, in the United States where unemployment between a trough and a peak can increase by 5 points when in France the increase is 2.5 points. GDP growth rates, however, vary mostly in sync (with that of France being lower by one point during expansions). During our analysis period, France has witnessed two recessions, in 1993 and 2008 but one expansion, around 2000 when the US had more expansion years.

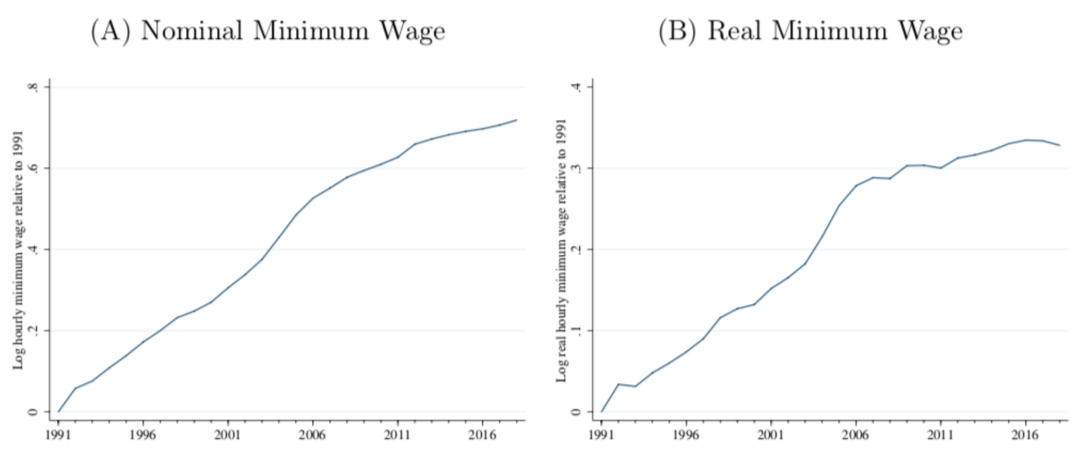
### Minimum Wage and Labor Cost

**Figure 1.11:** The French Business Cycle



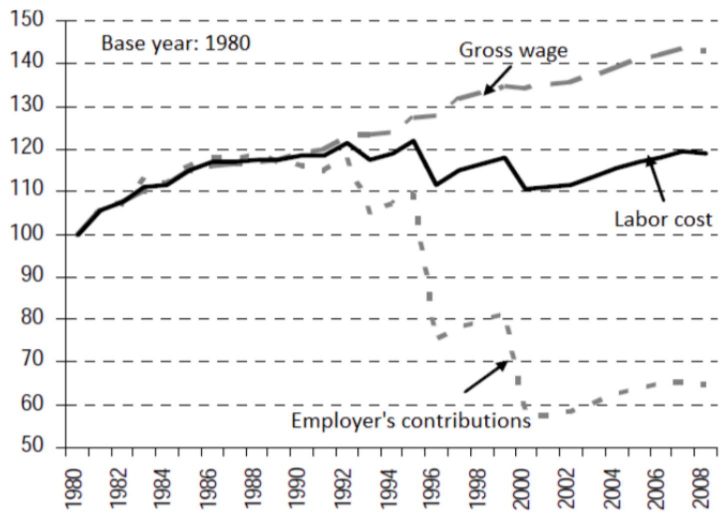

*Note:* Figure 1 plots against time: (a) the unemployment rate, (b) the growth rate of the GDP, for France and the United States. Source: INSEE and BLS for the unemployment rates. The World Bank for the GDP.

**Figure 1.12:** Nominal and Real Minimum Wage Over Time




*Note:* Figure plots against time: (a) the nominal hourly minimum wage, (b) the real hourly minimum wage in France. All statistics are normalized to 0 in 1991.
*Source:* INSEE.

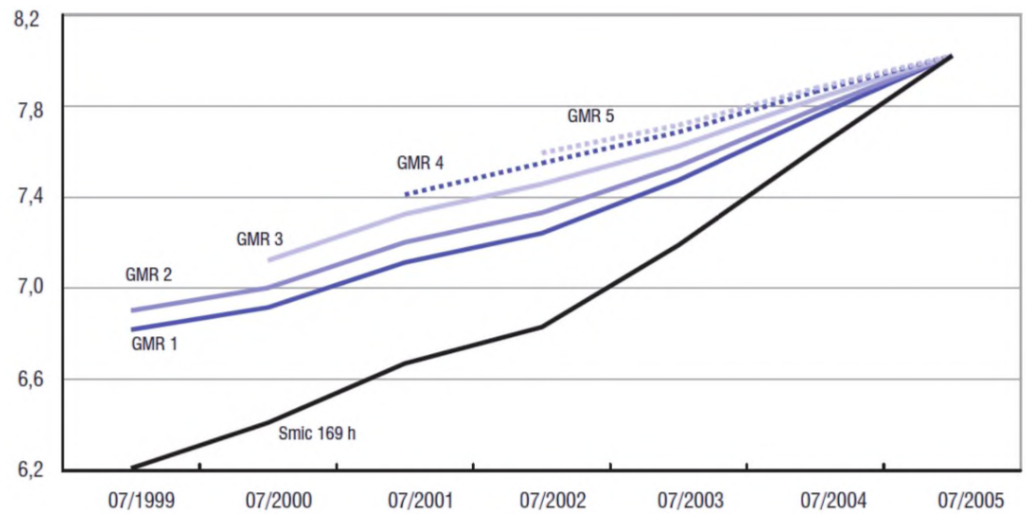
**Figure 1.13:** Labor Cost at Minimum Wage




*Note:* The labor cost is the sum of the gross wage and of the employer's contributions. All statistics are normalized to 100 in 1980.
*Source: "Rapport à la Commission des comptes de la sécurité sociale de juin 2009"*.

### Reducing Workweek Hours

**Figure 1.14:** The Several Minimum Wages in the 2000s

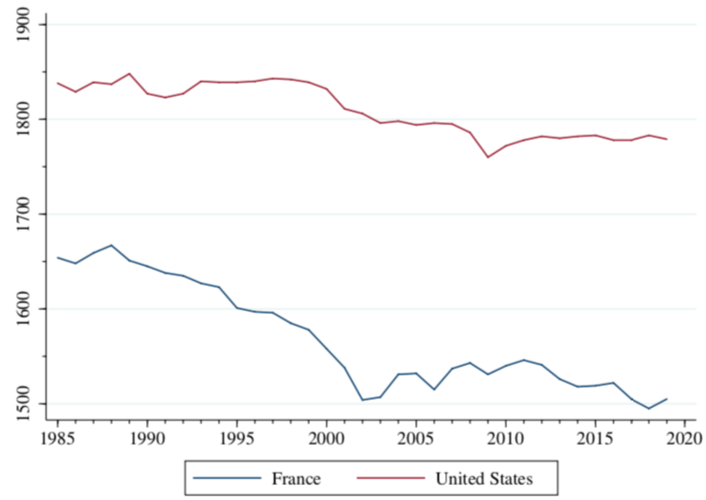



*Note:* Figure plots against time the five hourly minimum wages (GMR) for workers working 35 hours a week and the hourly minimum wage for workers working 39 hours a week (Smic 169h). Values are expressed in euros. GMR stands for "*Garantie Mensuelle de Rémunération*" (Monthly wage garanty). GMR 1-5 are applicable to firms which started reducing their worker's workweek respectively between: (1) 06/1998-06/1999 (2) 07/1999-06/2000 (3) 07/2000-06/2001 (4) 07/2001-06/2002 (5) in 07/2002.
*Source:* Malik Koubi and Bertrand Lhommeau, "Les salaires en France", 2007, Ministry of Labor.

At the end of the 1990s, the Jospin government, with Martine Aubry as Minister of Labor, decided to fulfill an electoral promise and to go to 35 hours. Discussions between the government, which included the green party, and business unions were tense. Negotiations started within various industries and firms. But, at some point, Martine Aubry enacted a law essentially forcing firms above 20 employees to come up with some agreement with their workers' unions or delegates. In addition, various incentives and subsidies were proposed at different moments in time. For instance, in June 1998, the so-called Aubry I laws gave establishments incentives to reduce their workweek and create or preserve employment in exchange for large subsidies. In order to receive these subsidies, firms had to reduce hours by at least 10% in order to attain an average weekly duration of 35 hours. In such a case, employment creation had to amount to 6% of total employment. A "defensive" aspect also allowed firms to receive subsidies to avoid economic separations or collective dismissals. The 2000 law, Aubry II, offered payroll tax subsidies for all firms that decided to go to 35 hours per week. Hence, among firms with more than 20 employees, at the beginning of the 21st century, various agreements prevailed. Some firms were still at 39 hours and had to pay overtime, others went to 35 hours between June 1998 and January 2000 and received incentives and subsidies, others refused the incentives (but received some "structural" subsidies) even though they went to 35 at similar dates (the so-called Aubry II forerunners). Firms also went to 35 hours after January 2000, receiving only the "structural" subsidies. Finally, remaining firms went to 35 hours and decided to receive no subsidies.

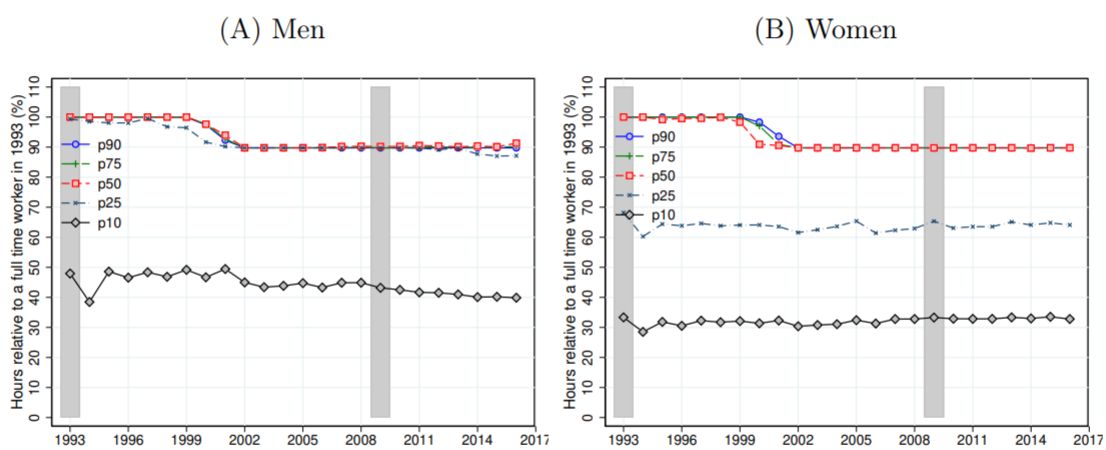
Figures 1.16 and 1.17 plot respectively the median annual number of hours worked, as a share of a full time job, and the median hourly wage for various percentiles of the earnings distribution.

**Figure 1.15:** Average Annual Working Time



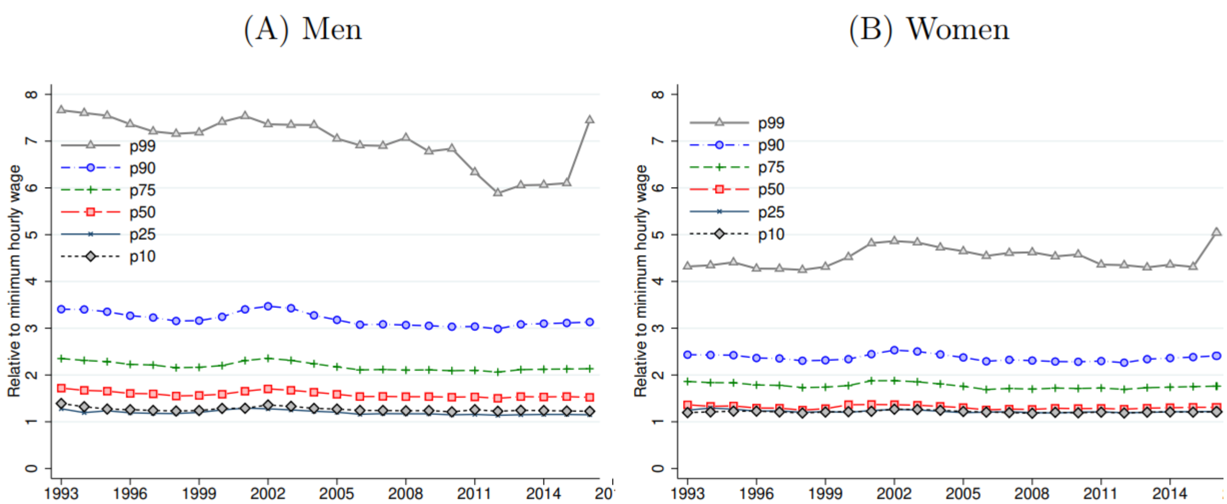

*Note:* Figure plots against time the five hourly minimum wages (GMR) for workers working 35 hours a week and the hourly minimum wage for workers working 39 hours a week (Smic 169h). Values are expressed in euros. GMR stands for "*Garantie Mensuelle de Rémunération*" (Monthly wage garanty). GMR 1-5 are applicable to firms which started reducing their worker's workweek respectively between: (1) 06/1998-06/1999 (2) 07/1999-06/2000 (3) 07/2000-06/2001 (4) 07/2001-06/2002 (5) in 07/2002.
*Source:* Malik Koubi and Bertrand Lhommeau, "Les salaires en France", 2007, Ministry of Labor.

**Figure 1.16:** Annual Working Time By Earning Percentile




*Note:* Figure plots against time, for 5 percentiles of the real raw earnings distribution, the following variables: (a) Men: median annual number of hours worked, (b) Women: median annual number of hours worked. For each of the 5 percentiles, we compute and plot the median number of hours worked as a share of a full time job in 1993.
*Dataset:* Panel DADS.

**Figure 1.17:** Hourly Wage By Earning Percentile

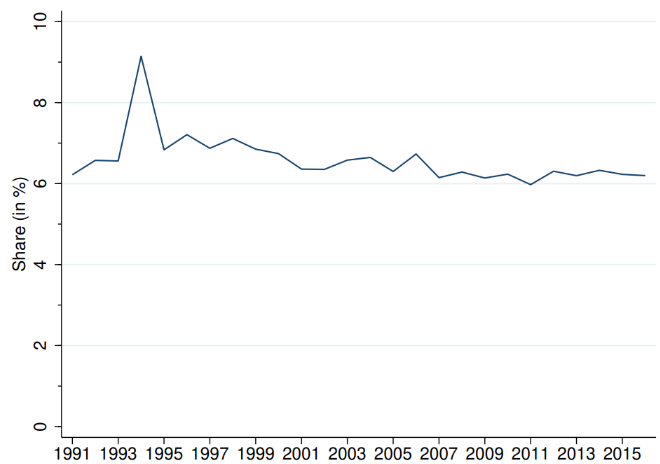



*Note:* Figure plots against time, for 5 percentiles of the real raw earnings distribution, the following variables: (a) Men: median hourly wage relative to the French minimum wage, (b) Women: median hourly wage relative to the French minimum wage. For each of the 5 percentiles, we compute and plot the median hourly wage divided by the national minimum wage.
*Dataset:* Panel DADS.

## 1.B Data and Descriptive Statistics

### Threshold values

**Figure 1.18:** Share of Observations Below the Minimum Earnings Threshold




*Note:* Figure displays for each year the share of observations with annual earnings below the minimum labor income threshold displayed in Table 1.2.
*Dataset:* Panel DADS.

### 1.B.1 French Cities and Territories

As we mentioned in the previous Section [**WARNING: check**], social movements that took place in France over the last 20 years appear to have a local origin (see Appendix **??** for more details). Hence, we will directly examine how different cities and territories are affected by earnings inequality.

As a first and preliminary step, we present now the concepts that will be used to characterize this spatial dispersion. In particular, we will use repeatedly categories that should help us approximate cities or territories. Indeed, the empirical analyses of Section **??** rely on across-cities and across-territories comparisons. In particular, we will characterize inequality between "urban areas" or between urban and rural "territories". Hence we define these concepts in the following paragraphs. However, first we present some basic facts about French local administrative structure.

France is mostly seen as a centralized country with a capital, Paris, much larger than any other French city and where most of its administration is located. However, France is also a country with

**Table 1.2:** Minimum Earnings Threshold and Minimum Wage

| Year | 1991 | 1992 | 1993 | 1994 | 1995 | 1996 | 1997 | 1998 | 1999 | 2000 | 2001 | 2002 | 2003 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Threshold | 1,850 | 1,913 | 1,909 | 1,941 | 1,964 | 1,992 | 2,024 | 2,077 | 2,100 | 2,111 | 2,153 | 2,182 | 2,220 |
| SMIC | 8,964 | 9,215 | 9,555 | 9,764 | 10,065 | 10,398 | 10,729 | 11,047 | 11,229 | 11,484 | 11,903 | 12,285 | 12,758 |
| Year | 2004 | 2005 | 2006 | 2007 | 2008 | 2009 | 2010 | 2011 | 2012 | 2013 | 2014 | 2015 | 2016 |
| Threshold | 2,295 | 2,385 | 2,444 | 2,468 | 2,466 | 2,505 | 2,506 | 2,498 | 2,529 | 2,539 | 2,553 | 2,574 | 2,585 |
| SMIC | 13,468 | 14,232 | 14,833 | 15,206 | 15,664 | 15,952 | 16,125 | 16,409 | 16,944 | 17,163 | 17,345 | 17,490 | 17,599 |

*Note:* Table displays for each year the minimum labor earnings threshold in euros 2018. The SMIC corresponds to the value of annualized minimum wage, accounting for changes that occurred during the year.

**Table 1.3:** Characteristics of the Workers in the Sample (CS).

| Year | Age shares (%) | | | Education Shares (%) | | | |
|---|---|---|---|---|---|---|---|
| | 25-35 | 36-45 | 46-55 | High-School Drop-out | High-School Degree | College Degree | Higher than College |
| 1995 | 37.7 | 35.1 | 27.2 | 67.9 | 13.8 | 10.7 | 7.7 |
| 2015 | 34.9 | 32.9 | 32.2 | 43.8 | 21.5 | 18.8 | 15.9 |

*Note:* Table shows descriptive statistics for CS sample. *Dataset:* panel DADS.

the highest number of municipalities in Europe: almost 35,000 against around 11,000 in Germany or in the United Kingdom. Most municipalities are small, with an average size of 1,800 inhabitants and a median of less than 500. For years, there has been some political desire to reduce this number.[28] However, this desire did not convert into real actions.

To perform comparisons across geographical units, in particular cities, we aggregate municipality-level data at the “Urban Area” level using the boundaries defined by the French Statistical Institute (INSEE) as of 2010. An urban area comprises a core center with at least 1,500 jobs and adjacent municipalities among which at least 40% of the employees work in the core center. Urban areas are typically smaller than commuting zones except for the largest cities such as Paris, Lyon, and Marseille.

There were 771 urban areas in 2015 which included 85% of the population.[29] Their size ranged from 2,500 inhabitants in 2015 to 11M in Paris. Due to its size, much larger than any other French urban area, and its strong heterogeneity, we run a specific analysis for Paris’s urban area. More precisely, we divide Paris’s urban area into three parts: the municipality of Paris, the close and the distant suburbs.[30]

Urban areas provide a reasonable approximation of cities and are useful to compare their relative convergence or divergence when examining earnings inequality. Nevertheless, these urban areas (as their name indicates) do not cover the whole territory and exclude rural areas. In addition, this view of cities as urban areas precludes making differences between city centers and suburban areas for example.

To get a comprehensive view of labor earnings dynamics between territories, we divide French municipalities into five groups of “territories”: rural areas, suburban areas, remote municipalities, central municipalities, and Paris.

We explain now how these territories are defined. A city might include one or multiple municipalities. When there is only one municipality, the city is classified as a “Remote Area”. When it includes several municipalities, these are divided into Central and Suburban municipalities based on their size.[31] Due to its prominent size and role in the French economy, we exclude Paris from Central municipalities and study it separately.

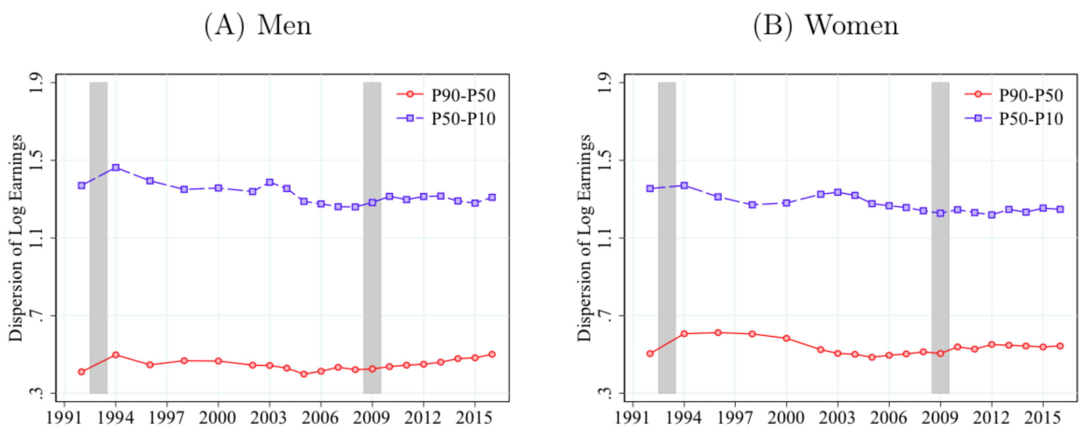
28 See **?**

29 We exclude areas “000”, “997” and “998” which include municipalities that do not belong to urban areas.

30 The municipality of Paris is composed of its twenty districts. The close suburbs include all the municipalities of the Urban Unit (*Unite urbaine*) of Paris, while excluding the Paris municipality. Finally, distant suburbs are composed of the municipalities of the urban area which are not included in the Urban Unit of Paris. We define urban units in the following paragraphs.

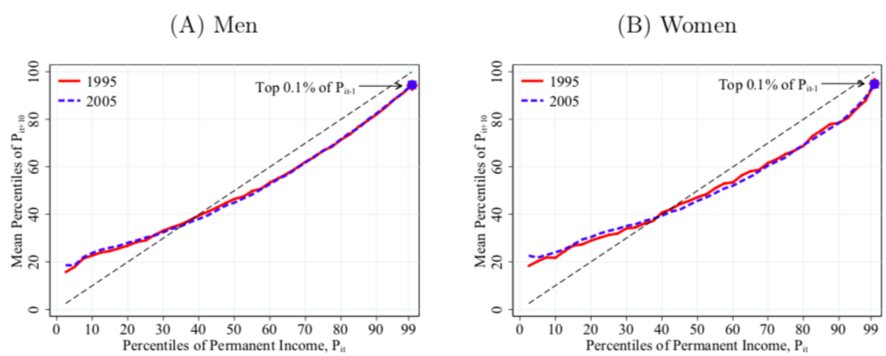
31 Central municipalities are either the biggest municipality if its population comprises at least 50% of the city’s population or the biggest municipality and all the municipalities with a population at least equal to 50% of the biggest municipality.

**Figure 1.19:** Initial Earnings Inequality (at age 25)




*Note:* Using real raw log earnings and the CS+TMax sample, Figure 8 plots against time the following variables: (a) Men: P90-50 and P50-10 at age 25, (b) Women: P90-50 and P50-10 at age 25. Shaded areas are recessions. *Dataset:* Panel DADS.

**Figure 1.20:** Evolution of 10-Year Mobility Over Time




*Note:* Using permanent income, Figure **??** plots the average rank-rank mobility in 1995 and 2005 separately for men and women. *Dataset:* Panel DADS.

Municipalities are classified by the French statistical institute into the above 5 categories based on the urban unit they belong to.[32] An urban unit is defined as a group of municipalities with a total population of at least 2,000 inhabitants and with a continuity of the built-up area. Rural municipalities are municipalities which do not belong to an urban unit.

Since people may have different jobs in different locations, we compute our statistics based on the residency location, measured at the municipality-level. If people change within a year, we define residency as the one associated with the highest paying job. The variable on municipality of residency is available since 1993 so we start our geographic analysis at this date.[33]

[32] As a result, Paris refers to the Urban Unit of Paris when considered for across-territories comparisons.

[33] For this specific Section, we exclude people living in Corsica because the precise place of residency is not available every year. Hence, our across cities and across territories statistics will restrict to continental France.

## 1.C Additional Sources of Spatial Heterogeneity

We show in what precede that the timing of the earning changes within territories strongly correlated with the implementation of two major reform of the French labor market. We have shown that the territories that experienced higher earning growth are also the one with more workers affected by the reforms. Using a earning decomposition, into hours worked and hourly wage in the precedent section, also goes in the same way. However, it is possible that other factors explained the pattern describe in this section (the catch-up of low-wage women in low-density areas). For instance, as low-density areas have been greatly integrated in metropolitan labor market it might have induced an increasing job-to-job mobility that favored wage growth. It might also have induced an increasing in residential mobility with urban residents moving to live in low-density areas, thus bringing their individual urban wage premium in these territories (**?**).

We investigate the heterogeneity of labor market access to opportunities between residential locations, since previous results might reflect a better integration of remote locations to the labor markets and lower migration costs. Recent research have focus on frictional labor mobility (**?**; **?**) showing that workers face spatial frictions and mobility costs that decrease their ability to compete for distant job opportunities. Frictions affect the frequency of job transitions, while residential mobility costs impact the distribution of accepted wages.They highlight that the urban wage premium is driven by better opportunities for local job-to-job transitions in larger urban areas, while migration reduces lifetime inequalities by providing insurance against unsatisfactory initial location draws. In the following we discuss this two competing explanation. We look first at the job-to-job mobility by territories of residence and then residential mobility between territories. Our results reveal however that no substantial changes in job-to-job or residential mobility occurred over our period of study.

### 1.C.1 Job-to-job mobility

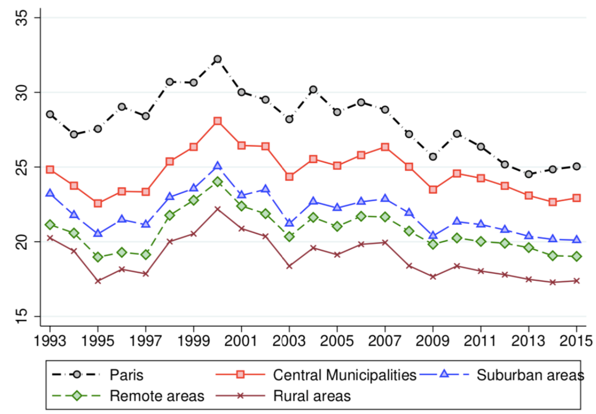
As documented in the literature, job-to-job mobility is a substantial driver of earning growth (e.g. the urban wage premium decomposition by **?**). Switching job allows workers to bargain for higher wages. If remote locations are more integrated to the rest of the national labor market, it should then be reflected in higher job-to-job mobility over time compare to urban municipalities. Figure 1.21 presents job-to-job mobility statistics by territory of origin, i.e. including both within and between territories moves, from 1993 to 2015. A move entails a change of establishment between $t$ and $t+1$.

As expected for a dense territory with more jobs and firms, mobility is highest in Paris' urban unit. It is lowest in rural areas, again an unsurprising result in a less dense territory that have access to less job opportunities. For all territories though, the changes are strictly parallel and pro-cyclical (with a maximum in 2000). Job-to-job mobility steadily decreases from 2001 on in all territories. No clear convergence seems to arise between territories. The higher access to opportunities in rural locations seems not to explain much of the decreasing inequality between territories, implying that frictional labor mobility across residential locations has remain substantial over time.

We test alternative specifications, with full-time workers and firm-to-firm mobility, to challenge this result. Employees working full-time[34] in $t$, experience similar trends but have a much lower

[34] We define full-time workers as worker working full-time at least 350 days during the year.

**Figure 1.21:** Job-to-Job Mobility by Territory




*Note:* Figure plots the evolution over time of the share of workers who change plant between two consecutive years. Each share is computed based on the place of residency before the change happened, irrespective of the place of residency after the change. For each individual in the data, we consider only the highest paying job. Territories and Paris are defined using urban units. *Dataset:* Panel DADS.

mobility rate: around 14% in Paris and 9% in other territories. This large gap results from the inclusion in our main sample of many low-earners, most likely to experience periods of unemployment and to be hired under very short-term contracts. Using firm-to-firm mobility does not alter the above results (mobility larger in Paris, smaller in rural areas) with between 13% and 20% of workers moving every year.

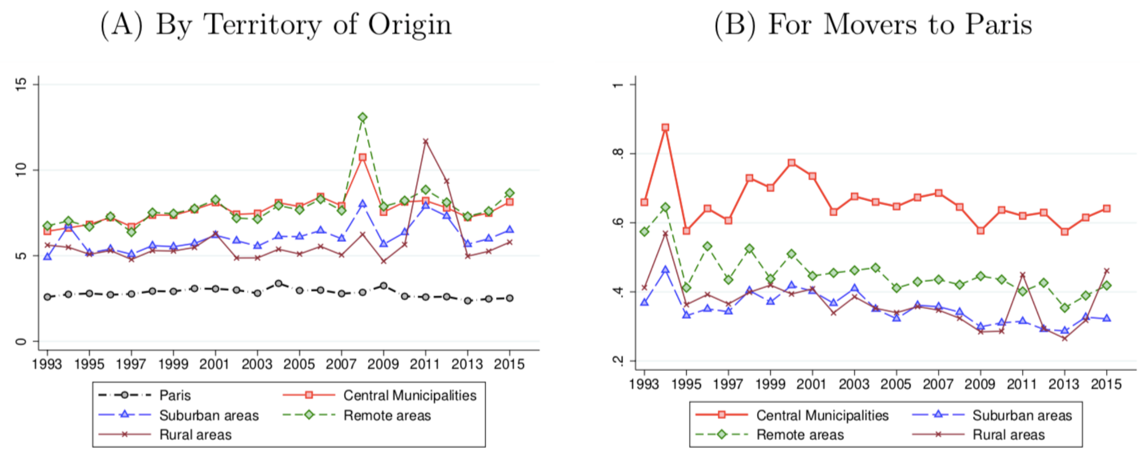
## 1.C.2 Residential mobility

Migration costs is also a source of inequality across territories that might affect earning dynamics, since they prevent workers born in deprived places to move to better opportunities. Geographic mobility mostly takes place between municipalities rather than between urban areas or regions. We first describe the mobility between territories: moves between, say, Paris and Rural areas and their impact on earnings.

Figure 1.22 display residential mobility rate by territories and across them. Mobility from Paris is half as likely (2.5% per year) as from each other territory (5-6% per year in 1993, slightly increasing to 6-8% in 2015). And mobility to Paris, the largest destination mostly comes from central municipalities rather than from more remote places. Over our period of study, we do not see structural changes in mobility patterns.

To give a fuller view of mobility, we contrast earnings growth for the stayers and for the movers. Figure 1.29 in appendix shows the 5-year (log) earnings growth distribution for stayers in Paris, other metropolis ("*Province*" hereafter) and low-density areas (also denoted rural areas in what follow). The distribution of earning growth is very similar in the rest of France: leptokurtic with negative skewness. The distribution in Paris is slightly less peaked and exhibits less negative skewness, suggesting more proximity with the Gaussian distribution and less dramatic shock (such that unemployment spell). Figure 1.23 show different distributions for movers (again between these three origins/destinations) with higher dispersion and lower kurtosis of 5-year earning growth. In contrast with stayers, moves to Paris induce a right shift in the 5-year (log) earnings growth distribution with the associated thickening of the right tail (upwards mobility) whereas mobility

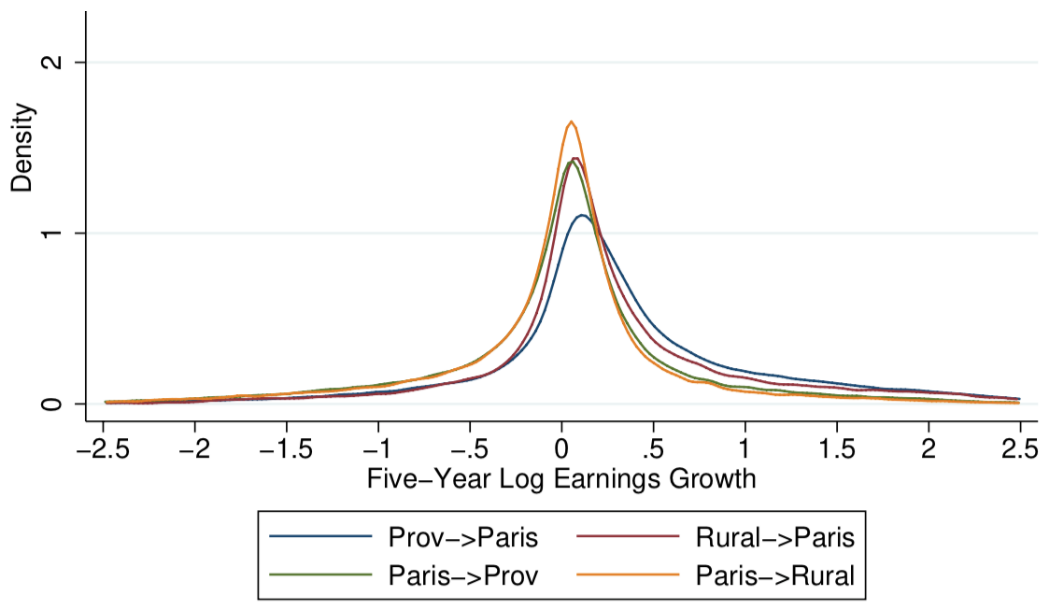
**Figure 1.22:** Residential mobility between territories




*Note:* Figure plots against time the following variables: (a) share of workers changing place of residency between $t$ and $t+1$ by territory of origin (b) share of workers moving to Paris between $t$ and $t+1$ by territory of origin. *Dataset:* Panel DADS.

from Paris has the opposite consequence (downwards mobility). It implies that moves to Paris are associated with higher earning gains on a 5 years period, while it is not the case for Parisian moving to rural areas. For movers to and from destinations other than Paris (not display here), the earnings distributions do not shift much, very similar to those observed for stayers, albeit with a decreased fraction of zero earnings growth (less peaked).

**Figure 1.23:** Five-Year Log Earnings Growth for Movers




*Note:* The Figure plots the density of the five-year log earnings growth separately for workers moving from or to Paris. A mobility is defined as a worker changing of place of residency between $t$ and $t+1$. *Dataset:* Panel DADS.

## 1.D Earning Growth Process

### 1.D.1 Presentation and Methodology

We have shown in the main text that life-cycle inequalities were substantial, with a higher level of earning inequality at the entrance in the labor market that fade with age for women, but increase for men during their 30s. To better understand this trend, we now propose to look at the earning dynamics process. Recent study by **?** has renew the debate on income risks, showing that labor earning dynamics was not linear and that the distribution of earning growth did not follow a Gaussian. Such observations are important to better understand the underlying process that lead to inequality growth or decline. It is one of the purpose of the Global Repository on Income Dynamics (GRID) to provide a systematic investigation of earning dynamics across countries and over-time, as well as to provide cross-countries comparison.

To study labor earning dynamics we compute one-year earning change at the individual level, which corresponds to the difference between the sum of total earnings from employment spell during year $t$ minus total earnings from employment spell during year $t-1$. It allows us to construct the whole distribution of earning changes between every two consecutive years. These distribution depart widely from normality and vary with earning level. Our approach is non-parametric, implying, as in **?**, that we do not make any assumptions on the underlying growth process. We therefor compute the first four moments of these distributions by earning percentiles and plot them to see how the shape of the earning growth distribution vary with permanent earnings.[35] That is, how earning change between two years for workers at the bottom of the distribution compare to workers with top earnings. We also decompose individuals by age-groups, since we observe that the earning process might vary over the life-cycle.

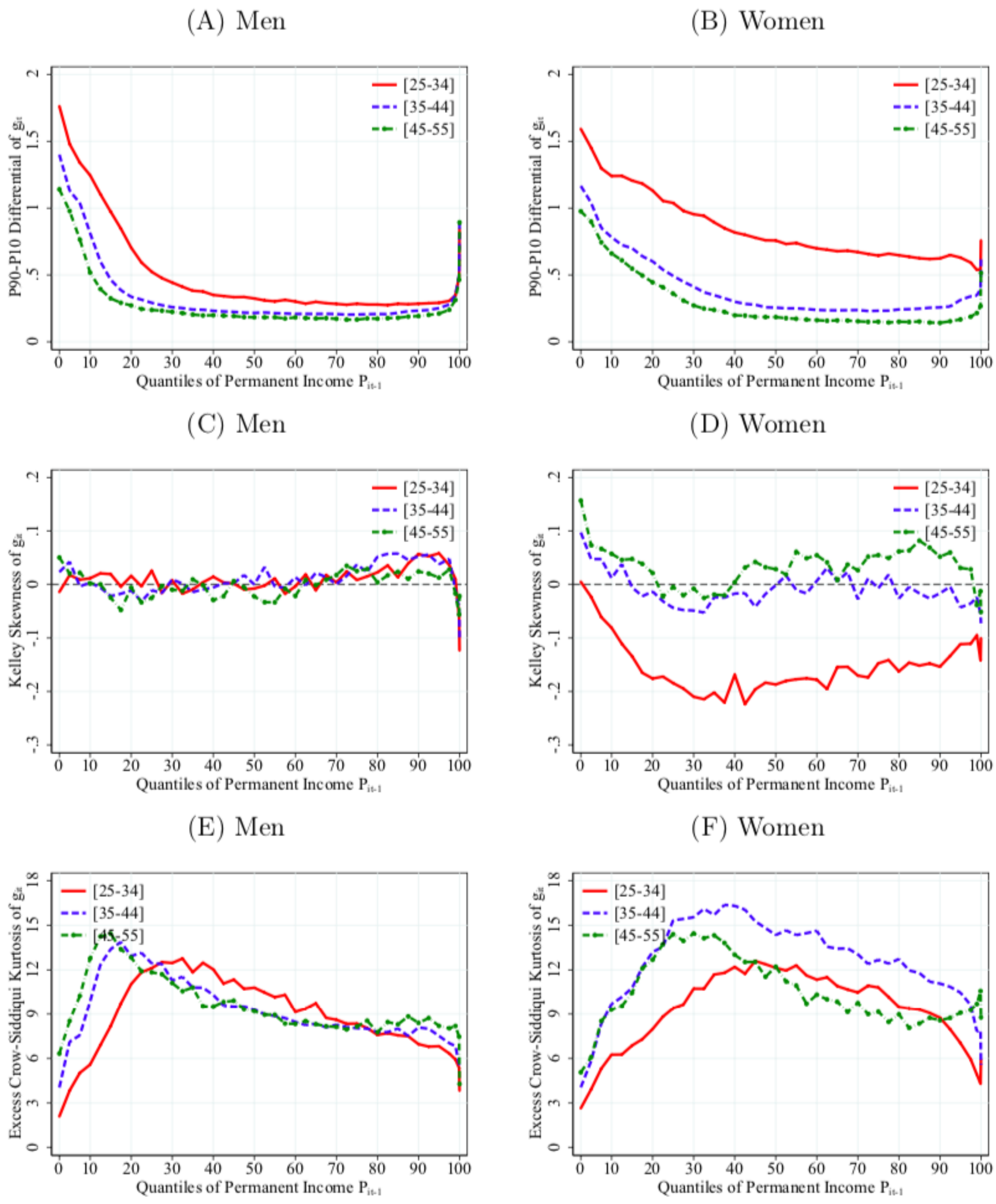
Figure 1.24, in appendix, display these statistics, with the P90-P10 ratio that measures the dispersion of the distribution of earning growth, the Kelly-Skewness, which measures the degree of asymmetry of the distribution that corresponds to the probability of positive shocks compare to negative ones, and the Excess Crow-Siddiqui Kurtosis that measures the "tailedness", which corresponds to the shape of the tail of the distribution, especially a high excess kurtosis means that the distribution asymptotically approach zero more slowly than a Gaussian. In addition, in the appendix, Figure 1.25 display the evolution of the P90-P10 ration over time, while Figure 1.26 display the evolution of the measure of Skewness and Kurtosis. These time series corresponds to the statistics for the whole distribution of earning growth for each years.

### 1.D.2 Results

First, the dispersion of one-year earning growth is higher for women than for men, and for young compare to aged workers (see panel A and B in Figure 1.24). That is, on average, women experience more important income variations, especially when they are young (below 35). It thus reveals less stable labor conditions for women, with higher variations at the extensive margin, with higher volatility of hours worked, especially for low-wage women, who work less than 35 hours a weak

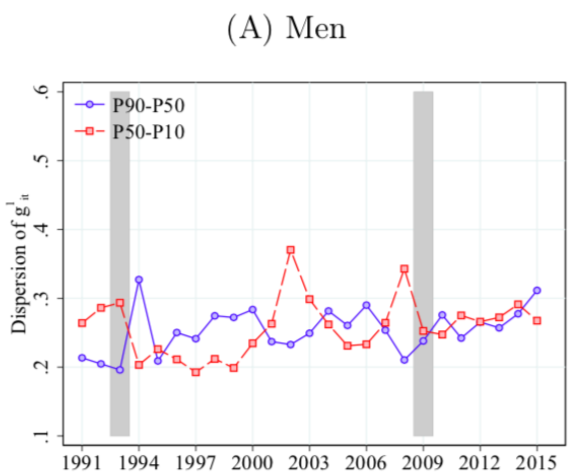
[35] Permanent earning is define as the sum of earnings during employment spell over a three year period. It thus extend previous measure of raw real earnings, that is computed on a yearly basis, to a raw real earnings computed on a three year basis.

**Figure 1.24:** Dispersion, Skewness and Kurtosis of One-Year Log Earnings Changes by Age Group




*Note:* Using residual one-year earnings changes and the H+TMax sample, Figure 1.24 plots against permanent income quantile groups the following variables for the 3 age groups: (a) Men: P90-10, (b) Women: P90-10, (c) Men: Kelley Skewness, (d) Women: Kelley Skewness, (e) Men: Excess Crow-Siddiqui kurtosis, (f) Women: Excess Crow-Siddiqui kurtosis. Excess Crow-Siddiqui kurtosis calculated as $\frac{P97.5-P2.5}{P75-P25} - 2.91$ here the first term is the Crow-Siddiqui measure of Kurtosis and 2.91 corresponds to the value of this measure for Normal distribution.
*Dataset:* Panel DADS.

**Figure 1.25:** Dispersion of One-Year Log Earnings Changes



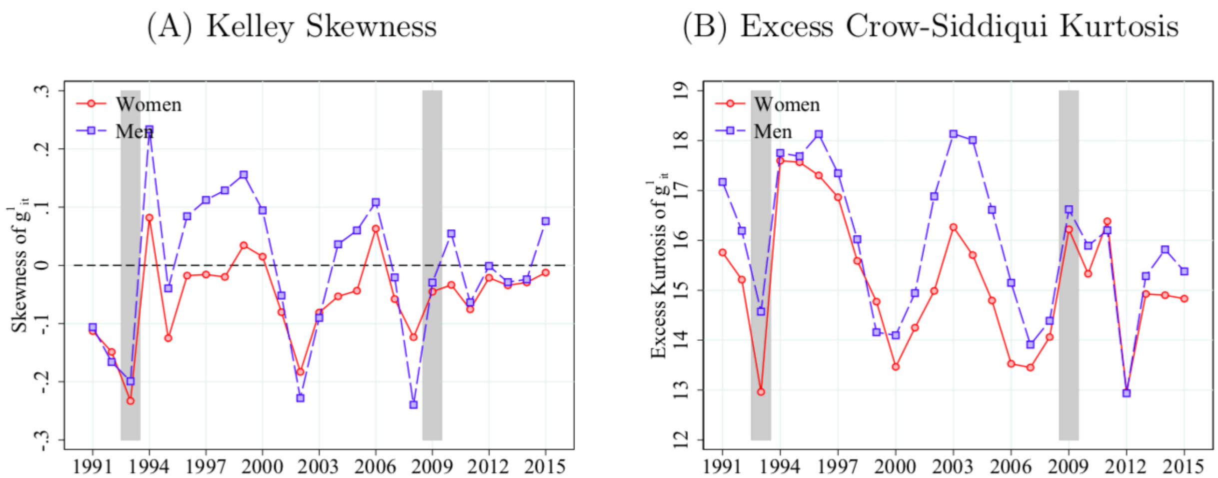

*Note:* Using residual one-year earnings changes (controlling for age) and the LS+TMax sample, Figure 10 plots against time the following variables: (a) Men: P90-50 and P50-10, (b) Women: P90-50 and P50-10. Shaded areas are recessions. *Dataset:* Panel DADS.

and less weak than men on average. Along earning percentile, earning growth rate is more disperse at the two ends of the distribution. Implying more uncertainty in the earning process for low and top-earners, while it is less true for female top-earners. It is also related to individual career path with higher wage growth during the first years of entrance into the labor market. Women experience larger uncertainty during their career than men, especially at ages between 25 and 34 y.o. Indeed, the earning dynamics is very disperse for low-paid and young women, much more than for men. It could result from a higher variation in the number of hours worked. But this also suggests a strong effect of maternity in earning growth at early stage of women career that could lead to strong reduction in hours worked and to flatten their wage prospects, as describe by **?** in Denmark and by **?** in additional countries.

Over time, the dispersion of earning growth has remained higher for women than for men (see Figure 1.25). Interestingly, we see an increase in the dispersion of earning growth during the labor market reforms, implying more disparities among workers in term of their earnings adjusting to the reforms. In particular lower (if not negative) earning growth (the 10th percentile of the earning growth) depart from the median value. It is less so for higher earning growth (P90-P50). The period before the Financial crisis was also associated with an increase in the dispersion of earning growth that the following recession stopped.

Second, the earning growth distribution is almost centered around the median value (display in previous figures) for all percentiles for men (see panel C in Figure 1.24). Except at the very top (above the 1th percentile), where it is left-skewed, implying that the probability of large negative (compare to median) income shock are greater than large positive (compare to median) income shock. But small positive income shock are in reverse more frequent than small negative shocks. The pattern is identical at all ages for men. By contrast, for women in the younger age group (25-34) and for almost all percentiles of the income distribution the earning growth distribution is left-skewed, implying large earning losses compare to the median (see panel D in Figure 1.24). Such shock arise between ages 25 and 34, and might be related to maternity. It is interesting to remark that all earning percentiles are concerned, and even more the highest ones. This suggest a stronger

**Figure 1.26:** Skewness and Kurtosis of 1-Year Log Earnings Changes




*Note:* Using residual one-year earnings changes (controlling for age) and the LS+TMax sample, Figure 11 plots against time the following variables: (a) Men and Women: Kelley skewness, (b) Men and Women: Excess Crow-Siddiqui kurtosis calculated as $\frac{P97.5-P2.5}{P75-P25} - 2.91$ where the first term is the Crow-Siddiqui measure of Kurtosis and 2.91 corresponds to the value of this measure for Normal distribution. Shaded areas are recessions. *Dataset:* Panel DADS.

penalty for high-skill female workers, since such occupations allow less flexibility in hours worked with nonlinear (convex) pay structure and reward more persistence in the position, as describes in **?**'s last chapter of the Grand Gender Convergence (**?**). In addition, the growth distribution is pretty centered, if not right-skewed, for ages between 35-44, which could correspond to a recovery process and an earning growth distribution with a thicker right-tail, associated with a substantial higher probability of high earning growth. For the older age groups, 45 to 55, the distribution is centered around the median, as for men.

Figure 1.26 display the evolution of the mean Skewness over years for men and women. It shows that the value of the Skewness increase during periods of expansion, implying that the probability of large negative socks decrease with economic growth. However, the distribution of earning risks is rarely right-skewed.

Finally, the excess kurtosis of earning growth is large for all percentiles and has an inverted U-shape across the percentiles of the permanent income distribution, with a maximum below the median (Panel E and F in Figure 1.24). It implies that earning growth is fat tailed for both men and women. It means that extreme values are much more frequent than what a normal distribution will predict. It is an important observations since it means that the dynamic of individual earning is much more abrupt than we thought, when we do not account for unemployment benefit. Fatter tails are observed around the 10th percentile for men and 25th percentile for women.[36] Thus, women experience larger earning shocks than men, with a higher excess Kurtosis for almost every percentiles, especially at ages between 35 and 44. Note that it can result from the gender pay gap and the lower earnings of women at all percentiles, which are thus more sensitive to smaller variations in relative value.

Time series of the excess kurtosis are display in Figure 1.26. It appears that the trends are pretty similar for men and women. We observe increasing kurtosis during recession, implying that

[36] It is important to recall here that we consider the *growth rate* distribution. Hence, in absolute value higher percentile might have higher gains, but not so in relative growth rate.

the magnitude of shocks are larger in bad time. Linked with negative skewness during this period, this reflects the fact that the dynamics of earning during recessions becomes very detrimental to the workers while the rest of the time it tends to recover a Gaussian pattern.

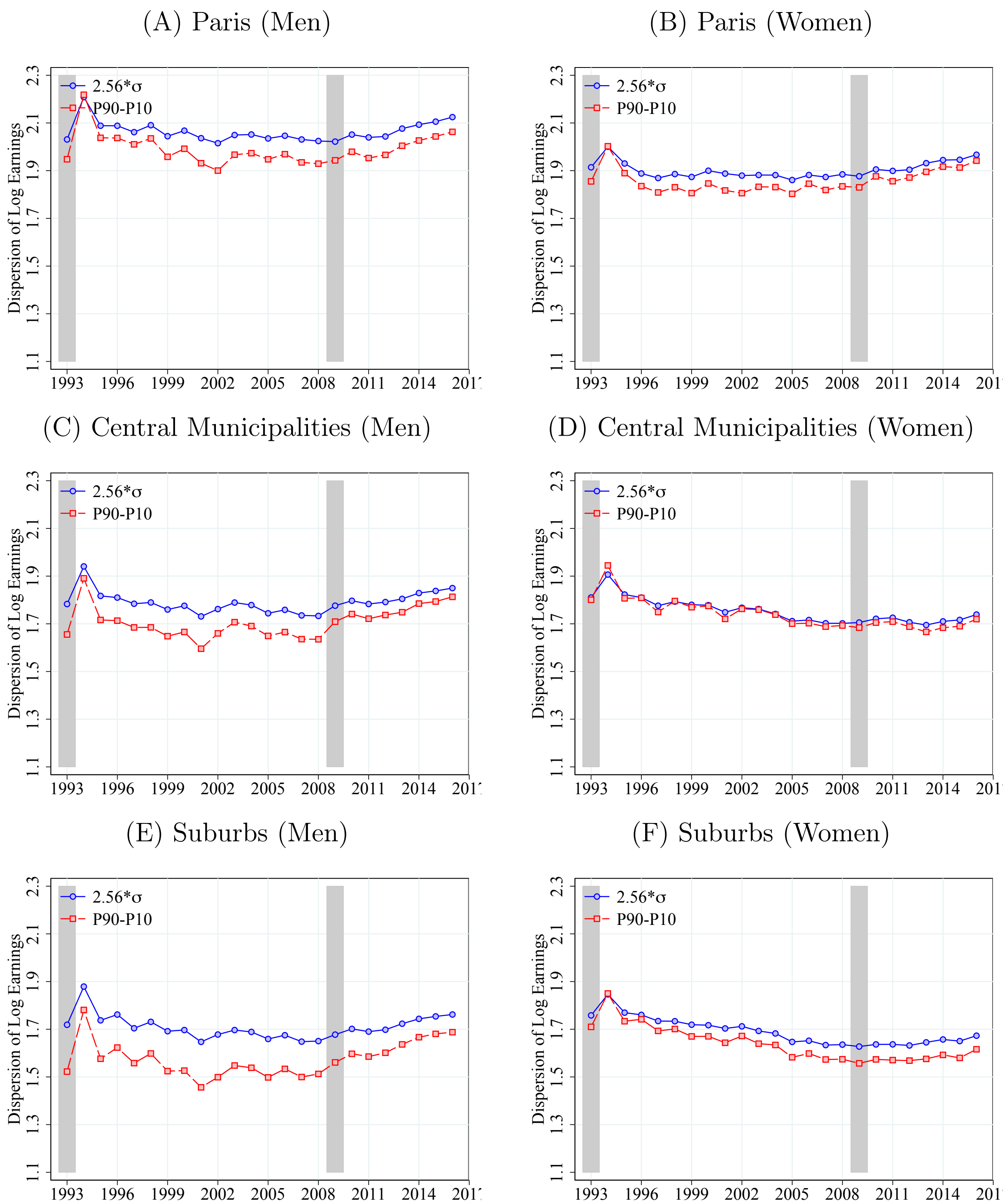
## 1.E Spatial Heterogeneity

### 1.E.1 Earnings Inequality by Territory

### 1.E.2 Earning Dynamics and Residential Mobility

### 1.E.3 Urban wage premium and individual characteristics

**Figure 1.27:** Earnings Inequality by Territory (1/2)

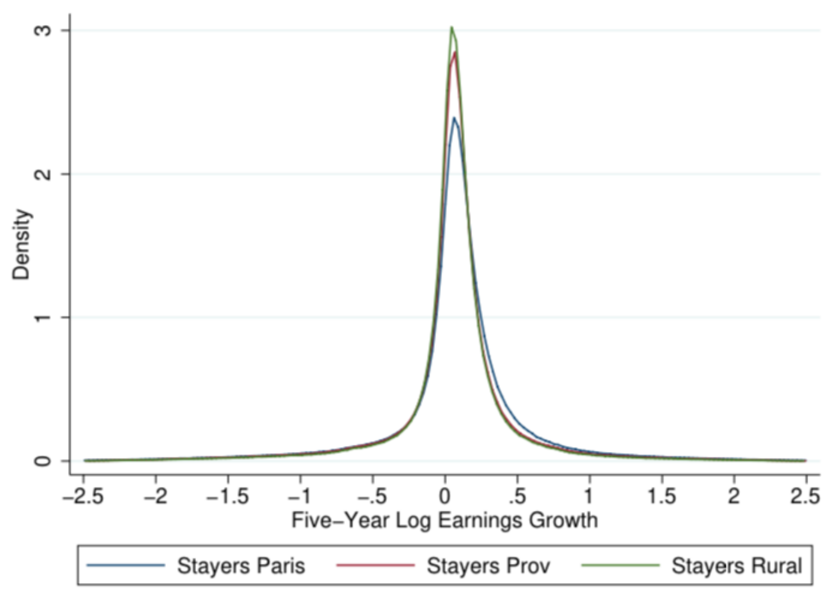



*Note:* Using real raw log earnings separately for men and women, Figure H.3 plots against time the P90-10 and 2.56*SD of log earnings for: (a-b) the urban unit of Paris, (c-d) central municipalities and (e-f) the suburbs. Territories and Paris are defined using urban units. Shaded areas are recessions. 2.56*SD corresponds to P90-10 differential for a Gaussian distribution. *Dataset:* Panel DADS.

**Figure 1.28:** Earnings Inequality by Territory (2/2)

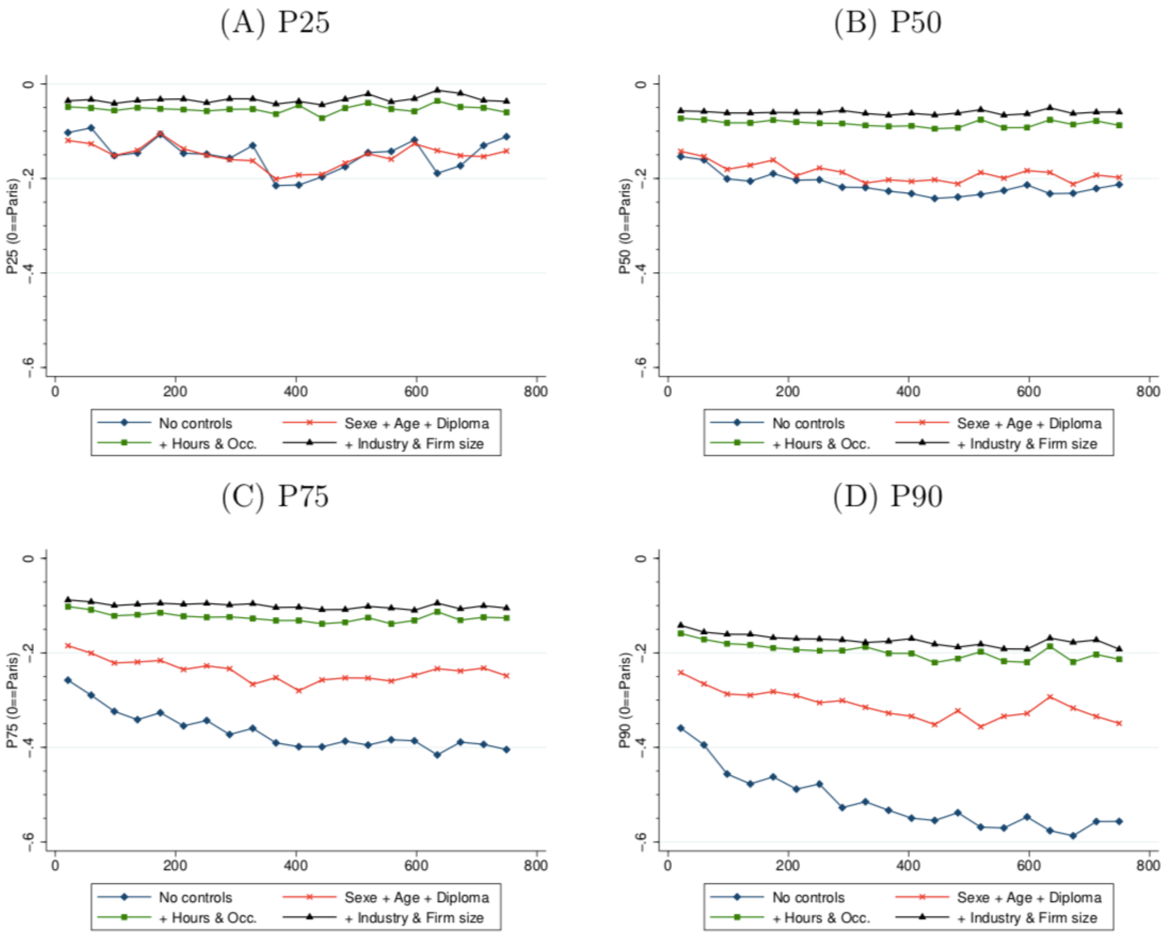

*Note:* Using real raw log earnings separately for men and women, Figure plots against time the P90-10 and 2.56*SD of log earnings for: (a-b) remote municipalities, (c-d) rural municipalities. Territories and Paris are defined using urban units. Shaded areas are recessions. 2.56*SD corresponds to P90-10 differential for a Gaussian distribution. *Dataset:* Panel DADS.

**Figure 1.29:** Five-Year Log Earnings Growth for Stayers




*Note:* Figure plots the density of the five-year log earnings growth separately for workers staying at least three consecutive years ($t-1, t, t+1$) in a given territory. The territory denoted "Prov" includes all territories except for the urban unit of Paris and rural territories.
*Dataset:* Panel DADS.

**Figure 1.30:** Between-Urban Areas Inequality After Controlling for Observable Characteristics




*Note:* Using residual log earnings for men and women, the Figure plots against city (urban area) rank (a) P25, (b) P50, (c) P75, (d) P90 for the period 2014-2016. City rank is based on 2015 census population. Controls include age dummies, four education dummies, the total number of hours worked, 2-digit occupation dummies, 4-digit industry dummies, and the size of the firm in head count. Observations are the 759 urban areas. They are displayed using a bin-scatter. *Dataset:* Panel DADS matched with EDP.

# CHAPTER 2

## Railway, Highway, and the Suburban Metropolis

*Evidence from Paris (1968-2014)*

with Corentin Trevien

**Abstract**

Using historical data on travel time and commuting flows for the Paris Region, we assess the impact of two major infrastructure programs: the Regional Express Rail (RER) and National Highways. Although the infrastructure programs contributed to a decrease in travel time in the overall network, individuals' commuting time has increased over the period. We use the structural gravity equation as a sufficient statistics to show the causal impact the two transportation programs had on commuting flows. If individuals substitute between the two modes to maximize their travel time, we show that the two infrastructures were complements in the making of the suburban metropolis. Faster rail lines have sustained significant commuting flows between downtown neighborhoods and the suburbs, in both ways; while roads have supported the rise of suburb-to-suburb connections, reducing detours.

**Keywords**: Commuting; Network; Urban planning; Mass-Transit; Highways.

**JEL Classification**: H54, N94, R12, R41.

Circulate under the following reference: Delemotte, Thomas and Trevien, Corentin. "*Railway, Highway, and the Suburban Metropolis: Evidence from Paris (1968-2014)*". Mimeograph (2022).

For comments that have improved this article, we are grateful to Nathaniel Baum-Snow, Ondine Berland, Clément Bosquet, Pierre Boyer, Léa Bou Sleiman, Laurent Gobillon, Florence Goffette-Nagot, Francis Kramarz, Raphaël Lafrogne-Joussier, Thomas Le Barbanchon, Isabelle Méjean, Tomasz Michalski, Ninon Moreau-Kastler, Louis-Daniel Pape, Roland Rathelot, Julien Sauvagnat, Chris Severen, James Siodla, Benoît Schmutz, Clémence Tricaud, and Yanos Zylberberg. We would also like to thank conference and seminar participants at UEA North American Meeting, IAAE Conference, EEA-ESEM Congress, Journées LAGV (AMSE), ERSA-European Commission (JRC), and CREST.

The authors gratefully acknowledge the Investissements d'Avenir (ANR-11-IDEX-0003/Labex Ecodec/ANR-11-LABX-0047) for financial support, the CASD (Centre d'Accès Sécurisé aux Données), the Quételet research network, the French National Institute of Geography (IGN), and INSEE for the access to confidential data.

The results presented in this chapter are the sole responsibility of the authors.

## Introduction

One of the major developments of recent decades has been the simultaneous concentration of economic activities in large metropolises and their dispersion from the city center, with the suburbanization of both population and jobs. This has led to an unprecedented concentration of wealth in the largest metropolises: Tokyo contributes 33% of Japanese GDP, Greater London 28% of British GDP, and Paris region 24% of French GDP.[1] It also implies a dramatic increase in commuting time and polluting emission. This induces major considerations on how to better organize cities to foster their sustainable growth.[2] Yet relatively little is known about the role of commuting infrastructures and alternative transportation modes in sustaining dense concentration of economic activities with large-scale commuting.

Heblich et al. (2020) has shown how the steam railway innovation in the XIXth century has shaped the making of the modern metropolis and the separation of residential location (in the suburbs) and workplaces (in downtown areas). The radial design of most of transportation infrastructures has lead to many study of the core-periphery structure of the metropolis. However, the emergence of individual car commute has fostered the rise of the suburban metropolis, which attracts workers outside the urban fringe, and the increase of suburbs-to-suburbs commuting, as we detail in this chapter. Therefore it is important to better understand how rail and road infrastructures interact in shaping large-scale commuting. Especially, are they complements or substitutes, and do they foster the extension of the metropolis in the same way?

This chapter focuses on two major infrastructure programs in Paris region affecting commuting in the second half of the twentieth century: the development of mass-transit and the building of highways. It therefore provides a topical case study to disentangle the contribution of both rail and road in shaping commuting in the suburban metropolis. Using individual-level data on workers from 1968 to 2014, we link commuting flows with travel time over this period for the entire Paris region. This data allows us to describe the substantial evolution of commuting pattern over the period. We then micro-found our empirical framework to study the incidence of the reduction in travel time on commuting flows, accounting for potential confounding factors such as place-based policies or targeted investments. We find that the two modes are complements in the making of the suburban metropolis. One the one hand, rail enables the sustaining of significant commuting flows along radial lines. On the other hand, roads foster suburb-to-suburb connections with few detours. At the individual level, we show that the two modes are substitutes: within residence-workplace pairs, longer travel time with rail cause additional commuting flows by car, while potential traffic on roads fosters the use of the public transport network to commute.

This chapter studies the Paris region (12 million inhabitants) to highlight the contribution of transportation infrastructures to the making of the suburban metropolis. First, as shown in the literature, transportation infrastructure improvements are a key explanation of the suburbanization process, and the Paris region experienced both the implementation of many new highways and a substantial improvement of its regional train system. We document the historical

[1] Data for city gross domestic product comes from the Brooking *Global Metro Monitor*, 2014. Data for national gross domestic product comes from the World Bank, 2014.

[2] For instance in France, major plan are set-up in order to promote low-externality transportation modes and to re-design transportation infrastructure. See e.g. SNCF et al. (2020), `https://www.ecologie.gouv.fr/sites/default/files/EF%26SEM-SD%20VF%2006%2004%202020.pdf`.

background of this urban planning and the two infrastructure programs in the first part of the chapter, while the data construction is described in the second part of the chapter. The use of French administrative data allows to recover both the places of residence and the place work for every inhabitant of the Paris region at the municipality level (1,287 distinct areas). We then compute travel time with public transport or car between any two municipalities, separately, for every years between 1968 and 2014. Although the infrastructure programs contributed to a decrease in travel time in the overall network, individuals' commuting time has increased over the period. This is due to a higher increase in commuting distance as compared to the amount of time saved. These changes have induced a shift in commuting patterns, with a decline of the share of commuting to the downtown area in favor of suburb-to-suburb commuting.

To study the causal incidence of the infrastructure improvements on commuting flows, we micro-found a structural gravity equation, with "three-way" fixed effects. It actually corresponds to a difference-in-difference, in which we identify the elasticity of commuting flows to travel-time from variations induced by the heterogeneous improvement of the transportation infrastructure system within the metropolis. In particular, we are able to control for confounding factors such as place-based policies, with location-specific fixed-effects, and targeted investment, with dyadic fixed-effects. Since their is different transportation modes available in the metropolis, we define a generalized cost of travel time to account for these alternatives.

We then show that road and rail infrastructures are substituted at the individual level within a given residence-workplace pair. Shorter travel time with public transport are associated with less car commuters. While higher potential traffic on roads increases the number of commuters using public transport. However, the two infrastructure programs have revealed to be complement in the making of the suburban metropolis. Improvement of the rail network have caused higher commuting flows between locations that are more distant along rail lines. Especially it has sustained substantial commuting flows between downtown neighborhoods and the suburbs. Conversely, the road improvement has fostered suburb-to-suburb commuting. For short distance car avoid waiting time associated with the use of public transport. For long distances it reduces the number of detours.

Finally our theoretical approach of the gravity equation allows us to structurally interpret our location-specific fixed effects. They encompass both the demography of places (variations in the numbers of residents or jobs) and the incidence of infrastructure improvement on places. The latter being the evolution of the relative accessibility of places in the new network.Our estimation shows, on the one hand, that residential locations beyond the urban fringe have benefited relatively more from infrastructure improvement to access employment opportunities. Central locations have conversely loose from their relative centrality. It suggests a better integration to the metropolitan labor market of formerly exclude municipalities. On the other hand, the places of work that benefited more from the new infrastructures in term of jobs are in the urban periphery and correspond to network nodes (e.g. Cergy, Massy, Roissy).

## Related literature

If the invention of the steam railway led to the first large-scale separation of workplace and residence during the XIXst century, the subsequent innovations in transportation technologies (cars and express rails) have reshaped metropolises with more complex commuting pattern. So far the literature has demonstrated the causal relationship between transportation infrastructures, both highways and railways, on urban sprawl with the growth in the periphery of the number of both residents and jobs (see for highways Baum-Snow, 2007 and Duranton and Turner, 2012; for railways Garcia-López et al., 2017 and Mayer and Trevien, 2017). However these works do not study the incidence of these infrastructures on the effective number of commuters, that provides a sufficient statistics to characterized the urban equilibrium. In addition, most of previous contributions rely on exogenous variations in travel time over radial infrastructures, connecting directly the metropolitan center with its periphery. Thanks to our data and our empirical framework, we are able to state the causal incidence of transportation infrastructures on effective commuting flows in any direction, beyond the core-periphery framework. It confirms previous results, but contrast them with heterogeneous incidence of travel time according to the targeted dyad. Furthermore, our results show that railways and highways did not have the same impact in the making of the suburban metropolis. The firsts have reinforced long distance along radial lines, while the second have contributed to the increase in the separation of residence from workplace in the suburbs.

We build our empirical strategy on quantitative spatial models (Ahlfeldt et al., 2015, Severen, 2020, Heblich et al., 2020, Tsivanidis, 2020). While this strand of papers rely on full estimation of the model parameters, often constraint by data limitation, since we observe effective commuting flows. Our empirical approach amounts, ultimately, to a difference-in-difference with a fuzzy design, where our treatment is the infrastructure improvement that reduces travel time. In addition, we are able to structurally interpret the fixed-effects in our setting. We show how to adapt recent advance in trade (Arvis and Shepherd, 2013, Fally, 2015) to the urban case. This approach allows us to clarify the role of the different fixed-effects in the estimation.

Finally this chapter contributes to the modal choice literature initiated by McFadden (1978) from a new lens, the network incidence of different transportation mode infrastructure. We show that in addition to different speed, different modes display different accessibility. The use of public transportation, with a discrete number of access point (stations), decline sharply with distance to the stations, and is used for a reduced number of routes that corresponds to rail lines, while car is much more flexible and allows to access remote places in the outskirt. Therefore, rail and road can be complement in the making of the suburban metropolis, while, at the individual level, they are substituted between each others. Our approach accounting for the generalized cost of commuting, confirm previous results showing that rail is used to avoid road traffic (Anderson, 2014).

The rest of the chapter is organized as follow. Section 2.1 describes the historical background and the two major infrastructure programs, before presenting the data and stylized fact in Section 4.1. The structural gravity approach is detailed in Section 2.3, with the estimation method and identification assumption. Finally Section 2.4 and 2.5 discuss our results, before the conclusion.

## 2.1 Motivating facts

The urban development of the Paris region between the 1960s and today is of particular interest because it is a typical example of urban sprawl but also because it has been the object of two major transportation infrastructure investments: the Regional Express Rail (RER) and a dense highway system.

### 2.1.1 A suburban metropolis

#### Urban sprawl

In the aftermath of the second World War, the Paris region (*Ile-de-France*)[3] experienced a dramatic demographic growth. Its population went from 6.6 millions inhabitants in 1946 to 8,5 millions in 1962. The additional 1,9 million residents have been hosted mostly in the surrounding area of Paris.

**Table 2.1:** Commuting pattern over time across the Paris Region (Center, Inner and Outer Suburbs).

| | Day location (workplace) | | | | |
|---|---|---|---|---|---|
| Night location (residence place) | Paris | Inner Suburbs | Outer Suburbs | Total | (%) |
| *Commuting flows in 1968* | | | | | |
| Paris | 1,127,136 | 145,080 | 18,936 | 1,291,152 | 0.306 |
| Inner suburbs | 552,164 | 1,159,132 | 49,360 | 1,760,656 | 0.417 |
| Outer suburbs | 232,120 | 151,432 | 789,604 | 1,173,156 | 0.278 |
| Total | 1,911,420 | 1,455,644 | 857,900 | 4,224,964 | 1 |
| (%) | 0.452 | 0.345 | 0.203 | 1 | 1 |
| *Commuting flows in 2014* | | | | | |
| Paris | 749,553 | 260,119 | 60,657 | 1,070,329 | 0.201 |
| Inner suburbs | 568,902 | 1,199,770 | 191,327 | 1,959,999 | 0.368 |
| Outer suburbs | 359,504 | 479,208 | 1,454,704 | 2,293,417 | 0.431 |
| Total | 1,677,960 | 1,939,098 | 1,706,688 | 5,323,745 | 1 |
| (%) | 0.315 | 0.364 | 0.321 | 1 | 1 |
| *Surface (in sq. km)* | | | | | |
| Total | 105 | 655 | 11,272 | 12,033 | 1 |
| (%) | 0.009 | 0.054 | 0.937 | 1 | 1 |

The spatial organization of Paris Region has, to some extent, a core-periphery structure (see table 2.1). In 1968 as well as in 2014 the city of Paris (*i.e.* the core of the Paris Region) hosted more jobs than (employed) residents, with a ratio of day to night population of 1.48 and 1.57, respectively.[4]. Between the end of the 1960s and the 2010s both jobs and residential population decline in Paris: Paris hosts only 31.5% of job and 20.1% of (employed) residents in 2014 compared to 45.2% and 30.6%, respectively, in 1968. Conversely, employment and population significantly increase in the suburbs – especially in the outer suburbs – and commuting flows within suburbs rise sharply over the period. In other words, concentration of both jobs and residents declines, since emerging suburban workplaces and residential locations are less land-constraint and allow for larger real estate (at lower floor-unit costs): larger plants and larger (individual) housing.

**Table 2.2:** Orientation of commuting flows

| Origin/dest. | Orientation | Dyads | | Commuters | | % Commuters | | 2014/ |
|---|---|---|---|---|---|---|---|---|
| | | All | % >0 | 1968 | 2014 | 1968 | 2014 | 1968 |
| Paris ↔ Paris | Any | 400 | 1.000 | 1,127,136 | 749,553 | 0.267 | 0.141 | 0.665 |
| Paris → Suburbs | Centrifugal | 25,340 | 0.350 | 164,016 | 320,776 | 0.039 | 0.06 | 1.956 |
| Suburbs → Paris | Centripetal | 25,340 | 0.783 | 784,284 | 928,407 | 0.186 | 0.174 | 1.184 |
| Suburbs ↔ Suburbs | Centrifugal (high) | 380,764 | 0.080 | 87,612 | 259,866 | 0.021 | 0.049 | 2.966 |
| | Centrifugal (low) | 313,828 | 0.091 | 114,916 | 324,643 | 0.027 | 0.061 | 2.825 |
| | Centripetal (high) | 380,764 | 0.161 | 217,832 | 643,538 | 0.052 | 0.121 | 2.954 |
| | Centripetal (low) | 313,828 | 0.128 | 223,740 | 560,489 | 0.053 | 0.105 | 2.505 |
| | Peripheral | 214,838 | 0.159 | 259,080 | 610,694 | 0.061 | 0.115 | 2.357 |
| | Same municipality | 1,267 | 1.000 | 1,246,348 | 925,780 | 0.295 | 0.174 | 0.743 |

*Dyads >0 : at least one commuter in at least one Census year between 1968 and 2014*

### Edge city or monocentric city

For a better understanding of suburb-to-suburb commutes, we beak down the 1,656,359 home-to-work flows (or dyads),[5] within the 1,287 municipalities of the Paris Region, into 9 categories (see table 2.2). We construct two "radiality" and "peripherality" indexes, measuring whether commuting flows are core-periphery oriented or note (see details in section 2.A page 125, appendix). Home-to-work trips that follow the core-periphery axis, moving away from Paris, are called "centrifugal", while those in the opposite direction are called "centripetal". Trips that are orthogonal to the center-periphery axis are called "peripheral". First, we noticed the granularity of commuting patterns, as only a small number of (oriented) municipality pairs are effectively used by commuters, implying a majority number of zeros in the commuting matrix, especially between suburban municipalities. We also notice that suburb-to-suburb commuting flows are growing faster than suburb-to-Paris commuting trips. Other indication of the fact that Paris is less monocentric, the share of workers leaving in Paris and working in the suburbs practically doubles. However, among suburb-to-suburb commuting flows, core-periphery-oriented trips are increasing the fastest, which is probably due to the spatial organization of the transportation network.

### Urban planning and place-based policies

The spatial organization of the Paris region in the second half of the 20th century is not only the result of urban sprawl, common to most western cities. It is also largely influenced by strong and numerous public policies. Several types of public interventions have been carried out, initially by the central Governement, then more and more by the local authorities.

One of the first large public intervention dates back to Napoleon III's and Haussmann's

[3] Since 1968, it is composed of 7 "*départements*" (administrative units with local jurisdiction): *Paris* (75, administrative identifier); 3 inner suburbs "*départements*": *Hauts-de-Seine* (92), *Seine-Saint-Denis* (93), *Val-de-Marne* (94); 4 outer suburbs "*départements*": *Seine-et-Marne* (77), *Yvelines* (78), *Essonne* (91), and *Val-d'Oise* (95).

[4] It implies that downtown Paris was hosting 3 jobs for 2 residents.

[5] 1,287 residence places × 1,287 work places

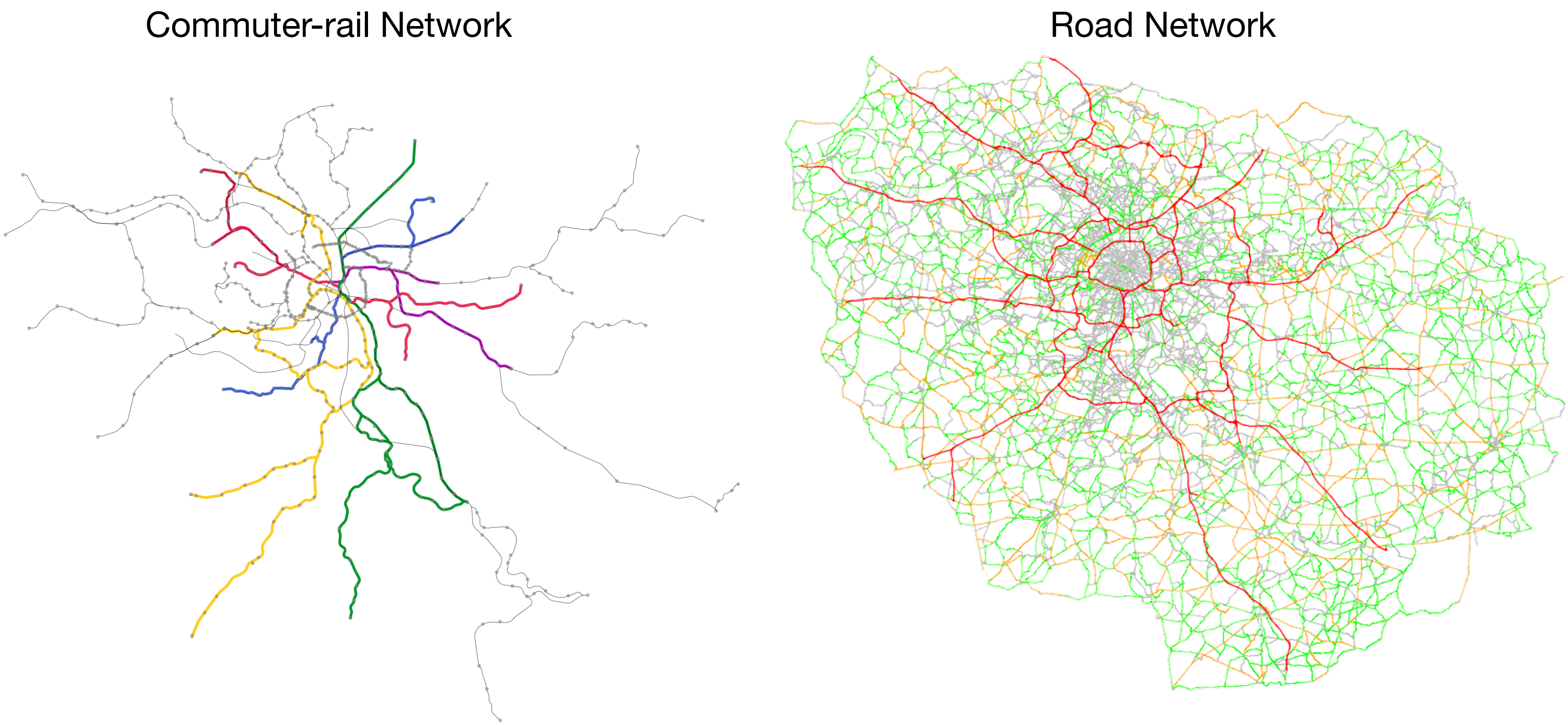
modernization of Paris, in the second half of the 19th century, paving the way for a French tradition of urban and regional planning. From the 1960s, several master plans follow one another to address the uncontrolled urban development of the Paris Region in the period following the Second World War. The "PADOG" plan in 1960, the "SDAURP" plan in 1965, the "SDAURIF" plan in 1976 and the "SDRIF" plan in 1994 and 2013, initiated the construction of new towns, the highway network and the mass transit network. However, the actual achievements are often far below the initial objectives. In addition, the objectives of the policy-makers vary over time. Some master plans aim to limit the growth of the Paris region in favor of the rest of the country, while others see the development of the Paris region as an opportunity for the whole country. Another point of divergence over time, the objective of densifying the inner suburbs replaces the construction of new towns in the outer suburbs.

Government intervention in urban development is not limited to broad master plans, but also includes many targeted programs. First, public intervention in the housing market is very important. Social housing represents nearly a quarter of the total housing stock in the Paris Region. This direct investment in the housing market is in addition to strict land use regulations. In addition, a major part of social housing is built in a limited number of neighborhoods delimited by public authorities in the 1960s ("*Zones à urbaniser en priorité*"). The spatial concentration of social housing in a limited number of municipalities leads to a spatial sorting on income. The resulting spatial mismatch and negative neighborhood effects called for additional public policies. Since the 1980s, the French government implemented numerous place-based policies to support distressed neighborhoods: employment zones (*"Zones Franches urbaines"*), urban renewal fund (*"Plan national de rénovation urbaine"*), Priority Security Zone (*"Zones de sécurité prioritaires"*) with extra funding for the police, Urban contract for social cohesion (*"Contrats urbains de cohésion sociale"*) securing subsidies to local authorities in order to improve education and access to employment and so on. Place-based policies do not only target disadvantaged neighborhoods. The business district of *La Défense*, one of the largest in Europe, is the result of a specific public intervention. In order to promote innovation, the central government has also supported the development of business clusters (*"Pôles de compétitivité"*), by subsidizing companies and universities located inside. The overlapping of all these public policies makes it very difficult to evaluate the effect of a specific policy on the overall urban development of the Paris region. This is one of the reasons why we focus on home-work flows, which offer more options for identifying causal effects.

### 2.1.2 Transportation network improvements

The improvement of transportation is one of the most important public policies of the second half of the 20th century in the Paris region. The two major investments focus on the highway network (613 km in 2014) and the *Regional express rail* (*"Réseau Espress Régional"* (587 km in 2014), a mass transit system between subway and commuter train. Figure 2.1 displays the rail and road networks in 2014. We observe that the road network is highly meshed while the rail network is still based on a hub-and-spoke model, centered on Paris. These maps suggest that the road network is more likely to promote a low density urban sprawl and decentralized commuting patterns, while rail is more usually associated with "ribbon" sprawl along rail lines.

**Figure 2.1:** Rail and Road Transportation Networks in Ile-de-France.




Left: Commuter-rail network, *RER* lines in colors; Suburban train (*Transilien*) lines in gray.
Right: Road network, green:municipal, orange:department, red:national.

**Regional Express rail**

Before the Second World War the Paris region was already the subject of various urban project, aiming at upgrading the regional rail network. They were but never put in place. The SDAURP plan 1965, initiated by President de Gaulle, planned the inception of a brand new rail network to connect the future regional system of new towns severed by fast and high-capacity trains, the *Regional Express Rail* system.

Due to budget constraints, subsequent president Pompidou decided to only upgrade the existing network of suburban trains, limiting the construction of new railroads to the strict necessary. It led to improving some existing routes and connecting them to airports and new towns with few additional branch line. The whole network efficiency was nevertheless improved by new trains and higher frequencies but only a few dozen new stations were built. Indeed, a major feature of the new network has been to connect Paris' railway stations together by tunneling *underground tracks* to connect them in order to design transversal lines: East-West (line A), North-South (line B), West-South (line C), and North-East (line D).[6] It has allowed significant decline in travel time, especially to access the center of Paris and to travel under Paris between suburbs.

In 2014, the RER network account 257 stations, among them 33 are located in Paris intra-muros. The *Regional Express Rail* system corresponds to a substantial part of Paris supply in public transportation and one of the largest Mass-Transit system in the world. In addition it is well connected to the 220km subway network as well as the rest of suburban trains that were left aside of the RER project, latter rebranded *Transilien*, that runs 1,299 km.[7] As a comparison, the *Shanghai Metro* runs over 700km, London *Underground* is 402 km long and London *Overground* 167km,[8] while New-York City *Subway* runs 380km.[9]

**Highway network**

In the 1960s, Paris region, as other Western cities, experiences a driving boom, with a dramatic increase in the use of individual vehicles. In France, car ownership went from 2.3 millions in 1950, to 6.7 millions in 1960 and 13.7 millions in 1970. However, in 1960, France was only equip with 10 kilometers highways (very low compare to other European continental countries like Germany of Italy). President de Gaulle decided to improve the French road network to foster economic growth and inter-city connections. The Motorway Master Plan (*"Plan directeur autoroutier"*) established the design of a brand new network of national highways to connect the French system of cities, and especially Paris with the other French cities. At the national level, the highway network reach 3,100km in 1975, 6,900km in 1990 and 11,800km in 2020. While it did not first intent to promote car commute, it eventually end up being used for daily commute, which foster urban sprawl (as in the

---

6 These tunnels are between *Nation - La Défense; Gare du Nord - Denfert; Gare du Nord - Gare de Lyon; Gare du Nord - Gare Saint Lazare; Invalides-Orsay*. Note that *Gare Montparnasse, Gare de l'Est* and *Gare de Bercy*, and their suburban trains have not been connected to the RER system.

7 Note that in the empirical study we will take into account the whole set of Paris region commuter-rail network, stations and lines, including the métro and the Transilien. It corresponds to 811 rail stations plus 176 tramway stops. It is worthy mentioning that those commuter-rail network gives also access to trans-regional trains (TER) and city-to-city express trains (Intercité).

8 Note that Greater London also benefit from 2,129km of routes with the *Great Western Railway* and 778km with the *Thameslink*.

9 An important remark is that New-York Metropolitan Area account 3,897km more with the additional transport facilities (*NJ Transit*, *Metro North Railroad*, and *Long Island Railroad*).

US, see Baum-Snow, 2007).

In addition to these radial highways, three ring-highways are built in the Paris region (*Périphérique* on the historical boundaries of the city of Paris, the A86 and, the farthest from the center of Paris, the *Francilienne*). The purpose of these circular highways was to facilitate commutes by car, as well as to speed up the crossing of the Paris region for inter-city transport. Unlike the public transportation system, based on a hub-and-spoke model, centered on Paris, ring-highways are likely to encourage the development of sub-centers, by-passing the monocentric organization of the metropolis. In 2014, the Paris region accounts 613 kilometers of highways, with around 300 access ramps.

### Rail, car and commuting

Table 2.3 shows that investments in transportation networks result in a sharp decrease in travel time between municipalities of the Paris region, over the period 1968-2014. The reduction in road travel time is much greater than in public transport travel time. This is due to the fact that, while both networks are of similar length, the RER is mainly an improvement of existing infrastructure, while the highway network is an entirely new infrastructure. We also note that the travel time within Paris is practically unchanged, whatever the means of transportation.

**Table 2.3:** Duration by mode

| | | **Duration (min.)** | | | | **Weighted duration (min.)** | | | |
|---|---|---|---|---|---|---|---|---|---|
| Origin/dest. | Orientation | Rail | | Road | | Rail | | Road | |
| | | 1968 | 2014 | 1968 | 2014 | 1968 | 2014 | 1968 | 2014 |
| Paris ↔ Paris | Any | 33.3 | 33 | 6.2 | 6.2 | 31 | 16.7 | 5 | 2.4 |
| Paris → Suburbs | Centrifugal | 115.6 | 113.1 | 47.5 | 40 | 51.5 | 56.1 | 12.4 | 15 |
| Suburbs → Paris | Centripetal | 116.6 | 114.2 | 47.5 | 40.1 | 59.6 | 64.4 | 17.1 | 18.8 |
| Suburbs ↔ Suburbs | Centrifugal (high) | 181.4 | 175.7 | 73.6 | 59.1 | 54.1 | 70.5 | 11.3 | 15.9 |
| | Centrifugal (low) | 185.5 | 181 | 64 | 53.5 | 54.2 | 67.1 | 11.2 | 15 |
| | Centripetal (high) | 182.1 | 176.2 | 73.4 | 59.2 | 56.6 | 78.3 | 13.3 | 19.5 |
| | Centripetal (low) | 186 | 181.3 | 64 | 53.5 | 56 | 68.4 | 12.3 | 16.3 |
| | Peripheral | 177.1 | 175 | 42.9 | 40.5 | 47.9 | 56.2 | 8.6 | 11.8 |
| | Same municipality | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |

On the contrary, the mean travel time between municipality weighted by the number of commuters, increase over the period. These findings are in line with the results of the French mobility surveys on which a small sample of people reports their commuting time. François (2010) notices a rise of commuters' "time budget" due to an increase in the distance between home and place of work that exceeds the time saved by faster transport. It suggests an endogenous relocation of both residents and jobs that benefits from better infrastructure by moving to more remote locations, fostering urban sprawl and large-scale commuting, but also the attraction of additional residents and jobs in the Paris region. We also notice that the travel time by road is much lower than the travel time by public transport. This is due to the fact that our calculation method does not take into account traffic jams (see subsection 2.2.2 for details on how congestion is taken into account in our econometric model and table 2.8 page 126 in Appendix for descriptive statistics on speed).

Table 2.4 shows that the modal share of public transport strongly depends on the orientation of the home-to-work trip in reference to Paris. In 2014, 75.3% of workers commuting from suburbs to Paris use public transport, compared to only 24.2% for peripheral suburb-to-suburb trips. Public transport seems to reinforce the monocentricity of the Paris region as the increase in the number of commuters using public transport is always greater than the increase in the number of commuters using cars for trips to or from Paris. In contrast, for suburb-to-suburb trips, growth in the number of commuters using public transit is the smallest for peripheral trips. This is probably due to the fact that public transport infrastructures often require to transit through Paris. Table 2.9 (page 126 in Appendix) indicates that the ratio between public transport route and the straight line distance between two municipalities increases as the peripherality index rises, making it less efficient. Conversely, car trips do not require more detours, whether they are radial or peripheral.

**Table 2.4:** Commuting flows by mode

| | | **Modal share (%)** | | | | **Commuters** | | | |
|---|---|---|---|---|---|---|---|---|---|
| Origin/dest. | Orientation | Rail | | Road | | 2014 | | 2014/1968 | |
| | | 1968 | 2014 | 1968 | 2014 | Rail | Road | Rail | Road |
| Paris ↔ Paris | Any | 0.376 | 0.62 | 0.071 | 0.079 | 464,416 | 58,874 | 1.252 | 0.843 |
| Paris → Suburbs | Centrifugal | 0.614 | 0.699 | 0.249 | 0.22 | 224,128 | 70,453 | 2.403 | 1.862 |
| Suburbs → Paris | Centripetal | 0.702 | 0.753 | 0.208 | 0.184 | 697,415 | 170,420 | 1.369 | 1.128 |
| Suburbs ↔ Suburbs | Centrifugal (high) | 0.376 | 0.34 | 0.303 | 0.606 | 87,960 | 156,821 | 3.268 | 7.233 |
| | Centrifugal (low) | 0.345 | 0.319 | 0.315 | 0.62 | 103,016 | 200,341 | 3.035 | 6.467 |
| | Centripetal (high) | 0.404 | 0.365 | 0.335 | 0.592 | 233,710 | 379,129 | 3.057 | 5.975 |
| | Centripetal (low) | 0.366 | 0.306 | 0.354 | 0.637 | 170,758 | 355,435 | 2.324 | 4.991 |
| | Peripheral | 0.292 | 0.242 | 0.33 | 0.686 | 146,885 | 416,408 | 2.146 | 5.393 |
| | Same municipality | 0.049 | 0.195 | 0.079 | 0.385 | 179,937 | 354,647 | 3.152 | 3.824 |

## 2.2 Data

In our study we choose to focus on the whole Paris region, called *Ile-de-France*, for two main reasons: first it is a stable geographic unit over decades and second it corresponds to a relevant administrative unit for the implementation of urban planning policies. French data are particularly detailed and allow to study the evolution of urban development at a quite precise geographic scale. The smallest geographic unit, the "*Communes*" or municipalities, date back to the French revolution (1789), where they corresponded to parishes, and has not changed much up to now, since municipal mergers remained rare in France, compared to other European countries. This implying that the number of municipalities (or the 20 *arrondissements* of the city of Paris) in Paris Region is both large (1,278) and stable over time.

### 2.2.1 Population and Commuting Flows

At each Census, the French national statistics institute (INSEE) collects both the residence and the workplace, for a (randomly selected) quarter of person in employment, at the municipality level. This allows us to calculate commuting flows between any two municipalities in the region[10] for each census year between 1968 and 2014 (1968, 1975, 1982, 1990, 1999, 2009, and 2014). In addition, the Census provides the means of transportation for commuters, but only in 1968, 1999 and after. Besides, we also calculate commuting flows for specific segments of the population (occupational category, gender, business sector) in order to explore a possible heterogeneous effect of transportation.

### 2.2.2 Travel Times and Commuting Times

Yet small geographic unit, travel time between two municipalities is not unique but depends on the precise location of both housing and workplace. We cannot take into account this heterogeneity, since our data does not provide a more precise location than the municipality scale, for both homes and jobs. We thus calculate travel time, using the location of town-halls ("*Chef-lieu*"). Town-hall is often located in the dense and long-standing urbanized part of the municipality, and seems to be the best approximation to locate both residents and workers at a unique point.

**Travel time with car**

The travel time by car is calculated with a online routing engine of the French geographic institute (IGN)[11] that provides the distance and duration of the trip, as well as the exact route between the departure and arrival. First, we retrieve precise routes between each pair of municipalities in the Paris region, using town-halls as a point of departure and arrival. Secondly, by overlaying the set of calculated routes, we recover the set of elementary segments of the road network by identifying common parts of the different trips. We also retrieve the traveling speed on each of these segments[12]. Third, we flag the opening date of each highway segment, using a collaborative website (`routes.fandom.com`). We lastly run the standard Dijkstra shortest path algorithm based on this road network graph, considering only segments opened for a given Census year. As the highway network develops, faster connections are possible and travel time decreases for most of the dyads. We obtain a database containing 11,594,583 observations (1,287 origins $\times$ 1,287 destinations $\times$ 7 years).

The construction of a complete traffic model is beyond the scope of this chapter. Nevertheless, it is not possible to ignore road congestion when looking at car travel in a city as large as Paris. We have already pointed out that the travel time based on the maximum allowed speed is much lower than the real one. To tackel down this issue, we build a proxy that we call "potential car traffic" based

10 Note that we drop workers for which either the residence or the workplace do not locate in Paris region
11 based on the BD TOPO database.
12 We notice that the IGN routing engine does not use the actual maximum speed but a theoretical estimate based on the size of the road, the administrative category and whether it is in or out of town. This one-size-fits-all approach is not well adapted to the Paris region where speed limits are often stricter than in the rest of the country. For example, this theoretical speed amounts to more than 100 km/h on the Paris ring road, whereas it has never exceeded 90 km/h and is now limited to 70 km/h. We thus lower all freeway speeds to 80 km/h in Paris and 100 km/h in the inner suburbs, and to 50 and 60 km/, respectively, on other roads.

on the hypothetical situation where all commuters use the car. We then calculate the total car flow of each segment and finally the mean car flow of each route between two municipalities.

**Travel time with public transport**

Travel time with public transport is mainly based on the rail network: subway (*Métropolitain*), suburban train (*Regional Express Rail* and regular service) and tramway. In the same way as road network, we calculate the travel time between each pair of municipalities for each Census year over the period 1968-2014. We base our calculation on open data displayed by the two public companies running Paris transport network (RATP and SNCF). These data provide us with the time spent in the train[13], the tramway or the subway, but also the transfer time between two lines in a station. As for the road network, we flag the opening date of each public transport service.[14] Since we do not observe the frequency of service, the waiting time is set to a uniform duration (3 minutes). Lastly, unlike the car journeys that can start directly from the center of each municipalities, public transport trips necessarily begin and end with another means of transportation, to reach a train, metro or tram station. Since we do not observe the evolution of the bus network, we make basic assumptions based on the road itineraries between municipality centers and the neighboring stations. The road distance is calculated with the IGN routing engine. We approximate the bus travel time by applying a uniform 12 km/h speed and adding a uniform waiting time of 12 minutes. We also take into account the fact that workers can bike or walk. For this purpose, we use a lower travel speed (6 km/h), but without waiting time. As a result, trips of less than about 2 km are made on foot or by bike, and longer trips are made by bus.

In short, travel time by public transport can be categorized into four different types of duration : (i) time spent in the train, (ii) the transfer time between two lines in a station, (iii) the waiting time in the station, (iv) travel time on foot, by bike or by bus to reach a station. Only duration (i) vary in time, but duration (iv) can also change for a given route since the worker can select another departure or arrival station, after the introduction of a new public transit service. We combine the four different types of duration to compute the public transport travel time, using Dijkstra shortest-path algorithm.[15] Note that this approach allows us to calculate a travel time for all 1,656,359 dyads (1,287 origins $\times$ 1,287 destinations), including those where there is no train, tram or subway station in departure or arrival municipality.

## 2.3 Empirical Strategy

We have described the two major infrastructure programs that have been implemented in the Paris region during the second half of the XXst century. Simultaneously the metropolis has experienced dramatic changes in its urban organization, with the rise of urban sprawl and a shift in commuting patterns. We investigate in what follow if these infrastructure improvements have caused this

[13] Note that within the network over a line some travel stop to every stations, while other travel over the same line jump some stations in order to reduce travel time between remote but important stations. We take into account such jumps.

[14] Note that the changes in travel duration between 1968 and 2014 are only due to these openings, and not to speed increases on existing lines.

[15] This approach, while standard in the transportation literature, is not the only one. Allen and Arkolakis (2021) propose an approach that relies on the adjacency description of the network to compute *path-integral* formulation of travel time. As if commuters were randomly picking commuting routes according to their relative speed.

suburbanization process and what were their respective contributions in individuals' commuting choices.

In this section, we present micro-found our empirical framework. This approach allows to structurally interpret our estimated coefficients. Our goal is to study the incidence of the reduction in travel time on commuting flows, accounting for potential confounding factors. Indeed the implementation choice of infrastructure investments is not random and targeted investment might bias naive estimates with reverse causality. Furthermore, place-based policies, such as public housing or free-tax zones, can be correlated with infrastructure investments, resulting also in reverse causality. Finally we describe how we account for travel time, allowing alternative modes, to define a generalized cost of travel.

### 2.3.1 Theoretical Background

We have shown that, if average travel time within the metropolis has decline due to major infrastructure investments, individuals' commuting time, on average, have increased substantially over the period. This suggests endogenous relocation of both residents and jobs. We define here a structural gravity equation, to state the incidence of travel time on commuting flows, that accounts for this general equilibrium effects. Once we observe commuting flows and travel time, this approach only need to estimate a three-way fixed effects equation to describe the effect of the infrastructure improvements on the organization of the metropolis.

#### Setup

We consider a closed economy, which corresponds to a single metropolis, populated by a mass $L$ of individuals that defined an integrated labor market. The metropolis is comprised of a large number of locations, indexed $i$,$j$ or $k \in N$.[16] where workers can reside or work. Those locations are characterized by their population of residents, $R_i$, and the total number of occupied jobs, $L_j$. They also offer a specific level of amenity, $u_i$, enjoyed by residents, and a specific productivity shifter, $a_j$. Each location supplies a fixed quantity of land, used for housing, $h$, and host a single representative firm that contribute to produce the numéraire $c$ of the economy, offering a competitive market-level wage, $w_j$, to workers.[17]

Transportation infrastructure network allow workers to commute within the metropolis against commuting cost $\kappa_{i,j}^m$, where $m \in M$ denotes the transportation mode. We speak of costs by abuse of language, $\kappa_{i,j}^m$ actually corresponds to a dis-utility since it discount utility. As we will describe later in this section, $\kappa_{i,j}^m$ will furthermore corresponds to a generalized cost of commuting, including travel time and opportunity costs.

The continuum of workers $\omega \in \Omega$ that inhabits the metropolis enjoy utility $U_{i,j}^m(\omega)$ by choosing to reside in location $i$, while working in $j$. Their utility is assumed Cobb-Douglas, with $\beta$ the share of housing spending in individuals' budget. Workers' maximisation problem is thus defined as follow:

16 This discretization of the geographic area is an empirical requirement. It also reflect the occupation constraint of space. These blocks can corresponds to *census tracks*, the finest level of data availability. In the French context it will correspond to *municipalities* (*communes*, in French), which are sufficiently granular to describe commuting pattern.

17 Production in this model can be seen as location-representative firms producing a differentiated variety of a freely-traded good (Armington condition) that are aggregated by consumers in a constant elasticity of substitution way to form the numéraire of the economy.

$$U_{i,j}^m(\omega) = \max_{(c,h),(i,j),m} \left(\frac{c}{1-\beta}\right)^{1-\beta} \left(\frac{h}{\beta}\right)^{\beta} \frac{u_i}{\kappa_{i,j}^m}\ z_{i,j}^m(\omega)$$

where individuals consume housing $h$ and a freely-trade composite good $c$. $u_i$ is an utility shifter freely provided by location $i$ to its residents. In addition, individuals observe an idiosyncratic draw $z_{i,j}^m(\omega)$ for each possible residence-workplace pairs $(i,j)$ by commuting mode $m$.[18]

To express the associated indirect utility faced by workers, let assume that rent price $q_i$ is paid to reside in $i$ and consume $h$ and that working in $j$ provide a wage $w_j$ shifted by location-specific productivity $a_j$. Since utility is assumed Cobb-Douglas, that we defined the set of endogenous price that controls the urban equilibrium in each location, $\{q_k, w_k\}_{k\in N}$, knowing commuting costs $\{\kappa_{i,j}^m\}_{\forall(i,j)\in N^2}$, solving for the workers' problemyields the following indirect utility function:

$$v_{i,j}^m(\omega) = \frac{u_i w_j a_j}{q_i^{\beta}\,\kappa_{i,j}^m}\ z_{i,j}^m(\omega)$$

**Commuting probability**

Lets now assume that idiosyncratic preferences are drawn from a Fréchet distribution, $F(z) = e^{z^{-\epsilon}}$. We normalize the scale parameter to one, as it enters isometrically to residence and workplace attracitivity. The shape parameter, $\epsilon > 1$, regulates the degree of preference heterogeneity among individuals. The smaller the shape parameter, the greater the heterogneity in idiosyncratic preference, and the less sensitive are individuals' location choice to economic variables (rent, wage, commuting cost, amenity or productivity).

Integrating over the idiosyncratic preference shocks gives the probability to chose among residence-workplace pairs by mode. Under the Fréchet assumption, the probability $\pi_{i,j}^m(\omega)$ that a worker chooses to reside in $i$ and work in $j$ is given by:[19]

$$\mathbb{P}\left[v_{i,j}^m(\omega) \geq v_{i',j'}^m(\omega)\right] \equiv \pi_{i,j}^m(\omega) = \frac{(u_i w_j a_j)^{\epsilon}(\kappa_{i,j}^m q_i^{\beta})^{-\epsilon}}{\sum_{i',j',m'}(u_k w_l a_l)^{\epsilon}(\kappa_{k,l}^m q_{R_k}^{\beta})^{-\epsilon}} \tag{2.1}$$

The commuting probability between locations $i$ and $j$ depends on the relative attractiveness of $i$ as a residential location and the relative attractiveness of $j$ as a workplace, which depends on their level of amenities, respectively $u_i$ and $a_j$, and endogenous prices, $q_i$ and $w_j$.It also depend on the bilateral commuting cost $\kappa_{i,j}^m$, which depends on the transportation mode $m$, and a multilateral resistance term (the denominator) that depends on the cost of all alternative bilateral commuting pairs.

18 Note that it introduces heterogeneity in the labor supply for two purposes in the model. First, if this idiosyncratic component is drawn from a Fréchet distribution, then we will recover standard results for multinomial choices with a multiplicative description of flows (see Eaton and Kortum, 2002). Second, it allows to provide a more realistic economic model, since heterogeneity arise among workers for different reason, e.g. productive matching between $\omega$ and $j$, individual-specific amenity preferences between $\omega$ and $i$, ..

19 Replacing $z$ in $F(z)$ according to the indirect utility expression leads to $G(v) = \mathbb{P}\left[v_{i,j}^m(\omega) \leq v\right] = 1 - \exp\left(-\left[v\frac{q_i^{\beta}\kappa i,j^m}{u_i w_j a_j}\right]^{\epsilon}\right)$. The chosen commuting pair with transportation mode is then the maximum of a sequence of Fréchet distributed random variables, which is itself a Fréchet. Such that given the fact that a worker's commuting choice $(i,j,m)$ corresponds to her probability to realize a higher utility in $(i,j)$ with commuting mode $m$ than in any other commute $(i',j',m') \neq (i,j,m)$, the probability $\pi_{i,j}^m(\omega) \equiv \mathbb{P}\left[v_{i',j'}^m(\omega) \leq v_{i,j}^m(\omega), \forall(i',j',m') \neq (i,j,m)\right]$ that the worker chooses the residence-workplace pair $(i,j)$ and mode $m$.

In light of previous results from the trade literature (Anderson and van Wincoop, 2003) we can define the two multilateral resistance terms that shows up in the urban network flows:

$$\Pi_{k,m}^{-\epsilon} = \sum_{j \in N} \left( \frac{w_j a_j}{\kappa_{k,j}^m} \right)^{-\epsilon} \quad \text{and} \quad P_{k,m}^{-\epsilon} = \sum_{i \in N} \left( \frac{u_i q_i^{-\beta}}{\kappa_{i,k}^m} \right)^{-\epsilon}$$

where $\Pi_k^{-\epsilon}$ corresponds to the "inward" resistance term, which is the relative attractivity of $k$ as a residence location , and $P_{k,m}^{-\epsilon}$ the "outward" resistance term, which corresponds to the relative attractivity of $k$ as a workplace.

These two terms allow to rewrite the commuting probability define in equation (t yields a *structural gravity equation*,[20] with $l_{i,j}^m$ the commuting flows between $i$ and $j$ using transportation mode $m$ defined as follow:

$$l_{i,j}^m = \pi_{i,j}^m L = \frac{R_i}{\Pi_i^{-\epsilon}} \frac{L_j}{P_j^{-\epsilon}} (\kappa_{i,j}^m)^{-\epsilon} \tag{2.2}$$

It provides a sufficient statistics that characterized the urban equilibrium by the magnitude of flows between places within the metropolis. These flows depends on the relative accesibility and demography of both the places of origin and destination of the flow, as well as the bilateral commuting costs to go from origin to destination. It exhibits an explicit multiplicative form in term of the spatial distribution of economic opportunities ($R_i$ and $L_j$ the number of residents and jobs) and the topography of the metropolis defined by the transportation networks (bilateral $(\kappa_{i,j}^m)^{-\epsilon}$ and multilateral commuting costs $\Pi_i^{-\epsilon}$ and $P_j^{-\epsilon}$). Commuting flows are hence expected to be larger between locations with more residents and more jobs ($R_i, L_j$), while they are expected to decrease between locations that are more distant ($\kappa_{i,j}^m$) from each other or remote compare to the rest of the metropolis ($\Pi_i, P_j$).

This sufficient statistics show that the local incidence of infrastructure improvements can be decomposed into two components: a direct effect (the variation in bilateral cost) and an indirect effect (the variation of the multilateral terms). In addition to the direct access to residents or jobs, it is thus as if any residential location faces a unique labor demand incidence of network's commuting costs $\Pi_i^{-\epsilon}$, which summarizes its overall access to attractive workplace (job amenity times wage discounted by commuting costs).The multilateral resistance terms provides a theoretically consistent way to aggregate commuting costs at any location in the city-economy, relative to workplaces' price index. This will proved useful to structurally interpret the coefficient estimated in the reduced-form (see section2.5). Conversely, it is as if any workplace faces a specific and unique labor supply incidence of commuting costs $P_j^{-\epsilon}$, which corresponds to the potential of "imported" workers from outside locations.

**Generalized costs of transport**

We now come back on the definition of the generalized costs of transport $\kappa_{i,j}^m$, left aside in what precede. First, using a transportation mode $m$ requires to pay a fixed cost $\alpha_m$, to buy a car for

[20] See Anderson 2011 or Head and Mayer 2014 for a discussion on structural gravity and an overview of the approaches that lead to such equation.

instance. Second, the value of time $\tau_{i,j}^m$ spend in transport $m$ is $\phi^m \tau_{i,j}^m$, with $\phi^m$ the value of time. Third, given the origin $i$ and the destination $j$, it exists alternative modes. Commuters hence account for the opportunity cost of not using the alternatives. The opportunity cost encompass a fixed costs, which is accounted for in $\alpha_m$, and the alternative cost of travel time with mode $m'$, that is $\theta^{m'} \tau_{i,j}^{m'}$, with $\theta^{m'}$ the value per unit of time. In addition, potential traffic matters for commuters when they commute by car. Therefor we also include a term $p_{i,j}$ that captures the potential peak of traffic a commuter can meet along the way (and that do not depend on the specific flows between $i$ and $j$). This peak corresponds to the segment with the maximum potential commuters (that is accounting for other commuting within the metropolis that may congest the path). Finally, we include origin-destination specific time-invariant characteristics, $\delta_{i,j}$, like the distance as the crow flies, elevation or past infrastructure investment. This decomposition of the different factors that translate commuting into an (dis)utility (cost) component defines the generalized costs of transport:

$$\kappa_{i,j}^m = \alpha_m \times (\tau_{i,j}^m)^{\phi^m} \times (\tau_{i,j}^{m'})^{\theta^m} \times (p_{i,j})^{\psi^m} \times e^{\delta_{i,j}}$$

### 2.3.2 Estimation Method

The derivation of the theoretical model of the metropolis allows to show that the urban equilibrium is fully described by equation (2.2). We show in what follow how to properly estimate it and explicit our identification hypothesis.

To estimate the structural gravity equation, we interpret observed data on residence-workplace commuting as a finite sample from the continuum probabilistic model. Given the observed bilateral commuting costs $\{\kappa_{i,j}^m\}_{\forall(i,j,m)}$, the equation parameters can be estimated by maximum likelihood. The log-likelihood function derived equation (2.2) is [21]

$$\ln \mathcal{L} = \sum_{i,j,m} l_{i,j}^m \, \ln(\pi_{i,j}) = \sum_{i,j,m} l_{i,j}^m \, \ln \left[ \frac{R_i L_j}{L} \frac{(\kappa_{i,j}^m)^{-\epsilon}}{\Pi_{i,m}^{-\epsilon} P_{j,m}^{-\epsilon}} \right] \tag{2.3}$$

As shown by Guimaraes et al. (2003), the maximum likelihood function in equation (2.3) is numerically equivalent to a Poisson pseudo-maximum likelihood (PPML) estimator[22] with the following specification:

$$l_{i,j}^m = \exp \left[ \delta_{i,m}^R + \delta_{j,m}^L - \epsilon \left( \phi^m \ln(\tau_{i,j}^m) + \theta^m \ln(\tau_{i,j}^{m'}) + \psi^m \ln(p_{i,j}) + \delta_{i,j} \right) \right] \times \varepsilon_{i,j}^m \tag{2.4}$$

where we explicit the generalized commuting cost, defined previously. In the empiric specification, the mode-specific fixed-cost boil out in the oriented location-specific fixed effects, since we assume no variation of the fixed-cost across locations of the metropolis we are not able to estimate mode-specific fixed-cost separately. The oriented fixed effects capture both location demography, including the incidence of place-based policies, and the network incidence of commuting, with $\delta_{i,m}^R = \frac{R_i}{\Pi_{i,m}^{-\epsilon}}$ and $\delta_{i,m}^L = \frac{L_j}{P_{j,m}^{-\epsilon}}$. Finally, $\varepsilon_{i,j}$ is the stochastic error term that arise from

[21] Note that it corresponds to the canonical conditional-logit likelihood of McFadden (1974) applied to multinomial location choices, as in McFadden (1978).

[22] Note that the PPML does not require the depend variable to be Poisson distributed (Santos-Silva and Tenreyro, 2006). The estimation procedures is fairly easy to implement and robust to misspecifications (Gourieroux et al., 1984).

data observation and that is assumed to not carry any systematic information about trade costs.[23] PPML estimator for gravity regressions allows to account for heteroscedasticity and to leverage the information contained in the zero commuting flows. Finally, note that our estimates of the generalized cost of transport will include the level of idiosyncratic preferences among workers, $\epsilon$. Since we assume that it is the same across modes, we are able to compare generalized costs estimate, up to this constant.

### 2.3.3 Identification

Equation (2.4) explicit the specification of our empirical strategy that derived from the theoretical model. We now discuss the identification hypothesis that allows us to estimate the causal relationship between the infrastructure improvements and additional commuting flows. In practice we rely on panel data and the variables in (2.4) are time-varying, except $\delta_{i,j}$. We do not introduce time previously for the sack of clarity. The introduction of time variations implies that the urban equilibrium described before is not static. It follows the changes in travel time induced by infrastructure improvements, as well as location-specific amenities, productivity and housing supply that changes over time. Since our framework is a characterization of these urban equilibria, the estimation is done by comparing different urban equilibrium of the same metropolis over time.

There is no doubt that policy makers clearly intended to treat specifically some dyads. One of the stated goals of the Regional Express Rail project was to connect the city of Paris, downtown area, to the New Towns (*Villes nouvelles*), in the outskirts. We can argue that, even if public authorities built housing in *Cergy-Pontoise*, new town, office towers in *La Défense* and rebuild a regional rail line between the two places, there is no specific way to bias the functioning of the labor market that is the number of commuters by residence-workplace pairs. Public authorities cannot (and probably do not want to) promote the hiring of *Cergy-Pontoise* inhabitants by companies located in *La Défense*. But the causal estimation of transportation effect will be (upward) biased if new infrastructures come together – more or less systematically – with urban development projects at each end. Therefor we state clearly below the five sources of bias that can threat identification in our setting. We then describe how our approach addressed these different issues to provide a convincing identification of the parameters of interest. Our approach can be seen as a difference-in-difference estimation with fuzzy design, where we compare different level of treatment among dyads, treatment being the variation in travel time induced by the improvement of the infrastructure. Causal interpretation hence derived from the standard "parallel trends" assumption being satisfied.

In our network setting, different sources of bias are threat to the identification. First, travel time is a product of both distance and speed. The number of residents in a given location who work in an other one is strongly correlated with the distance from the two locations, and since travel time, by both rail and road, is determined first and foremost by distance, the (strong) correlation with commuting flows can in no way be considered as a causal impact. In addition, speed depends strongly on infrastructure investments, which are determined by existing infrastructures and public

[23] Especially the identification with the PPML estimator is ensured under the condition that $\mathbb{E}[\varepsilon_{i,j}|\delta_i^R, \delta_j^L, \kappa_{i,j}] = 1$. Note that the error term could also enter in an additive way, and the estimation would be valid too (see Santos-Silva and Tenreyro 2006): $y_i = exp(X_i\nu) + \eta_i$ can be expressed as $y_i = exp(X_i\nu) \times \zeta_i$, with $\zeta_i = 1 + \eta_i/exp(X_i\nu)$ ($\zeta_i \geq 0$) and $\mathbb{E}[\zeta_i|\mathcal{X}] = 1$ instead of $\mathbb{E}[\eta_i|\mathcal{X}] = 0$. Actually, defining economic quantities in terms of $\mathbb{E}[y|X]$ circumvents the issue of whether the 'disturbances' in the model are additive of multiplicative (Wooldridge 1992).

investments (targeted investment). Second, the incidence of infrastructure improvements on travel time are spatially correlated, since segment improvements, of rail or road, can benefit different residence-workplace pairs if their connections rely on a mutual segment. Therefor commuting flows are interdependent via the network structure of transportation infrastructure. In addition, it implies, third, that the disutility of travel time between two locations is actually relative to outside opportunities. We do not expect the same incidence of a mass-transit infrastructure in a central locations compare to a remote one. For the same population in the two locations, it should generate more flows from the latter compare to the former. Fourth, location specific demography dynamics can bias the estimate of commuting costs if infrastructure improvements are correlated with place-based policies that shift either the local provision of residential amenities (e.g. park and recreations), productivity (e.g. free-tax zones) or housing supply (e.g. social housing programs). Fifth, their is a sample selection issue in the data, since most of the dyads display zero commuting flows.

The "three-way" fixed effects specification, which derived from the theoretical model, actually allows to get ride of these different identification issues. First, the residence-workplace fixed effect control for any dyad-specific time-invariant characteristics (observable or not and that do not changed over time), which would be correlated with travel time reduction like distance or targeted investments. We thus estimate the effect of a *change* in travel time -that do not depend on dyad-specific characteristics- on a *change* in the number of commuters. It also captures spatial auto-correlation, the second source of bias described above, since the relative location of the dyad in the network enters the dyad-specific characteristics, captured by the fixed effects. Third, the dynamic location fixed effects, $\delta^R_{i,t}$ and $\delta^L_{j,t}$, captured potential general equilibrium bias. Equation (2.2) makes clear that network interdependence boils out in the two multilateral terms, $\Pi^{-\epsilon}_{i,t}$ and $P^{-\epsilon}_{j,t}$, that finally only depends on either the residential location $i$ or the workplace $j$. Actually these two terms discount the bilateral commuting costs and thus control for the dyad relative remoteness. In addition, the two location-specific fixed effects also capture location-specific demographic dynamics, the fourth source of bias. They indeed control for time-varying idiosyncratic shocks at the origin and the destination of the dyayde. For example, the construction of the business district of La Défense is a strong positive shock on jobs in municipalities where it is located (*Puteaux*, *Nanterre* and *Courbevoie*). Conversely, the establishment of New Towns is a massive shock on housing market of formerly rural villages.[24] Such shocks increase (or reduce) the supply for commuters. Furthermore, such shocks are correlated with development of transportation infrastructures, due to the coordination of public policies. *La Défense* is one of the first RER stations, new towns are closely connected to the highway network. The gravity framework captured this mechanic increase of commuting flows with location size in the location-specific fixed effects. Finally the fifth threat is addressed by using PPML estimator that leverage the information contain in the zeros (see Santos-Silva and Tenreyro 2006).

One last thing that we need to discuss with the introduction of fixed effects is the incidental parameter problems (see Neyman and Scott 1948; Lancaster 2002). Such issue arise when estimation noise from estimates of fixed effects and other “incidental parameters” contaminates the scores of the main parameters of interest, inducing bias.[25] Weidner and Zylkin (2021) shows that the point estimates produced by three-way fixed-effects PPML estimator in gravity settings are actually

[24] the population of *Cergy*, that is one of the two centers of *Cergy-Pontoise* one of the new towns, blows from 2,895 inhabitants in 1968 to 48,226 in 1990.

[25] This bias can manifest in a variety of different ways. A generic characterization by Fernández-Val and Weidner (2018)

asymptotically consistent (while it does not hold for other PML estimators). It shows nonetheless that estimates are asymptotically biased, meaning that the asymptotic distribution of the estimates is not centered at the truth as the number of observation grows. In other words, it approaches the truth “at an angle” asymptotically. However, using analytical bias corrections to address each of these biases[26] do not change much our point estimates and, while larger, our confident intervals remains far from zero.

Finally, we challenge our identification strategy with several robustness checks (see section 2.4.3), estimating our model on sub-samples that are less likely to be concerned with the "intention-to-treat" issue (e.g. targeted investment, place-based policy). To achieve this, we apply standard approaches in transportation project evaluation. First, we exploit the differences between the network that was actually built and the one that was initially planed, changes emerging from budget restriction (similarly to Donaldson (2018)). Second, we restrict our sample to municipalities that were coincidentally connected to the network because they are located in between urban subcenters (following Banerjee et al. (2020)). Third we exclude dyads that are directly connected by the same line and we only keep routes need to transfer from one line to another in a hub station (similarly to Giroud (2013)), routes that are not directly treated by the policy maker.

## 2.4 Main Results

### 2.4.1 Baseline results

We now turn to the presentation of our main results. The estimation of the structural gravity equation (2.2) with the PPML estimor only requires to observe commuting flows and travel time between any two locations in the metropolis. Our baseline results are reported in Table 2.5. It corresponds to the specification described in equation (2.4), where we regress the total number of commuters, from municipality $i$ to municipality $j$ and mode $m$ at year $t$, on the generalized costs of transport that is composed of the travel time $\tau^{m}_{i,j,t}$ with mode $m$ (either car or public transport) at year $t$, the alternative travel time $\tau^{m'}_{i,j,t}$ offered by mode $m'$, the potential traffic on the road along the way $p_{i,j,t}$ and the time-invariant dyadic fixed effect. The columns in Table 2.5 corresponds to different specification of the generalized cost of transport.

The estimation confirms standard results in the literature that infrastructure improvement, by reducing travel time, caused additional commuting flows. With a simple definition of commuting costs with travel time (columns 1 and 2), we find a negative incidence of public transport time on public transport travel time (2.29), as well as a negative incidence of road travel time on the number of commuters using car (1.71). With our definition of the generalized cost of transport, the associated estimates for travel time are -2.42 for public transport and -1.22 for road. Both rail and road travel time are in minutes in our specification, so we observe that additional minutes by trains prevent more commuters than additional minutes traveled on roads. This is consistent with the observation of a slight decline of public transport use with commuting distance. In addition, more minutes by

gives: bias$(\hat{\beta}) = \frac{bp}{n} + o(p/n)$ and $std(\hat{\beta}) = \frac{c}{\sqrt{n}} + o(n^{-1/2})$, where $\hat{\beta}$ corresponds to the estimates of the parameter of interest, $n$ the total number of observations, $p$ the total number of parameters, inclusive of any fixed-effects. $b \in \mathbb{R}$ and $c > 0$ are constants that depend on the model being estimated.

[26] We use the Stata package `ppml_fe_bias`, provided by Weidner and Zylkin (2021).

**Table 2.5:** Regressions by mode

| | (1) | (2) | (3) | (4) |
|---|---|---|---|---|
| **Pub. transport commuters** | | | | |
| log(Pub. transport dura.) | $-2.29^{***}$ | | $-1.725^{***}$ | $-2.416^{***}$ |
| | 0.124 | | 0.114 | 0.107 |
| log(Road duration) | | $-1.703^{***}$ | $-1.323^{***}$ | 0.017 |
| | | 0.09 | 0.104 | 0.093 |
| log(Potential road traffic) | | | | $0.765^{***}$ |
| | | | | 0.014 |
| **Road commuters** | | | | |
| log(Pub. transport dura.) | $-2.254^{***}$ | | $-1.243^{***}$ | $0.418^{***}$ |
| | 0.137 | | 0.13 | 0.118 |
| log(Road duration) | | $-1.717^{***}$ | $-1.992^{***}$ | $-1.222^{***}$ |
| | | 0.117 | 0.092 | 0.086 |
| log(Potential road traffic) | | | | $-0.117^{***}$ |
| | | | | 0.012 |
| **Other commuters** | | | | |
| log(Pub. transport dura.) | $-2.95^{***}$ | | $-1.602^{***}$ | $-0.934^{***}$ |
| | 0.131 | | 0.123 | 0.116 |
| log(Road duration) | | $-2.81^{***}$ | $-2.561^{***}$ | $-1.567^{***}$ |
| | | 0.107 | 0.094 | 0.09 |
| log(Potential road traffic) | | | | $0.282^{***}$ |
| | | | | 0.015 |
| Pseudo R2 | 0.8929 | 0.8936 | 0.8954 | 0.9238 |
| No. obs. | 2,007,905 | 2,007,905 | 2,007,905 | 2,007,905 |
| No. fixed effects | | | | |
| Orig. × Dest. | 184,770 | 184,770 | 184,770 | 184,770 |
| Orig. × Year | 5,100 | 5,100 | 5,100 | 5,100 |
| Dest. × Year | 5,113 | 5,113 | 5,113 | 5,113 |
| Mode × Year | 11 | 11 | 11 | 11 |
| Years | 1968, 1999, 2009 and 2014 | | | |

Notes: The table provide the estimation results of equation (2.4). The columns corresponds to different specification of the generalized cost of transport, with column (4) corresponding to the baseline described in the theoretical part. Columns from (1) to (3) do not account for potential road traffic. In addition, column (5) interacts road travel time with potential traffic to capture the discounting factor of potential traffic on travel time.

public transport are also often associated with line switch. We also observe that potential road traffic matters for the number of car commuters. More potential traffic (form other dyads) is thus associated with less commuters using cars. In comparison with other studies, we lie in the range of the estimated values for the elasticity of commuting flows to travel time. While studying the implementation of steam railway in London in the XIXth century, Heblich et al. (2020) instrument travel time with distance and find a significant elasticity of -5.02. Looking at Berlin, Ahlfeldt et al. (2015) find an elasticity of -0.07 in cross-section from self-reported travel times for year 2008.

In addition, the opportunity cost includes in our specification of the generalized cost of transport, account for the travel time offered by the alternative commuting mode. The estimation Column (4) show for car commute that a reduction in the alternative travel time, public transports, is associated with less car commuters. It implies that investment on public transport reduced the attractivity of car to commute. However, quantitatively less than travel time with the used mode. This does not hold for public transport, once we control for potential road traffic. The incidence of road travel time boils out, while potential road traffic has a significant positive effect on commuters using public transport in the dyad. It suggests that potential congestion on roads increase the number of commuters relying on public transports, while the raw travel time by car does not affect the reliance on public transport. Indeed, since car use allows more flexibility in the commuting direction, travel time by car, *per se*, does not provide a sufficient statistics to predict public transport commuting flows.

Table 2.10 in appendix display the same results but with all type of commuting flows as outcome, without considering the transportation mode used to commute. It allows to include data from census survey where transportation mode were not reported. It display similar results, except that we cannot interpret any opportunity costs. The magnitude of public transport travel time effect on overall commuting flows is -2.08 and the effect of road travel time is -1.19, which pretty close to our baseline specification results. We will use this broader specification of commuting costs, with overall commuting flows (without distinguishing them by mode), in what follow for the sack of simplicity, since it does not alter our estimates compare to our baseline specification, and to be able to include all census years.

In Table 2.11 display we perform our baseline regression, but we introduce the different fixed effects successively.[27] First, it allows to observe the sharp decline in the number of observation that contributes to the estimation, going from 11.6M to 1.6M, since the model becomes pretty constraint with the three-way fixed effects specification.[28] Travel times are always negatively related to commuting flows in a significant manner. However, we observe contrasted evolution, while adding fixed effects, between public transport and car travel time. For the former, its incidence increase (in absolute value) with the number of fixed effects. Especially controlling for time-unvarying residence-workplace characteristics instead of residence and workplace-specific fixed effects increase the estimate in absolute value from .41 to 1.55, suggesting that investments are targeting more remote links and that controlling for dyadic fixed effects matters substantially for public transport commute. In complement, location-specific dynamics fixed effects only play a minor role, suggesting that place-based policy are not so important in driving flows or that infrastructure investments are not so often correlated with those more local policies. Interestingly, we observe that the introduction of fixed effects dampen the estimate of car commuting elasticity to road travel time. Indeed, introducing the dyadic fixed effects instead of the location-specific fixed effects reduced the estimate of travel time, in absolute value, from -2.75 to -0.83, suggesting that in contrast with public transport, road infrastructure target strategic axes. Location-specific fixed effects play only a little role in this case too.

### 2.4.2 Heterogeneity

Baseline results validate our theoretical hypothesis, with a high pseudo-R-square and a significant impact of the generalized cost of transport on commuting flows. These results confirm that reduction in travel time, allowed by infrastructure improvement, have caused additional commuting flows in Paris region, for both public transport and car. We now discuss heterogeneity of our estimates

[27] We control for observable characteristics when we drop the fixed effects: distance for the residence-workplace fixed effects, the number of residents for origin fixed effects and the number of jobs for destination fixed effects.

[28] Table 2.12 display the same results but conditioning on the dyads that can be estimated in the three-way fixed effects specification. It display similar estimates.

Table 2.6: Regressions - period

| | All commuters | | |
|---|---|---|---|
| | (1) | (2) | (3) |
| $\log$(Pub. transport dura.) | $-1.18^{***}$ <br> $0.105$ | $-0.92^{***}$ <br> $0.079$ | $-0.819^{**}$ <br> $0.327$ |
| $\log$(Road duration) | $-0.675^{***}$ <br> $0.072$ | $-0.722^{***}$ <br> $0.05$ | $-0.418^{***}$ <br> $0.115$ |
| $\log$(Potential road traffic) | $0.406^{***}$ <br> $0.017$ | $0.205^{***}$ <br> $0.012$ | $0.269^{***}$ <br> $0.028$ |
| Pseudo R2 | 0.9681 | 0.9559 | 0.9538 |
| No. obs. | 314,907 | 482,740 | 327,312 |
| No. fixed effects | | | |
| Orig. $\times$ Dest. | 105,306 | 160,924 | 163,662 |
| Orig. $\times$ Year | 3,856 | 3,861 | 2,573 |
| Dest. $\times$ Year | 3,845 | 3,850 | 2,568 |
| Years | 1968-1982 | 1982-1999 | 1999-2014 |

according to the sub-period of time used for estimation, the geometry of the dyad, if it is radial or peripheral, and finally with sub-groups of raw distance between origin and destination. To this end, we will gather flows for all modes and used the specification of Table 2.10 in appendix rather than the specification that distinguishes flows by mode for the sack of simplicity and to add all census years from 1968 to 2014.

The infrastructure programs underwent different phase, with major improvements occurring in the 1980s. Table 2.6 provides results of our estimates for different time sub-period: 1968 to 1982, 1982 to 1999, and 1999-2014. We observe small variations of the estimates over time, with a decrease of the effect, in absolute value, for the two modes over time, especially at the end of the period. It is important to note that restricting the sample to sub-period of time reduces the number of municipalities that contributes to the estimation. The location-specific fixed effects decrease from around 9,000 to less than 4,000. However, the number of dyads fixed effects is just slightly lower. Suggesting that the connections with remote municipalities contribute less in this estimation than in the baseline one. This could explain why the estimate of the incidence of time on flows is lower than with all years put together. Finally, note that the goodness of fit of the regression remains very high for all sub-period.

We discuss in Section 2.1 the making of the subruban metropolis in light of observed commuting flows. Our empirical strategy has confirmed the causality between the infrastructure programs and the relocation of commuters within the metropolis. Let now dig further the incidence of transportation infrastructure according to the direction of the flows. In particular, the modern metropolis followed a core-periphery structure, with most of the jobs in downtown area and residence in the surrounding suburbs. Along which direction went its extension into a suburban metropolis?

Table 2.7 display the estimation of the elasticity of commuting flows to the travel time for different directions: peripheral or radial in column (1) and distinguishing centrifugal from

**Table 2.7:** Regressions - flow geometry

| | All commuters | |
|---|---|---|
| | (1) | (2) |
| Peripherality $\times$ $\log$(Pub. transport dura.) | $-0.517^{**}$ <br> $0.229$ | $-0.507^{**}$ <br> $0.226$ |
| $\times$ $\log$(Road duration) | $-0.694^{***}$ <br> $0.125$ | $-0.682^{***}$ <br> $0.124$ |
| $\times$ $\log$(Potential road traffic) | $0.27^{***}$ <br> $0.016$ | $0.267^{***}$ <br> $0.016$ |
| Radiality $\times$ $\log$(Pub. transport dura.) | $-2.11^{***}$ <br> $0.104$ | |
| $\times$ $\log$(Road duration) | $-1.098^{***}$ <br> $0.059$ | |
| $\times$ $\log$(Potential road traffic) | $0.373^{***}$ <br> $0.012$ | |
| Centrifugal rad. $\times$ $\log$(Pub. transport dura.) | | $-2.716^{***}$ <br> $0.168$ |
| $\times$ $\log$(Road duration) | | $-1.434^{***}$ <br> $0.126$ |
| $\times$ $\log$(Potential road traffic) | | $0.518^{***}$ <br> $0.017$ |
| Centripetal rad. $\times$ $\log$(Pub. transport dura.) | | $-1.781^{***}$ <br> $0.116$ |
| $\times$ $\log$(Road duration) | | $-0.828^{***}$ <br> $0.07$ |
| $\times$ $\log$(Potential road traffic) | | $0.29^{***}$ <br> $0.012$ |
| Pseudo R2 | 0.9551 | 0.9552 |
| No. obs. | 1,571,140 | 1,571,140 |
| No. fixed effects | | |
| Orig. $\times$ Dest. | 224,796 | 224,796 |
| Orig. $\times$ Year | 9,003 | 9,003 |
| Dest. $\times$ Year | 8,982 | 8,982 |
| Years | 7 census years from 1968 to 2014 | |

centripetal flows in column (2). It shows that for public transport, the incidence of travel time is four time higher over the radial dimension, compare to the peripheral one. It is particularly true for commuting from the periphery to the center (*centrifugal*) than reverse commuting (*centripetal*). This pattern could results from rail line being mostly radial, therefore commuting over peripheral directions implies line exchanges that are known to be (negatively) valued more by commuters (see Small and Verhoef (2007)). This result indicates that rail infrastructures have contributed to preserve the core-periphery structure of the metropolis. However, it also sustain reverse commuting and probably flows that cross the downtown area to connect distant suburban municipality thanks to the tunnels build under the central districts.

The same hold for car commuting but with lower magnitude. The higher flexibility of individual car induced a lower incidence of travel time and less contrast between radial and peripheral commuting. It is in line with an increase of the share of car commuter with distance, while in parallel the share of public transport commuters decrease slightly with distance (see Figure 2.4 in appendix). In addition, it supports the idea that road improvements have contributed more to the interconnection of suburban locations than rail.

Finally, Table 2.13 in appendix provides results distinguishing Paris downtown area and its suburbs and by sub-groups of raw distance between residence and workplaces (as the crow flies). It shows that the incidence of road travel time on commuting flows is higher when the workplace locate in the suburbs. It reinforces previous conclusion on the role of road improvement in connecting remote areas to the rest of the metropolis thanks to the higher flexibility of individual car. It also suggests that previous results showing a substantial centrifugal orientation of the flows fostered by road improvement are actually suburbs-to-suburbs commuting. Similar results hold for public transport. We find a greater elasticity when the workplace is in the suburbs fostering both reverse commuting and suburbs-to-suburbs flows. These findings are in line with the suburbanization of jobs in the suburban metropolis, with higher rate than residential suburbanization.

Turning to the raw distance between residence and workplace, we observe a higher incidence of road travel time for both short and long distance, while the value is the lowest for distances that are between 10 and 20km. It suggests that for mid-range commuting distance the flexibility of car commuting is more privileged compare to travel time than for other shorter and longer commutes. In addition, potential road traffic is gradually increasing with distance, which is in line with an additive accumulation of congestion in the resulting effective commuting time. The incidence of travel time offered by public transport travel on flows is also gradually increasing with distance. It is in line with previous results that public transport are privileged over direct lines. When the connection become crooked, with line exchange and additional waiting time, the reliance on public transport decline sharply.

### 2.4.3 Robustness Checks

**Identification strategy** As explained above, public authorities might intend to treat some specific dyads. To measure the magnitude of this issue, we build on identification strategies proposed by Mayer and Trevien (2017) to select two subsamples of municipalities that were unintentionally treated. We then estimate our gravity model, selecting origin, destination or both municipalities among the subsamples. The first strategy (columns 1 to 3) compares municipalities located in-between Paris and subcenters (airport and newtowns) and other municipalities left out of the network, because of a less favorable location (see figure 2.6). The second strategy (columns 4 to 6) is based on the fact that the initial Regional express rail plan is quite different from the network that was actually built (see figure 2.8 and 2.7). This turnaround is mainly due to budget cut, which led to favor rebuilt lines instead of brand new ones. With this strategy, we compare municipalities that were connected to the RER but that should not have, to municipalities that was originally planned to be connected and did not. In all cases, we find stronger elasticity than the baseline result, suggesting that out model is not biased by this potential source of endogeneity.

In the second range of tests, we check that indirect routes, that require a change of train or subway, are also affected by a decrease in travel time, since the "intention to treat" seems less clear for such dyads (see table 2.14 in appendix, page 132). The elasticity is not significant when restricting to direct routes, probably because of the low number of dyads in that case (3%). We divide indirect routes in two categories, based on the ratio of the distance with public transport and the distance as the crows fly. When this ratio is above 1.5, we consider the public transport route as irrelevant, because it requires a sizeable detour. The elasticity associated with travel time reduction is much more important for routes that we identify as relevant than for irrelevant ones.

Our last robustness test is less conclusive (see table 2.15, page 132 in the appendix). When exclude municipalities directly connected to the RER network, to keep only those that are indirectly connected, the elasticity of rail travel time fades (column 1) or disappears (columns 2 and 3). This must be due to the fact that, municipalities that are remote from the RER network have not encountered a detectable improvement of their accessibility with public transport network.

**Travel time variation construction** We tested several specifications for the rail travel time variable, for which we restrict to municipalities near a rail station. Table 2.16 (in the appendix) shows that the variable which includes the whole duration (walking to the station, waiting for the train, time spent on the train, and possible transfer), between the time of leaving home and the time of arriving at work, has the best explanatory power. Table 2.17 (in the appendix page 133) indicates that a weighted duration of all stations in origin and destitation municipalities (column 4) proves to be less satisfactory than selecting the station which is closest to the town hall (columns 1 to 3).

## 2.5 Additional Results

Previous section discuss the main results of our empirical strategy: the causal relationship between the improvement of transportation infrastructures and additional commuting flows, which has contributed to the making of the suburban metropolis. The theoretical framework allows us to clearly show how our empirical strategy provide a convincing identification of the parameter of interests. In this section, we propose to leverage this theoretical framework to interpret the information contain in the regression fixed effects.

### 2.5.1 Interpretation of the Fixed Effects

Our empirical approach is based on the estimation of a structural gravity equation (2.2), which allows to impose that the estimation, described in equation (2.4), satisfies the general equilibrium conditions. In particular, the oriented location-specific fixed effects captures demographic evolution of residential location and workplaces, as well as general equilibrium effect of network improvements, such that we show:

$$\delta^R_{i,t} = \frac{R_{i,t}}{\Pi^{-\epsilon}_{i,t}} \quad \text{and} \quad \delta^L_{j,t} = \frac{L_{j,t}}{P^{-\epsilon}_{j,t}}$$

with $R_{i,t} = \sum_j l_{i,j,t}$ and $L_{j,t} = \sum_i l_{i,j,t}$ the number of residents and jobs, respectively, that capture the demographic dynamics of places corresponds to idiosyncratic shocks on residential locations and workplaces that could be correlated with the infrastructure improvement, including place-based policies like social housing or free-tax zone that were put in place in Paris region during our period of study. The general equilibrium incidence of the infrastructure network improvement are the multilateral resistance terms that capture the relative centrality of residential location and workplaces over year, respectively $\Pi^{-\epsilon}_{i,t} = \sum_{j,m} \left(\frac{a_j w_j}{\kappa^m_{k,j}}\right)^{-\epsilon}$ and $P^{-\epsilon}_{j,t} = \sum_{i,m} \left(\frac{u_i q_i^{-\beta}}{\kappa^m_{i,k}}\right)^{-\epsilon}$. Hence, in theory, we can provide more information on the incidence of the infrastructure improvement with the fixed effects than the causal bilateral effect on flows associated with a reduction of commuting time. We show in what follow how the PPML estimation ensure a perfect match between the estimated and theoretical parameters.

The estimation of structural gravity (using observed residential populations and jobs) is in fact equivalent to including fixed effects and imposing the sum of fitted commuting flows to equal observed spatial economic activity distribution for each source and each destination so that $\sum_i \hat{l}_{i,j} = R_i$ and $\sum_j \hat{l}_{i,j} = L_j$. It corresponds to the adding-up constraint, that does not hold in general, except if it is explicitly taking into account in the estimation. It implies (in general) to either redefine residential populations and jobs (as their fitted values), or impose fitted commuting flows to sum up to observed residential populations and jobs. The PPML estimator is an exception, where fitted residential population and jobs always equal observed values, as long as residential location and workplace fixed effects are included in the estimation (Fally 2015). When they are no missing observations, we obtain: $\sum_i \hat{l}_{i,j} = \sum_i l_{i,j} = R_i$ and $\sum_j \hat{l}_{i,j} = \sum_j l_{i,j} = L_j$.

It directly derives from the first-order conditions associated with the PPML approach ($\sum_i l_{i,j} - \hat{l}_{i,j} = 0$). This property is specific to PPML, which is the only PML estimator that systematically yields the adding up property.[29]

Imposing observed spatial economic distribution is systematically satisfied with PPML, according to the additive property. The two multilateral-resistance terms can be recovered from the parameters estimation and the observed residential and employed population in location $k$

$$\hat{\Pi}_k^{-\epsilon} = L_0 R_k \exp(-\hat{\delta}_k^R) \quad \text{and} \quad \hat{P}_k^{-\epsilon} = \frac{L_k}{L_0} \exp(-\hat{\delta}_k^L) \tag{2.5}$$

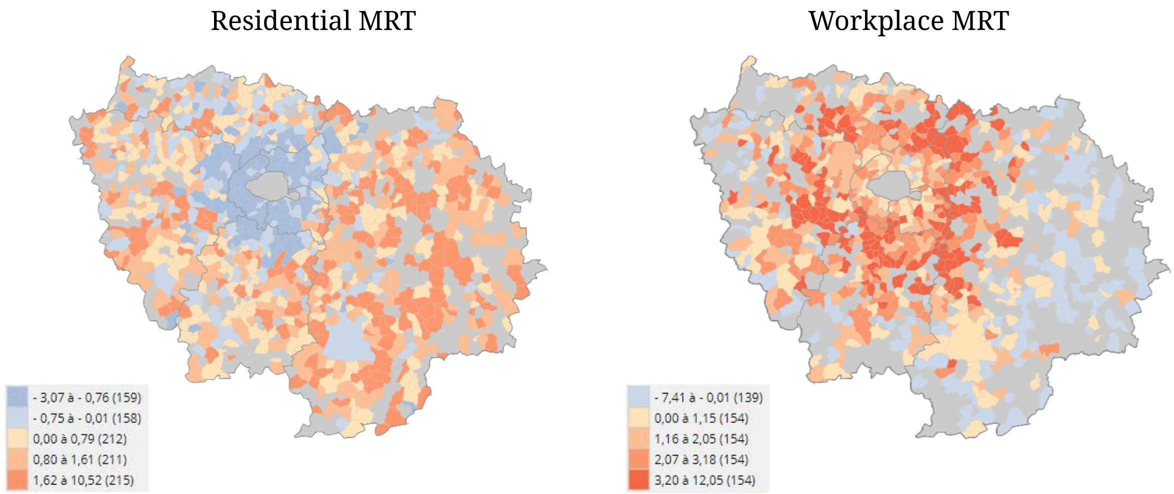
ensuring a perfect match between the structural gravity terms and the corresponding directional fixed effects. They are the unique solutions of the general equilibrium, up to a normalization term. It thus provide a theory consistent way to recover the multilateral resistance terms from the estimated fixed effects. It corresponds to the location-specific incidence of the improvements of the infrastructure network, since the attractiveness of a place depends on its relative accessibility in the metropolis.

In particular, they are theory-consistent aggregates of bilateral commuting costs[30] and they correspond to general equilibrium commuting costs indexes. Hence they capture the fact that a change in any bilateral commuting costs has (i) additional effect on commuting flows,(ii) it also affects all other locations within the citywith (iii) possible feedback effects on the original treated bilateral connection. In addition it highlights the fact that a reduction in the bilateral commuting costs between two locations needs to offset the overall improvement of commuting routes within the city to translate into an increase in commuting flows.

29 With OLS we obtain $\sum_i log(\hat{l}_{i,j}) = \sum_i log(l_{i,j})$, which do not imply equality between the sum *in level*. With non-linear least square approach (flows are in levels and we minimize the sum of squared errors) we obtain $\sum_i \hat{l}_{i,j} l_{i,j} = \sum_i l_{i,j}^2$. For Gamma-PML, the inclusion of fixed-effects implies that the ratio $\frac{l_{i,j}}{\hat{l}_{i,j}}$ averages to unity for each residential location and each workplace, which does not imply equality between the sum of flows at residential location or workplace.

30 It allows to collapse the $N \times N$ dimensions of feasible bilateral commutes within the city into $2 \times N$ dimensions of location-specific commuting cost indexes.

**Figure 2.2:** Multilateral Resistance Terms Growth Rate (1968-2014)




### 2.5.2 Multilateral Resistance Terms

The distributions of the multilateral resistance terms, estimated using the procedure described above, depart strongly from normality. Indeed, as expected, the distribution is fat-tailed and right-skewed, with few locations that display a very high level of centrality and most of the rest with comparable level of low accessibility. Since we are interested in the incidence of infrastructure improvement in reshuffling opportunities within the region, we will focus on the relative growth rate of the multilateral resistance terms between the beginning and the end of our period of study.

Figure 2.5 in appendix display the distribution of the growth rate of the multilateral resistance terms. The density of the two distribution has a shape that is close to normality, implying that there is a majority of locations that saw a variation in their accessibility to opportunities (either residents or jobs) close to the mean, with few outliers. We map the value of this growth rate of the multilateral resistance terms in Figure 2.2.

The maps highlight interesting patterns of the reshuffling of opportunities in the suburban metropolis. We observe first that residential locations that benefit the most the improvement of the regional transportation system are remote places, in the outer suburbs. Conversely, central residents are adversely affected by the improvement of the transportation network. Indeed remote locations are less remote thanks to the reduction in average commuting time within the transportation network. Especially long distance travels are the one that observe a larger reduction in associated travel time thanks to the infrastructure improvement with long RER lines and highway. Central location face a reduction in their relative access to jobs that could either corresponds to the suburbanization of jobs and/or the increasing competition associated with additional labor supply with the integration of remote local labor market to the metropolitan labor market through commuting.

Turning to evolution of the multilateral resistance terms for workplaces, we observe a peak of the increase in accessibility in a ring 20km away from the metropolitan core. It corresponds to transportation nodes (RER interconnections and highway branches) that are in the suburbs and magnified the labor supply incidence of commuting costs (e.g. Cergy, Massy, Roissy). We observe that the few negative values are counted at the regional fringe. It corresponds to place of work that are too far from the infrastructure improvement to benefit from them and access to a sufficiently supply of resident workers.

The study of the evolution of commuter market access shed light on the suburbanization process. Conversely to seminal results from the monocentric approach of the city (Alonso-Muth-Mills) that assumes the existent of a *unique* central business district, where all jobs concentrate, we show that actually jobs have suburbanized thanks to a larger improvement of firm access to workers close to peripheral transportation nodes (multi-line rail stations, highways interconnections) compare to the rest of the metropolis. The reduction in overall travel time within the region (magnified by the building of rail tunnels under Paris and ring-roads) increases the centrality of infrastructure nodes to attract more workers through commuting that enlarges their local supply of workers. On the other hand, the suburbanization of residents almost follow a monocentric pattern, with remote residential location benefiting most from infrastructure improvement (and reduction in travel time). However, those centripetal commuting flows reach less the downtown areas and connect the low-density with the urban suburbs. Remote locations -at the end of line- have seen a substantial increase in their relative access to attractive jobs through commuting, while central ones have lost from their comparative advantage to access close jobs and lower labor supply competition.

## 2.6 Conclusion

Suburban metropolises have emerged all around the world in the second half of the twentieth century, mostly fueled by the democratization of car commuting. Transportation infrastructure, like highways, have accompanied this development. Interestingly, in few metropolises like Paris, urban planner also developed major public transport infrastructure, Mass-Transit, that also led to urban sprawl. We have shown that this framework implied a relative complementary between rail and road, where the former maintained substantial commuting flows between the downtown area and the suburbs along radial lines, while the later supported the rise of suburb-to-suburb commuting.

It suggests that reducing the use of cars for commuting would impact more suburban residents and jobs than the ones located in central areas. In addition, existing public transport infrastructures lack of flexibility to represent a substitute for all commuters. An opportunity to address these issues could reside in favoring multi-modal commuting, that have been underused -and studied- so far.

# Bibliography

**Ahlfeldt, Gabriel M, Stephen J Redding, Daniel M Sturm, and Nikolaus Wolf**, "The economics of density: Evidence from the Berlin Wall," *Econometrica*, 2015, *83* (6), 2127–2189. ↑96, ↑113

**Allen, Treb and Costas Arkolakis**, "The welfare effects of transportation infrastructure improvements," *Working paper*, 2021. ↑105

**Alonso, William**, *Location and Land Use: Toward a General Theory of Land Rent* 1964. ↑121

**Anderson, James E**, "The gravity model," *Annual Review of Economics*, 2011, *3* (1), 133–160. ↑108

_ **and Eric van Wincoop**, "Gravity with gravitas: A solution to the border puzzle," *American Economic Review*, 2003, *93* (1), 170–192. ↑108

**Anderson, Michael L**, "Subways, strikes, and slowdowns: The impacts of public transit on traffic congestion," *American Economic Review*, 2014, *104* (9), 2763–96. ↑96

**Arvis, Jean-Francois and Ben Shepherd**, "The Poisson quasi-maximum likelihood estimator: a solution to the 'adding up'problem in gravity models," *Applied Economics Letters*, 2013, *20* (6), 515–519. ↑96

**Banerjee, Abhijit, Esther Duflo, and Nancy Qian**, "On the road: Access to transportation infrastructure and economic growth in China," *Journal of Development Economics*, 2020, *145*, 102442. ↑112

**Baum-Snow, Nathaniel**, "Did highways cause suburbanization?," *The Quarterly Journal of Economics*, 2007, *122* (2), 775–805. ↑96, ↑102

**Donaldson, Dave**, "Railroads of the Raj: Estimating the impact of transportation infrastructure," *American Economic Review*, 2018, *108* (4-5), 899–934. ↑112

**Duranton, Gilles and Matthew A Turner**, "Urban growth and transportation," *Review of Economic Studies*, 2012, *79* (4), 1407–1440. ↑96

**Eaton, Jonathan and Samuel Kortum**, "Technology, geography, and trade," *Econometrica*, 2002, *70* (5), 1741–1779. ↑107

**Fally, Thibault**, "Structural gravity and fixed effects," *Journal of International Economics*, 2015, *97* (1), 76–85. ↑96, ↑119

**Fernández-Val, Iván and Martin Weidner**, "Fixed effects estimation of large-T panel data models," *Annual Review of Economics*, 2018, *10*, 109–138. ↑111

**François, Dominique**, "Se rendre au travail : distances et temps de transport s'allongent," in "La mobilité des Français. Panorama issu de l'enquête nationale transports et déplacements 2008," Vol. 4 2010, pp. 83–98. ↑102

**Garcia-López, Miquel-Àngel, Camille Hémet, and Elisabet Viladecans-Marsal**, "Next train to the polycentric city: The effect of railroads on subcenter formation," *Regional Science and Urban Economics*, 2017, *67*, 50–63. ↑96

**Giroud, Xavier**, "Proximity and investment: Evidence from plant-level data," *The Quarterly Journal of Economics*, 2013, *128* (2), 861–915. ↑112

**Gourieroux, Christian, Alain Monfort, and Alain Trognon**, "Pseudo maximum likelihood methods: Applications to Poisson models," *Econometrica*, 1984, pp. 701–720. ↑109

**Guimaraes, Paulo, Octávio Figueirdo, and Douglas Woodward**, "A tractable approach to the firm location decision problem," *Review of Economics and Statistics*, 2003, *85* (1), 201–204. ↑109

**Head, Keith and Thierry Mayer**, "Gravity equations: Workhorse, toolkit, and cookbook," in "Handbook of international economics," Vol. 4, Elsevier, 2014, pp. 131–195. ↑108

**Heblich, Stephan, Stephen J Redding, and Daniel M Sturm**, "The making of the modern metropolis: evidence from London," *The Quarterly Journal of Economics*, 2020, *135* (4), 2059–2133. ↑94, ↑96, ↑113

**Lancaster, Tony**, "Orthogonal parameters and panel data," *The Review of Economic Studies*, 2002, *69* (3), 647–666. ↑111

**Mayer, Thierry and Corentin Trevien**, "The impact of urban public transportation evidence from the Paris region," *Journal of Urban Economics*, 2017, *102*, 1–21. ↑96

**McFadden, Daniel**, "The measurement of urban travel demand," *Journal of Public Economics*, 1974, *3* (4), 303–328. ↑109

_ , "Modelling the choice of residential location," *Spatial Interaction Theory and Planning Models*, 1978. ↑96, ↑109

**Mills, Edwin S**, "An aggregative model of resource allocation in a metropolitan area," *The American Economic Review*, 1967, *57* (2), 197–210. ↑121

**Muth, Richard F**, "Cities and Housing: The Spatial Pattern of Urban Residential Land Use," *Third Series: Studies in Business and Society*, 1969. ↑121

**Neyman, Jerzy and Elizabeth L Scott**, "Consistent estimates based on partially consistent observations," *Econometrica: Journal of the Econometric Society*, 1948, pp. 1–32. ↑111

**Santos-Silva, João and Silvana Tenreyro**, "The log of gravity," *The Review of Economics and Statistics*, 2006, *88* (4), 641–658. ↑109, ↑110, ↑111

**Severen, Christopher**, "Commuting, labor, and housing market effects of mass transportation: Welfare and identification," *FRB of Philadelphia, Working Paper*, 2020. ↑96

**Small, Kenneth and Erik Verhoef**, *The Economics of Urban Transportation* 2007. ↑116

**SNCF, Réseau, Direction Générale Stratégie, Programmation et Maîtrise d'Ouvrage, and Direction de la Stratégie du Réseau**, "Schéma Directeur: Etoile ferroviaires et Services express métropolitains," Technical Report, Ministère de la Transition écologique 2020. ↑94

**Tsivanidis, Nick**, "The aggregate and distributional effects of urban transit infrastructure: Evidence from bogotá's transmilenio," *Working paper*, 2020. ↑96

**Weidner, Martin and Thomas Zylkin**, “Bias and consistency in three-way gravity models,” *Journal of International Economics*, 2021, *132.* ↑111, ↑112

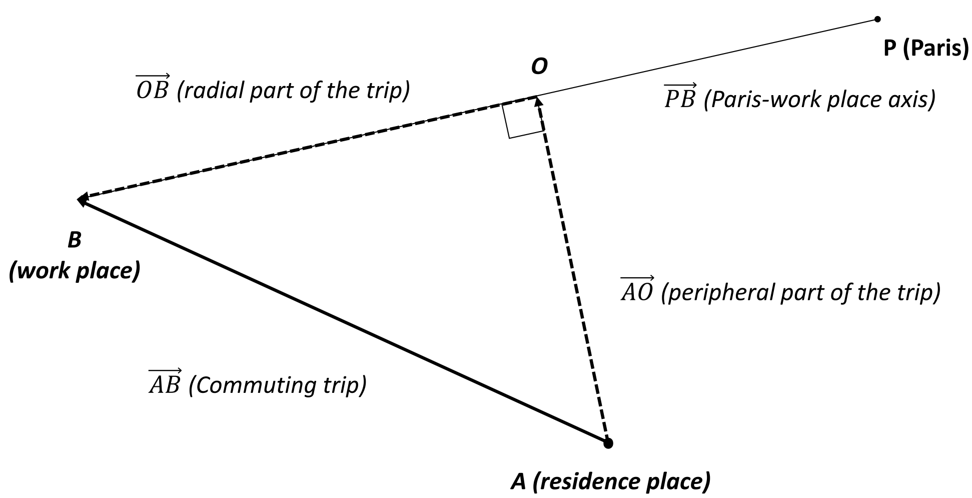
**Wooldridge, Jeffrey M**, “Some alternatives to the Box-Cox regression model,” *International Economic Review*, 1992, pp. 935–955. ↑110

## 2.A Orientation of commuting flows and supplementary statistics

Radial and peripheral indexes aims at measuring whether commuting flows between municipalities A and B are core-periphery oriented or not (see figure 2.3). There are both based on the vector $\vec{AB}$ projection on vector $\vec{PB}$ where $P$ designates the city of Paris (*i.e.* the Paris-workplace axis). The radial part of the trip from A to B is the orthogonal projection $\vec{OB}$ while the peripheral part of the trip from A to B is the orthogonal rejection $\vec{AO}$. Both index are normalized with the straight line distance between A and B, $\|AB\|$. Finally, the radiality index $iR$ is equal to $\frac{\|OB\|}{\|AB\|}$ and the peripherality index $iP$ is equal to $\frac{\|AO\|}{\|AB\|}$.

In order to obtain the same index the trip from A to B and the reverse trip from B to A, both vector$\vec{AB}$ projections on $\vec{AP}$ and $\vec{BP}$ are considered. We select the one that minimizes $\|AO\|+\|BO\|$. Last point, the radial index can be decomposed in two centripetallity and centrifugallity indexes, depending on whether the worker moves closer or further away from Paris, on the way from home to work.

**Figure 2.3:** Radiality and peripherality indexes



### 2.A.1 Supplementary descriptive statistics

**Table 2.8:** Speed by mode

| Origin/dest. | Orientation | Speed (km/h) | | | | Weighted speed (km/h) | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | Rail | | Road | | Rail | | Road | |
| | | 1968 | 2014 | 1968 | 2014 | 1968 | 2014 | 1968 | 2014 |
| Paris ↔ Paris | Any | 6.3 | 6.4 | 34.2 | 34.2 | 6.5 | 6.8 | 34.9 | 35.4 |
| Paris → Suburbs | Centrifugal | 21.5 | 22 | 52.4 | 62.2 | 9.4 | 10.9 | 39.5 | 48.9 |
| Suburbs → Paris | Centripetal | 21.4 | 21.8 | 52.4 | 62.1 | 11.8 | 14.6 | 42.2 | 51.2 |
| Suburbs ↔ Suburbs | Centrifugal (high) | 21.5 | 22.2 | 52.9 | 65.9 | 10.1 | 12.6 | 40.2 | 51.1 |
| | Centrifugal (low) | 17 | 17.4 | 49.1 | 58.8 | 8.1 | 9.7 | 39.1 | 48 |
| | Centripetal (high) | 21.4 | 22.1 | 53.1 | 65.9 | 10 | 15.4 | 43 | 53.4 |
| | Centripetal (low) | 16.9 | 17.3 | 49.2 | 58.8 | 8 | 10.5 | 40 | 48.3 |
| | Peripheral | 11.1 | 11.3 | 45.9 | 48.8 | 6 | 6.9 | 36.6 | 42 |

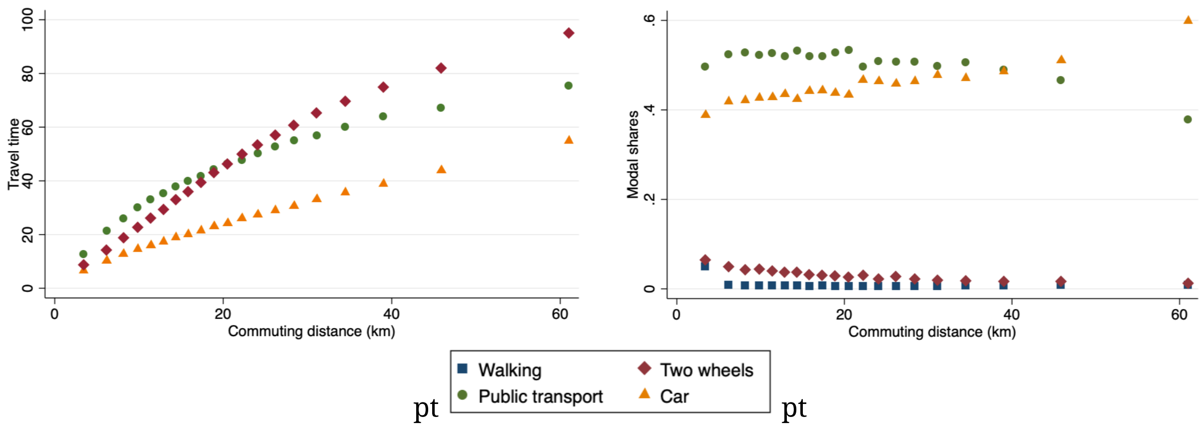
*Note: speed are calculated with straight line distance.*

**Table 2.9:** Distance by mode

| Origin/dest. | Orientation | Unweighted distance (km) | | | | | Weighted distance (km) | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | Straight line | Rail | | Road | | S. line | | Rail | | Road | |
| | | | 1968 | 2014 | 1968 | 2014 | 1968 | 2014 | 1968 | 2014 | 1968 | 2014 |
| Paris ↔ Paris | Any | 3.5 | 5.1 | 5 | 4.5 | 4.5 | 1.6 | 1.5 | 4.7 | 2.6 | 3.7 | 1.8 |
| Paris → Suburbs | Centrifugal | 41.5 | 56.6 | 56.5 | 51.3 | 52.5 | 8.7 | 10.5 | 11.6 | 14.8 | 10.4 | 16.1 |
| Suburbs → Paris | Centripetal | 41.5 | 56.6 | 56.6 | 51.2 | 52.5 | 12.3 | 15.5 | 16.3 | 21.6 | 14.9 | 20.6 |
| Suburbs ↔ Suburbs | Centrifugal (high) | 65 | 92.9 | 92.8 | 81.8 | 82.9 | 10.4 | 13.8 | 13.3 | 21.3 | 9.8 | 17.8 |
| | Centrifugal (low) | 52.4 | 94.4 | 94.9 | 67.7 | 71.8 | 8 | 11.4 | 11.8 | 18 | 9.5 | 16.4 |
| | Centripetal (high) | 65 | 93.2 | 93.1 | 81.3 | 83.1 | 10.8 | 18.2 | 13.9 | 29.1 | 12 | 22.7 |
| | Centripetal (low) | 52.4 | 94.6 | 95.1 | 67.7 | 71.9 | 8.1 | 12.5 | 12 | 19.9 | 10.6 | 17.9 |
| | Peripheral | 32.9 | 82.2 | 82.5 | 40.9 | 44.2 | 5 | 7.6 | 8.7 | 12.4 | 6.9 | 11.4 |

## 2.A.2 Modal shares

**Figure 2.4:** Mode shares and average travel time binned by row distance




**First graph** (*left hand side*) plots average travel time for commuting within Paris region, with respect to the great circle distance (not the infrastructure network specific distance) associated to the commute.
*Notes:* Travel time by transportation modes (without congestion), binned by row distances (20 bins).

**Second graph** (*right hand side*) plots the modal shares by distance bins. Transportation mode choice are self-reported (yet census is mandatory and every workers obliged to provide its personal information with the accuracy). Two wheels corresponds to any two wheels vehicles like bike or motorbike.
*Notes:* Modal shares by transportation modes, binned by row distances (20 bins).
*Source:* Census (year 2013), INSEE. Travel time and itinerary, API-IGN (2020). Own calculation of public transport network, from RATP, SNCF and STIF/IdF Mobilité open-data (2013).

## 2.B Supplementary estimation results

### 2.B.1 alternative specifications

**Table 2.10:** Baseline results, without mode distinction

| | (1) | (2) | (3) | (4) | (5) | (6) |
|---|---|---|---|---|---|---|
| log(Pub. transport dura.) | $-2.825^{***}$<br>0.099 | | $-2.158^{***}$<br>0.085 | $-2.076^{***}$<br>0.079 | $-2.042^{***}$<br>0.079 | $-2.026^{***}$<br>0.078 |
| log(Road duration) | | $-2.433^{***}$<br>0.063 | $-2.166^{***}$<br>0.057 | $-1.189^{***}$<br>0.045 | $-2.635^{***}$<br>0.092 | $-2.637^{***}$<br>0.092 |
| log(Potential road traffic) | | | | $0.427^{***}$<br>0.011 | $-0.018$<br>0.034 | $-0.018$<br>0.034 |
| log(Road travel time) $\times$ log(Potential road traffic) | | | | | $0.17^{***}$<br>0.01 | $0.17^{***}$<br>0.01 |
| log(Bus travel time) | | | | | | $-0.014^{*}$<br>0.0083 |
| log(Walking time) | | | | | | 0.008<br>0.013 |
| Pseudo R2 | 0.9536 | 0.954 | 0.9543 | 0.9551 | 0.9552 | 0.9552 |
| No. obs. | 1,571,140 | 1,571,140 | 1,571,140 | 1,571,140 | 1,571,140 | 1,571,140 |
| No. fixed effects | | | | | | |
| Orig. $\times$ Dest. | 224,796 | 224,796 | 224,796 | 224,796 | 224,796 | 224,796 |
| Orig. $\times$ Year | 9,003 | 9,003 | 9,003 | 9,003 | 9,003 | 9,003 |
| Dest. $\times$ Year | 8,982 | 8,982 | 8,982 | 8,982 | 8,982 | 8,982 |
| Years | 7 census years from 1968 to 2014 | | | | | |

Notes: The table provide the estimation results of equation (2.4) but without distinguished commuting flows by mode. The columns corresponds to different specification of the cost of transport.

**Table 2.11:** Regressions – fixed effects

| | (1) | (2) | (3) | (4) | (5) | (6) | (7) |
|---|---|---|---|---|---|---|---|
| log(Pub. transport dura.) | $-0.471^{***}$ | $-0.404^{***}$ | $-0.309^{***}$ | $-0.53^{***}$ | $-0.412^{***}$ | $-1.553^{***}$ | $-2.076^{***}$ |
| | 0.142 | 0.13 | 0.024 | 0.04 | 0.029 | 0.181 | 0.079 |
| log(Road duration) | $-2.262^{***}$ | $-2.741^{***}$ | $-2.699^{***}$ | $-2.27^{***}$ | $-2.752^{***}$ | $-0.827^{***}$ | $-1.189^{***}$ |
| | 0.081 | 0.077 | 0.02 | 0.035 | 0.021 | 0.122 | 0.045 |
| log(Potential road traffic) | $-0.076$ | $-0.068^{*}$ | $-0.037^{***}$ | $-0.013$ | $-0.066^{***}$ | $0.845^{***}$ | $0.427^{***}$ |
| | 0.048 | 0.035 | 0.0088 | 0.0081 | 0.0096 | 0.05 | 0.011 |
| log(Distance) | $0.471^{***}$ | $0.563^{***}$ | $0.482^{***}$ | $0.44^{***}$ | $0.567^{***}$ | | |
| | 0.058 | 0.051 | 0.016 | 0.019 | 0.018 | | |
| log(Population) | $0.477^{***}$ | $0.671^{***}$ | | $0.506^{***}$ | | | |
| | 0.015 | 0.031 | | 0.0053 | | | |
| log(Employment) | $0.756^{***}$ | $0.84^{***}$ | $0.862^{***}$ | | | | |
| | 0.025 | 0.037 | 0.0049 | | | | |
| Pseudo R2 | 0.9166 | 0.939 | 0.9353 | 0.9226 | 0.94 | 0.9468 | 0.9551 |
| No. obs. | 11,594,583 | 11,594,583 | 11,586,861 | 11,559,834 | 11,552,144 | 1,573,572 | 1,571,140 |
| No. fixed effects | | | | | | | |
| Year | 7 | 7 | | | | 7 | |
| Origin | | 1,287 | | | | | |
| Destination | | 1,287 | | | | | |
| Orig. × Year | | | 9,003 | | 9,003 | | 9,003 |
| Dest. × Year | | | | 8,982 | 8,982 | | 8,982 |
| Orig. × Dest. | | | | | | 224,796 | 224,796 |
| Years | 7 census years from 1968 to 2014 | | | | | | |

Notes: The table provide the estimation results of equation (2.4) introducing the fixed-effects successively.

**Table 2.12:** Regressions – fixed effects (restricted sample)

| | (1) | (2) | (3) | (4) | (5) | (6) | (7) |
|---|---|---|---|---|---|---|---|
| log(Pub. transport dura.) | $-0.414^{***}$ | $-0.33^{**}$ | $-0.274^{***}$ | $-0.402^{***}$ | $-0.341^{***}$ | $-1.557^{***}$ | $-2.076^{***}$ |
| | 0.143 | 0.137 | 0.022 | 0.036 | 0.027 | 0.187 | 0.079 |
| log(Road duration) | $-2.009^{***}$ | $-2.458^{***}$ | $-2.451^{***}$ | $-1.991^{***}$ | $-2.47^{***}$ | $-0.822^{***}$ | $-1.189^{***}$ |
| | 0.068 | 0.07 | 0.018 | 0.032 | 0.018 | 0.119 | 0.045 |
| log(Potential road traffic) | $-0.057$ | $-0.049^{*}$ | $-0.017^{**}$ | $-0.0035$ | $-0.048^{***}$ | $0.839^{***}$ | $0.427^{***}$ |
| | 0.044 | 0.028 | 0.0079 | 0.0071 | 0.0088 | 0.05 | 0.011 |
| log(Distance) | $0.36^{***}$ | $0.439^{***}$ | $0.382^{***}$ | $0.303^{***}$ | $0.446^{***}$ | | |
| | 0.056 | 0.053 | 0.014 | 0.016 | 0.016 | | |
| log(Population) | $0.449^{***}$ | $0.695^{***}$ | | $0.5^{***}$ | | | |
| | 0.01 | 0.028 | | 0.0059 | | | |
| log(Employment) | $0.703^{***}$ | $0.836^{***}$ | $0.803^{***}$ | | | | |
| | 0.019 | 0.036 | 0.0044 | | | | |
| Pseudo R2 | 0.8819 | 0.9144 | 0.908 | 0.8931 | 0.9163 | 0.947 | 0.9551 |
| No. obs. | 1,571,140 | 1,571,140 | 1,571,140 | 1,571,140 | 1,571,140 | 1,571,140 | 1,571,140 |
| No. fixed effects | | | | | | | |
| Year | 7 | 7 | | | | 7 | |
| Origin | | 1,287 | | | | | |
| Destination | | 1,287 | | | | | |
| Orig. × Year | | | 9,003 | | 9,003 | | 9,003 |
| Dest. × Year | | | | 8,982 | 8,982 | | 8,982 |
| Orig. × Dest. | | | | | | 224,796 | 224,796 |
| Years | 7 census years from 1968 to 2014 | | | | | | |

Notes: The table provide the estimation results of equation (2.4) introducing the fixed-effects successively, conditioning on dyades that can be estimated in the more constraint model (Column 7).

**Table 2.13:** Regressions - by distance

| All commuters | (1) | (2) |
|---|---|---|
| Paris↔Paris × log(Road duration) | $3.199^{***}$ <br> $0.795$ | |
| × log(Pub. transport dura.) | $-0.97^{***}$ <br> $0.323$ | |
| × log(Potential road traffic) | $0.906^{***}$ <br> $0.072$ | |
| Paris→Sub. × log(Road duration) | $-1.38^{***}$ <br> $0.188$ | |
| × log(Pub. transport dura.) | $-2.334^{***}$ <br> $0.222$ | |
| × log(Potential road traffic) | $0.612^{***}$ <br> $0.036$ | |
| Suburbs→Paris × log(Road duration) | $-0.626^{***}$ <br> $0.095$ | |
| × log(Pub. transport dura.) | $-1.216^{***}$ <br> $0.106$ | |
| × log(Potential road traffic) | $0.159^{***}$ <br> $0.022$ | |
| Suburbs↔Sub. × log(Road duration) | $-1.411^{***}$ <br> $0.051$ | |
| × log(Pub. transport dura.) | $-2.514^{***}$ <br> $0.102$ | |
| × log(Potential road traffic) | $0.436^{***}$ <br> $0.011$ | |
| <10km × log(Road duration) | | $-1.02^{***}$ <br> $0.076$ |
| × log(Pub. transport dura.) | | $-0.944^{***}$ <br> $0.091$ |
| × log(Potential road traffic) | | $-0.072^{***}$ <br> $0.01$ |
| 10-20km × log(Road duration) | | $-0.737^{***}$ <br> $0.044$ |
| × log(Pub. transport dura.) | | $-1.742^{***}$ <br> $0.067$ |
| × log(Potential road traffic) | | $0.193^{***}$ <br> $0.0085$ |
| 20-30km × log(Road duration) | | $-0.957^{***}$ <br> $0.056$ |
| × log(Pub. transport dura.) | | $-1.794^{***}$ <br> $0.09$ |
| × log(Potential road traffic) | | $0.272^{***}$ <br> $0.01$ |
| >30km × log(Road duration) | | $-1.447^{***}$ <br> $0.056$ |
| × log(Pub. transport dura.) | | $-2.862^{***}$ <br> $0.143$ |
| × log(Potential road traffic) | | $0.25^{***}$ <br> $0.011$ |
| Pseudo R2 | 0.9553 | 0.927 |
| No. obs. | 1,571,140 | 1,555,520 |
| No. fixed effects | | |
| Orig. × Dest. | 224,796 | 223,509 |
| Orig. × Year | 9,003 | 8,981 |
| Dest. × Year | 8,982 | 8,537 |
| Years | 7 census years from 1968 to 2014 | |

## 2.B.2 Robustness regressions

**Table 2.14:** Identification strategy – indirect routes

| | All commuters | | | |
|---|---|---|---|---|
| | (1) | (2) | (3) | (4) |
| log(Rail travel time) | $-0.306$ <br> 0.312 | $-0.23^{*}$ <br> 0.123 | $-0.69^{***}$ <br> 0.144 | 0.234 <br> 0.157 |
| log(Road travel time) | $-0.858^{***}$ <br> 0.144 | $-0.642^{***}$ <br> 0.058 | $-0.564^{***}$ <br> 0.068 | $-0.65^{***}$ <br> 0.078 |
| log(Rail travel time)× Dist. to stations | $-0.00025$ <br> 0.00033 | 0.00015 <br> 0.00012 | 0.0000026 <br> 0.00015 | $-0.000092$ <br> 0.00016 |
| No. obs. | 15,458 | 137,231 | 81,600 | 55,631 |
| No. f.e. | | | | |
| Orig. × Dest. | 2,658 | 27,830 | 17,148 | 12,311 |
| Orig. × Year | 2,122 | 2,154 | 2,146 | 2,077 |
| Dest. × Year | 2,008 | 1,943 | 1,766 | 1,879 |
| Routes | Direct | Indirect | Indirect & relevant | Indirect & irrelevant |
| Years | 7 census years from 1968 to 2014 | | | |
| Estimation | PPML | PPML | PPML | PPML |

**Table 2.15:** Identification strategy – RER stations only

| | All commuters | | |
|---|---|---|---|
| | (1) | (2) | (3) |
| log(Rail travel time) | $-0.199^{**}$ <br> 0.087 | 0.097 <br> 0.094 | 0.097 <br> 0.126 |
| log(Road travel time) | $-0.545^{***}$ <br> 0.047 | $-0.568^{***}$ <br> 0.047 | $-0.435^{***}$ <br> 0.067 |
| log(Rail travel time)× Dist. to stations | 0.000022 <br> 0.000026 | $-0.0000038$ <br> 0.000033 | $-0.000027$ <br> 0.000036 |
| No. obs. | 272,737 | 216,054 | 125,872 |
| No. f.e. | | | |
| Orig. × Dest. | 75,721 | 60,731 | 38,212 |
| Orig. × Year | 7,138 | 8,230 | 6,820 |
| Dest. × Year | 6,206 | 5,054 | 4,828 |
| Subsample | RER municipalities only | | |
| Origin | ✓ | | ✓ |
| Destination | | ✓ | ✓ |
| Years | 7 census years from 1968 to 2014 | | |
| Estimation | PPML | PPML | PPML |

**Table 2.16:** Alternative treatment variable – rail travel time

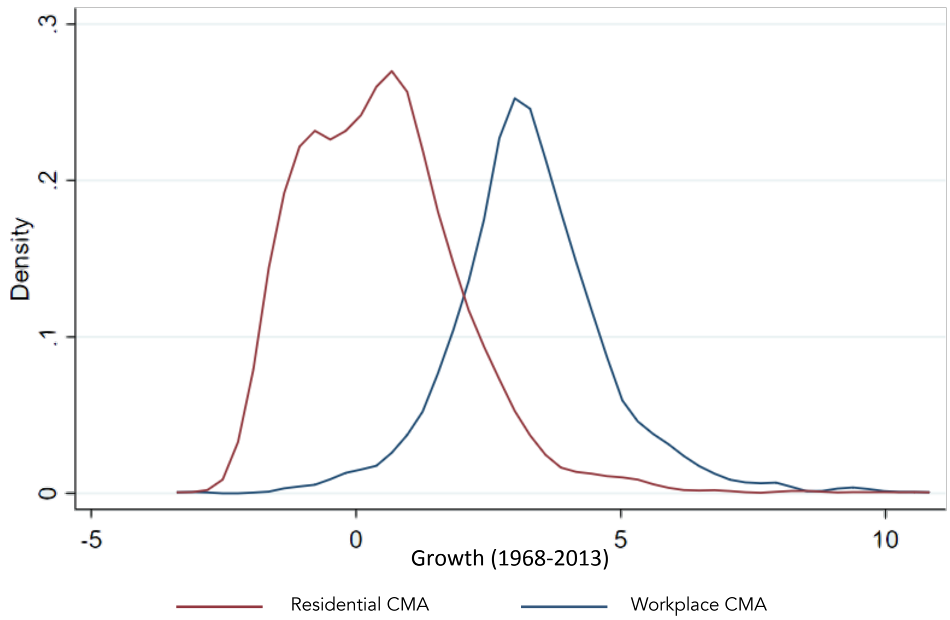
| | Commuting from and to municipalities near a station | | | |
|---|---|---|---|---|
| | (1) | (2) | (3) | (4) |
| log(Road travel time) | $-0.809^{***}$ | $-0.785^{***}$ | $-0.776^{***}$ | $-0.771^{***}$ |
| | 0.052 | 0.053 | 0.053 | 0.053 |
| log(Travel time on the train) | 0.019 | | | |
| | 0.031 | | | |
| log(Transfer + On the train) | | $-0.203^{***}$ | | |
| | | 0.043 | | |
| log(Waiting + Transfer + On the train) | | | $-0.284^{***}$ | |
| | | | 0.058 | |
| log(Walking to station + Waiting + Transfer + On the train) | | | | $-0.491^{***}$ |
| | | | | 0.084 |
| No. obs. | 152,689 | 152,703 | 152,703 | 152,703 |
| No. f.e. | | | | |
| Orig. × Dest. | 30,488 | 30,490 | 30,490 | 30,490 |
| Orig. × Year | 2,170 | 2,170 | 2,170 | 2,170 |
| Dest. × Year | 2,094 | 2,094 | 2,094 | 2,094 |
| Years | 7 census years from 1968 to 2014 | | | |
| Estimation | PPML | PPML | PPML | PPML |

**Table 2.17:** Alternative treatment variable – from stations to municipality

| | Commuting from and to municipalities near a station | | | |
|---|---|---|---|---|
| | (1) | (2) | (3) | (4) |
| log(Rail travel time) | $-0.129$ | $-0.284^{***}$ | $-0.259^{***}$ | $-0.048^{**}$ |
| | 0.087 | 0.058 | 0.043 | 0.021 |
| log(Road travel time) | $-0.625^{***}$ | $-0.776^{***}$ | $-0.767^{***}$ | $-0.812^{***}$ |
| | 0.087 | 0.053 | 0.043 | 0.039 |
| No. obs. | 51,851 | 152,703 | 278,823 | 360,578 |
| No. f.e. | | | | |
| Orig. × Dest. | 9,472 | 30,490 | 59,246 | 81,632 |
| Orig. × Year | 1,063 | 2,170 | 3,376 | 3,991 |
| Dest. × Year | 1,014 | 2,094 | 3,202 | 3,793 |
| Subsample | Max. dist. b/w station and city-hall | | | All station |
| | 500m | 1km | 2km | weighted |
| Years | 7 census years from 1968 to 2014 | | | |
| Estimation | PPML | PPML | PPML | PPML |

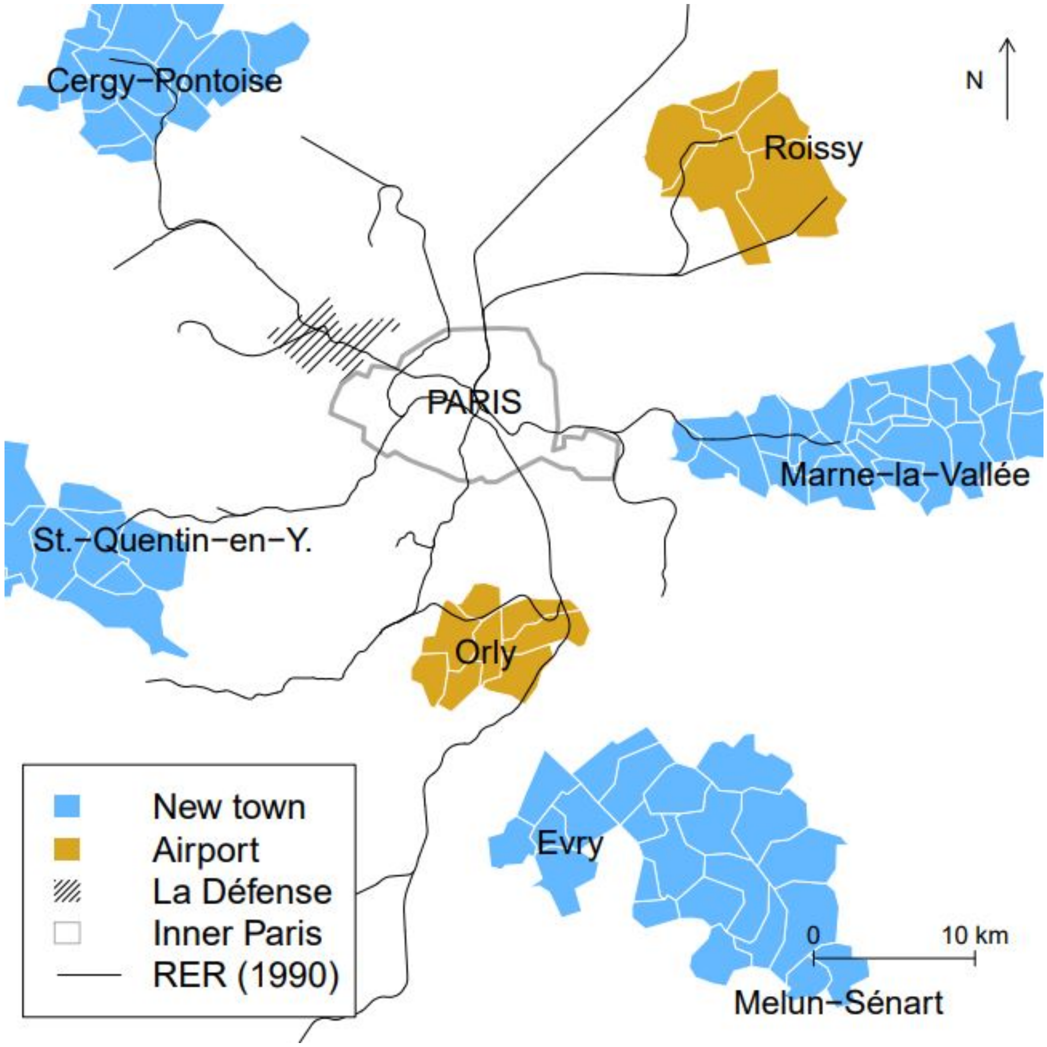
### 2.B.3 Distribution of the Fixed Effects

**Figure 2.5:** Multilateral Resistance Terms: *Growth rate (1968-2013)*



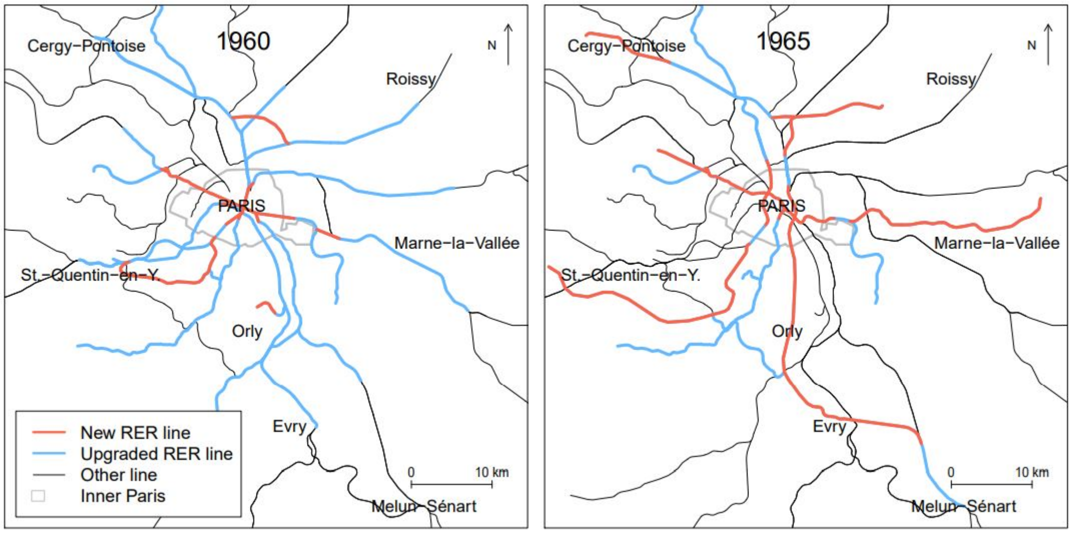
## 2.C Map of the different urban plan

**Figure 2.6:** The RER network and New Towns



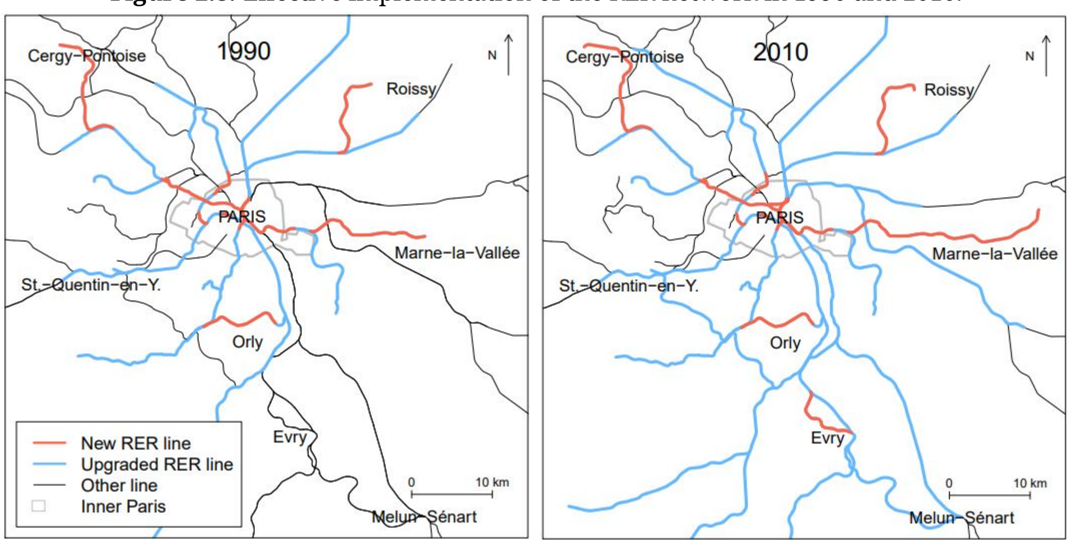

New Towns are areas that were chosen in the 1965 urban plan. The Subway Area highlights all municipalities that have at least one Parisian metro station. The borders inside those areas show the municipalities. The borders again represent municipalities, where in the case of Paris all 20 arrondissements are treated as separate municipalities.
*Source:* IAU - Ile-de-France; Mayer and Trevien (2017)

**Figure 2.7:** Urban Planning: RER projects in 1960 (PADOG) and 1965 (SDAURP).




PADOG: *Plan d'Aménagement et D'Organisation Générale de la Région Parisienne*.
SDAURP: *Schéma directeur d'aménagement et d'urbanisme de la Région Parisienne*.
*Note:* To contain the Paris Region growth, a plan was proposed in 1960, the PADOG, to limit urban development to the already built-up areas, with reorganization and equipment of this central part, partly thanks to transportation infrastructures, while the rest of the region should remain unbuilt. The coming to power of President De Gaulle is a turning point in the planning policy for Paris. In 1965 the SDAURP plan to organize the scattered and under-equipped suburbs, under the Paul Delouvrier's plan.
*Sources:* IAU - Ile-de-France; Mayer and Trevien (2015)

**Figure 2.8:** Effective implementation of the RER network in 1990 and 2010.




*Transilien* is the rebranding of the suburban trains not included in RER lines, that benefited from improvement after the success of the RER with modernization of stations and rolling stock.
*Note:* The RER project, followed by the *Transilien* starting in 1999, actually results in upgrading most of existing lines, with interconnections under the historical city core of Paris. It included the construction of few new *branch* lines towards airports and "New Towns". The five RER lines progressively opened between 1969 and 2004, and reaches the goals assigned by the 1965 SDAURP, with improved rolling stock and frequencies.
*Sources:* IAU - Ile-de-France; Mayer and Trevien (2015)

CHAPTER 3

# The Origins of the *Gilets jaunes* movement

with Pierre Boyer, Germain Gauthier, Vincent Rollet, and Benoit Schmutz

**Abstract**

This chapter presents the results of a geographical study on the areas where the Yellow Vests (*Gilets jaunes*) movement first emerged. This grassroots movement was organized at a local level but developed throughout France as soon as the protests began in November 2018. Using new Facebook data related to the movement, we show a strong geographical correlation between online mobilization (on Facebook) and offline mobilization (blockades of roundabouts). We precisely map the protests in France. Then, using data on roads where speed limits were lowered during the summer of 2018, and on the average commuting distance in French cities, we show that the issue of mobility is an important explanatory factor for the initial growth of the movement.

**Keywords**: Yellow Vests; Protests; Inequality; Urban Economics; Online Mobilization.

**JEL Classification**: F15, J40, J60, J80, C83.

Published under the following reference: Boyer, Pierre, Delemotte, Thomas, Gauthier, Germain, Rollet, Vincent, and Schmutz, Benoît. «*Les déterminants de la mobilisation des Gilets jaunes*», Revue économique 2020/1 Vol. 71, p. 109-138 (2020).

We would like to thank Thierry Kamionka and two anonymous peer reviewers, as well as Micael Castanheira, Julien Grenet, Fanny Henriet, Nolwenn Loisel, Phoebe MacDonald, Clément Malgouyres, Audrey Rain, Anasuya Raj, and Clémence Tricaud for their comments.

We would also like to thank Francis Kramarz for financing access to the data for two of the authors, the “Investissements d’Avenir” program (ANR-11-IDEX-0003/Labex Ecodec/ANR-11-LABX-0047), the producers of the data used in this study (INSEE), the Comité du secret statistique (Committee on Statistical Confidentiality) for the procedure for accessing these data, and finally the Centre d’accès sécurisé aux données (CASD) (Secure Access Data Center) for the practical details of the use of these data.

The results presented in this chapter are the sole responsibility of the authors.

# Introduction

At the end of 2018, France was shaken by a large-scale protest movement: the "*Yellow Vests*" (Yellow Vests in French). The movement initially brought together motorists angry about rising fuel prices. But it quickly turned into a general protest against government policy. This movement stood out due to its local character and its nationwide coverage. Its members were encouraged to block traffic as close as possible to their homes, resulting in an unprecedented number of blockades on the very first Saturday of protests. Facebook seems to have played an important role in the success of this decentralized mobilization.

Although the increase in the tax on energy products was the trigger for the movement, it does not seem to be the only explanation. Indeed, various protest movements took place at the beginning of Emmanuel Macron's presidential term without coalescing in this way. Besides dissatisfaction with the political choices made by the government, there were also fundamental questions about public policies that had been in place for several decades, in a context of declining public spending and increasing inequality, with significant territorial repercussions. Finally, the mobilization took place in the context of an apparent questioning of liberalism and social democracy in many modern democratic societies.

The Yellow Vests belong to a long history of social movements in France Neveu (1996). In some respects, they reproduced familiar historical patterns. Indeed, fiscal policies are often evoked to explain the origins of popular uprisings (Ponticelli). In France, from medieval peasant uprisings to the price of gas at the pump, fiscal revolts have appeared frequently throughout the history of taxation (Delalande, 2011, 2014). However, the Yellow Vests also differed from traditional protests in several ways. First, the movement was distinguished by its use of numerous decentralized gathering points, often at roundabouts, symbols of French car culture. On the very first Saturday of protests, there were 788 blockades. Second, the demonstrations seem to have been organized largely without the intervention of traditional intermediary bodies such as political parties or unions: the latter were slow to join the protest movement. Finally, social networks seem to have played a decisive role in the organization and media coverage of the movement. By mid-December, there were 1,548 Facebook groups with more than 100 members associated with the Yellow Vests movement. Although massive demonstrations had previously been launched and catalyzed by social networks around the world,[1] this was the first time that a mobilization of this magnitude had been sparked by social media in France.

Recent theoretical work has highlighted the potential importance of social networks in the emergence of large-scale protest movements (Edmond 2013; Little 2016; Barbera and Jackson 2018). It has long been known that coordination is key for effective collective action, but that it is limited by information asymmetries and communication channels (Sandler and Blume 1992; Russell 1982; Ostrom 2015). From this point of view, social networks are potentially game-changing for citizens and governments alike. They have the potential to facilitate access to other, sometimes more reliable sources of information, especially in autocratic regimes, and allow for better coordination of strategies among protesters. Some empirical studies seem to support predictions based on these models. In the context of the Arab Spring, for example, Twitter activity was a strong predictor of actual protests (Steinert-Threlkeld et al. 2015; Acemoglu et al. 2017). Similar results have been found

[1] The Arab Spring and Occupy Wall Street in 2011, La Manif pour tous in 2012, Nuit debout in 2016, etc.

in Chinese (Qin et al. 2017), American (Caren and Gaby 2011; Vasi and Suh 2013; Bastos et al. 2015), and Russian (Enikolopov et al. 2017) contexts.

The media coverage received by the Yellow Vests was extensive. On tele-vision, radio, and in newspapers, numerous interpretations were put forward to explain the emergence of the movement.[2] In the academic context, Sebbah et al. (2018) have analyzed the Yellow Vests movement through the prism of traditional media (newspapers) and social networks using textual analysis techniques (Facebook, Twitter, petitions on change.org). Their analysis highlights the importance of the themes of mobility and fiscal policy, as well as the discontent of the protesters. Using the CEVIPOF Political Trust Barometer, Algan et al. (2019) have studied the sociodemographic characteristics of Yellow Vests supporters. Their analysis shows that these supporters are mostly far-left and far-right voters, or nonvoters. It should be noted that the survey focuses on the political positioning of these supporters, but it did not question the protesters themselves. Finally, Bennani et al. (2019) have examined the local factors involved in digital participation in the "*grand débat national*" (great national debate). They show that median standard of living and level of education are the main determinants at the departmental level of overall participation in the "*grand débat en ligne*" (great online debate). To our knowledge, no study has examined the connections between physical blockades (at roundabouts) and online activity (Facebook).

In this chapter, we approach the Yellow Vests movement through the prism of geography. We aim to answer the following question: What are the characteristics of the geographical areas with a high level of mobilization at the beginning of the movement? In this, our study differs from the two works mentioned above in several ways. First, we are interested in the mobilization itself (and not its support in the wider population), on Facebook and in the physical territory (blockades). We document the Facebook groups associated with the Yellow Vests that we were able to locate. We thus highlight a strong correlation between online and offline activity, before mapping the mobilization in these two dimensions. Second, we built a geolocated database that combines administrative sources (jobs, income, voting history) with offline (blockades of roundabouts) and online(Facebook groups) mobilization indicators. Our econometric study highlights the strong link between mobilization and variables related to mobility: commuting distance and the reduction of the speed limit on secondary roads from 90 km/h to 80 km/h.

The rest of the chapter is divided into six parts. First, we briefly review the history of social movements in France and the major events during Emmanuel Macron's presidency that led to the emergence of the Yellow Vests. Second, we describe in detail the collection of our data and we map the movement. Next, we discuss the factors that could explain mobilization, which we group into four categories: political preferences, government decisions, socioeconomic factors, and geographical constraints associated with the territories. We then present the results of the econometric analysis. Finally, we conclude by discussing prospects for future studies.

[2] A keyword search for "Yellow Vests" gives 181,563 journalistic articles published between October 1, 2018, and July 1,2019 (*Source:* Factiva).

## 3.1 An Atypical Social Movement

By disrupting the established political order, protest movements have the potential to bring about major economic and political changes.[3] For social scientists, it is important to understand the origins and consequences of these movements. However, their erratic nature and their mutation over time make them a challenge for research.

By adopting the definition provided by **?** we conceive a social movement (or mobilization) as “an intentional acting-together, marked by the protagonists’ explicit plan for concerted mobilization. This acting-together is centered around a demand, the defense of an interest, or a ’cause’".[4] In this chapter, we focus on the different modes of coordination implemented in order to organize collective action, in particular the link between online(Facebook) and offline(blockades of roundabouts) means, by mapping these two dimensions of the movement and by highlighting the links between the territory and these forms of mobilization.

### 3.1.1 Historical dynamics of social movements

The second half of the twentieth century was a period of major change in social movements, linked to social and economic transformations (deindustrialization of Western economies, end of the Cold War, and globalization).[5] Thus, the workers’ movement, characterized by its acts of opposition within factories (pitting workers and unions against the owners of the means of production), declined from the 1960s onward, with the end of “workingclass consciousness” and the rise of the service economy. 1968 marked a turning point with the appearance of “new social movements” (Touraine 1968; Parkin 1968; Crouch1978), which saw the emergence of new figures of collective action such as students, regionalists, feminists, and the LGBTQ community. These groups made cultural demands, in what were now transversal struggles (Foucault1982) that were not limited to one country in particular and that addressed the effects of power as such.[6] These new social movements were accompanied by the affirmation of the subject and of the global dimension of struggles,[7] which led, from the 1980s onward, to the emergence of *global* movements (Della Porta et al. 2015) structured around non-governmental organizations (such as the alter-globalization movement) and based on the combination of very local actions and a global structuring.

When we look at social movements from this historical perspective, it is clear that what makes the events of late 2018 stand out is the Yellow Vests’ use of new tools (social networks) and the actors involved in the action (motorists). In particular, this mobilization did not relate to means of production, either in its modes of action or in its demands; nor did it, like the new social movements,

---

3 Numerous case studies have highlighted that protests structure voters’ political preferences. In the nineteenth century in the United Kingdom, the widespread uprising of agricultural workers influenced the propensity of voters to vote for reform of the British electoral system (Aidt and Franck 2015). More recently, the 2009 Tea Party protests in the United States significantly contributed to the increase in the vote for the Republican Party (Madestam et al. 2013). Furthermore, the threat of protests can impact a government’s public policies. Typically, in the case of fiscal policies, governments may move away from a system of optimal taxation in order to avoid tax revolts (Passarelli and Tabellini 2017).

4 **Translator’s note:** Unless otherwise stated, all translations of cited foreign-language material in this chapter are our own.

5 A more detailed general overview of these historical and social dynamics can be found in Wieviorka (2008).

6 This last point involves the disappearance from protestors’ demands of an identified oppo-nent. New social movements questioned the forms of domination and power in society, in both the public sphere (politics, work) and the private sphere (violence against children, women’s place in the home, etc.).

7 The individual became the driving force of action, in their very choice to *join* the collective action, thereby giving rise to ethical considerations and a new perception of the total individual as a “rights-bearing subject” (Touraine 2013).

involve minority figures;[8] nor did it possess the supranational structuring characteristic of global movements. Nevertheless, we can associate with it certain characteristics of recent movements, which together seem to form a turning point, with the use of social networks for the organization, structuring, and mediatization of collective action. A unique aspect of the Yellow Vests movement seems to lie in the connection between local gatherings and an effective national coordination.

### 3.1.2 The context of the Yellow Vests

Various mobilizations have taken place in France since the 2000s, often taking the government and the president as their adversary.[9] One action that echoes the demands of the Yellow Vests is the opposition to the ecotax in Brittany against the backdrop of layoffs in the agri-food industry, which gave rise to the "*bonnets rouges*" (red caps) movement in 2013. Since Emmanuel Macron took power, various mobilizations have hit the headlines: against the easing of the wealth tax, against the increase in the generalized social contribution, against labor law reform in 2017, and then, in the spring of 2018, against the opening up of railways to competition[10] and against the reform of the system for admission to higher education (Parcoursup). None of these mobilizations, which involved up to several hundred thousand demonstrators, resulted in changes to government policies.

In the study of collective action, motorists and motorcyclists are atypical populations. These groups often organize for recreational purposes, but they also monitor changes in road regulations and sporadically mobilize to influence government policies (the use of speed cameras, reduction of speed limits, fuel taxes). Demonstrations took place in January 2018 to oppose Prime Minister Édouard Philippe's plan to lower the speed limit on secondary roads from 90 km/h to 80 km/h, which brought together several thousand demonstrators in different cities.[11] On May 29, 2018, a motorist, Priscillia Ludosky, launched a petition on change.org, "For a decrease in fuel prices at the pump!," against the backdrop of a continuous increase since 2016.

### 3.1.3 Chronology of events

#### The emergence of the movement

**Gas prices, speed limit, and general discontent.** In 2015, then-President François Hollande decided to gradually implement a carbon tax on top of the existing gas tax, in order to make diesel and gasoline after-tax prices converge. The carbon tax was confirmed in 2017 by newly-elected President Emmanuel Macron, even though oil prices had been increasing since 2016 and car-related

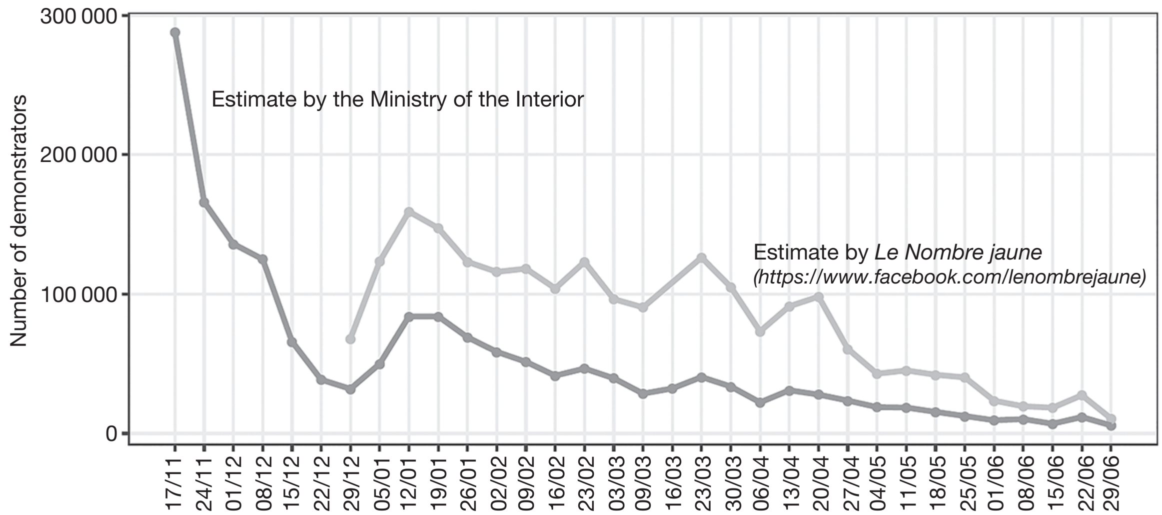
8 It is also interesting to note that the movement involved the (re)emergence of a majority actor, claiming to represent "the people." This characteristic connects the Yellow Vests to the contemporary populist dynamic and can probably help to explain the (initial) rejection of the movement by progressive forces.

9 These include, in chronological order: the so-called "banlieue riots" in 2005, which began in Clichy-sous-Bois following the death of two teenagers, Zyed Benna and Bouna Traoré; opposition to the "contrat première embauche" (CPE) (first employment contract) in 2006; and opposition to pension reforms in 2010. Opposition to the "El Khomri" labor law in 2016 lasted from March to September and gave rise to various innovations in collective action, with Nuit debout using a variant of the "city square movement" in the Place de la République, and the presence of "black blocs" in the leading group. These different mobilizations had major effects on public policies, including the amendment or withdrawal of the policies in question, and even the implementation of a state of emergency in the case of the 2005 riots.

10 This reform was decided on before Emmanuel Macron came to power, but it took effect during his term of office.

11 R. Bx with AFP, "Vitesse limitée à 80 km/h: des milliers de motards en colère contre 'la Sécurité rentière,' " *Le Parisien*, January 27, 2018. It is interesting to note that the article mentions that, "in addition to the 80 km/h limit, the demonstrators were also protesting against the high cost of living and the increase in the CSG [generalized social contribution], while singing *La Marseillaise*."

**Figure 3.1:** Estimation of the number of protesters per act.




*Note:* Number of estimated demostrators, November 17, 2018 to June 20, 2019. *Sources:* Ministry of Interior. Le Nombre jaune.

expenses had been increasing for several years.[12] A few months later, in January 2018, Prime Minister Édouard Philippe decided to decrease the speed limit on secondary roads from 90 km/h to 80 km/h, citing road safety concerns. This latter decision was not part of Emmanuel Macron's campaign manifesto and triggered the organization of many traffic slowdown protests throughout the country.[13] The new 80 km/h regulation went into effect on July 1, 2018.

By the end of the summer holidays, the yearly increase in the carbon tax was confirmed in the 2019 budget despite growing discontent,[14] particularly among motorists.[15] A well-known association of car users, "40 millions d'automobilistes" launched the initiative "Coup de Pompe" encouraging everyone to send their gas bill to the President. Several petitions were also launched online to alert the government of the impact of gas prices on purchasing power. Other initiatives surfaced on social networks, including videos totaling millions of views.[16] Although these various individual initiatives were pointing out the same discontent, they failed to coordinate and gain momentum.

**The Seine-et-Marne cluster.** On October 12, a local newspaper in the Seine-et-Marne département (located in the Greater Paris area) reported on a petition launched by a local motorist, Priscilla Ludosky.[17] The petition was initially created in May 2018 on the platform Change.org[18] and had garnered fewer than 1,000 signatures by the time of the article. The day following the article's publication, the number of signatories in Seine-et-Marne tripled. Meanwhile, an association of car users called "Muster Crew" was planning to block the Parisian ring road to protest against the

12 In 2018, the Automobile Club Association estimated that car-related expenses (including gas, insurance, tolls, fines, technical control) had increased by 3% to 4.6% in a single year (link).

13 See, for example, here.

14 See, for instance, here.

15 See, for example, here, here, or here.

16 The video that received the largest audience was uploaded by Jacline Mouraud and totaled 6 million views by the end of year 2018; see here.

17 See here.

18 See here.

increase in compulsory car-related expenses.[19] The protest was scheduled on November 17. Eric Drouet, a Seine-et-Marne resident, was the leader of this initiative. He shared the Press article about Ludosky's petition on the Facebook account of his association.

**The movement goes national.** On October 21, *Le Parisien*, a national newspaper, wrote an article about the petition, where it was explicitly linked with the planned blockade.[20] The petition was then reported on extensively in the media and the number of signatures skyrocketed, reaching 724,225 by November 16. The call for a mobilization on 11/17 started to generate interest on the Internet, where several videos were posted, mostly on Facebook or YouTube, to urge people to join the movement. A video published by Frank Buhler, from the southern département of Tarn-et-Garonne, soon went viral.[21] On October 24, Ghislain Coutard, a resident of the southern city of Narbonne, suggested that supporters should put their high-visibility vest under their windshield as a rallying sign.[22] This vest has been mandatory to carry in cars since 2008 and is called a "*gilet jaune*" in French. The large responses both to the petition and to the call to block the Parisian ring rapidly prompted people to plan their own local events for 11/17.

A website (link) was created to coordinate the mobilization. It provided a map of the organized blockades, updated in real time. As of November 16, the map documented 788 proposed blockades. Most locations for their potential to block traffic and economic activity, but the main targeted infrastructure was the roundabout (*rond-point* in French). Some large cities had multiple blocking points – for example one in the city center and another in the outskirts, close to a shopping mall. The demonstrators wore the gilet jaune, giving the movement a strong visual identity. Many blockades took place in areas with no history of demonstration. We show the substantial correlation of online mobilization in predicting these blockades in the next chapter. This online organization allows to reach remote locations, beyond the urban fringe, where most of blockades took place during the first Act. In the absence of a national coordination, police officers simply recorded undeclared demonstrations and reported traffic violations. The Ministry of Interior reported that approximately 300,000 demonstrators had participated in the first Saturday of action.[23] The success of this day fostered the planning of following events, subsequently referred to as "Acts", on both existing and new Facebook groups.

**A month of major mobilization.**

The first Act was the peak of the offline mobilization, and was followed by a steady decrease in participation on each Saturday following 11/17. The Ministry of Interior reported an important number of law violations in low-density areas and small cities during the first Saturday of protests.

The events of December 1 received considerable media coverage. Almost nobody signed the Change.org petition afterwards (see detailed in chapter

19 Note that the group's name and members have changed since the initial announcement of the event (link).
20 See here.
21 See here.
22 See here.
23 "France Police - Policiers en colère" (a police association supporting the movement) reported over a million of protesters.

**Follow-up and decline of the movement.**

After the holidays, the government decided to evacuate the remaining camps on the roundabouts and protesters began to concentrate in large cities, where their numbers were much lower than during the first month (despite a temporary increase on January 12). Impressed by the political impact of the movement, labor unions had started supporting the mobilization at the end of 2018, and offered to supervise the organization of the demonstrations in the absence of a clear organizational structure. Since the beginning of the movement, however, participants had rejected the idea of having a leader or a spokesperson. This feature made it particularly difficult for the government to negotiate with the demonstrators and put a clear end to the protest.[24]

From the onset of the movement, the demands and composition of its protesters evolved significantly. Initially fueled by anger towards fiscal reforms, the movement progressively evolved into a full-blown protest against the government and the political class. Concerns related to direct democracy and parliamentary oversight emerged and received some support from the population. On January 15, the government launched the Grand Débat, the purpose of which was to acknowledge grievances on four broad topics proposed by Emmanuel Macron: taxes, ecology, democracy and public services.[25] Some Yellow Vests reacted by launching an alternative platform with their own debate called *Vrai Débat* (literally, "True Debate"). The Grand Débat ended on March 15. The next day, an unauthorized demonstration on the Champs Elysées led to major acts of vandalism, including the burning of a famous restaurant (the Fouquet's) and the ransacking of many luxury stores. However, this surge was short-lived and the following Saturdays were far more quiet. On April 25, Emmanuel Macron held a press conference following up on the Grand Débat and, among other proposals, announced a reduction of the income tax and the re-indexing of small pensions to inflation.

In the following months, symbolic actions tended to replace weekly demonstrations and often targeted Emmanuel Macron's public appearances.[26] At the end of the year, groups of Yellow Vests took active part in the widespread demonstrations against the pension reform.[27]

**The movement fails to create a political party.** The European election campaign kicked off at the beginning of 2019, but it did not seem to generate much interest among the Yellow Vests. Since its inception, the movement had been reluctant to transform into a more classical political organization. On November 29, 2018, some members released an official statement announcing that based on a survey of about 30,000 members, they would bring forward 42 requests to the government.[28] Most demands were aimed at increasing the purchasing power of the middle class, and the Prime Minister accepted to meet with delegates of the movement. However, after those delegates began receiving threats from other protesters who wanted to avoid association with the government, the meeting was canceled.[29] We come back in chapter 4 on the rise of negative sentiment and antagonist messages of Yellow Vests' Facebook pages. The European elections took

24 A branch of the Yellow Vests that was seeking for a more structured movement created the so-called "Assembly of the Assemblies" (*Assemblée des Assemblées* in French), which brought together representatives from dozens of delegations, without major result.

25 See here.

26 See here and here.

27 See here.

28 See here.

29 See here.

place on May 26 and the two lists making explicit reference to the Yellow Vests collected less than 1% of the votes.

## 3.2 Description and Measurement of the Mobilization

To quantify offline mobilization, we used a map of the blockades planned for November 17, 2018. We also identified more than 1,500 Facebook groups created before December 13, 2018, whose activity we measured in order to quantify the size of the online mobilization. From this information, we were able to build a database of Yellow Vests mobilization.

### 3.2.1 Offline mobilization

After the call for a national blockade on October 10, and owing to the increasing number of rallies planned throughout the country, a website was set up to list the different actions planned and to facilitate coordination of the November 17 mobilization.[30] This website offered an interactive map of the rallies, updated in real time. In this chapter, we use the map of rallies planned for November 17 (which was recorded on the evening of November 16). This map refers to 788 rallies announced in mainland France, which are geocoded and can each be associated with a particular *commune* (municipality). These are declarationsof an *intention to demonstrate*, made by the Yellow Vests themselves on the eve of the mobilizations. To our knowledge, there is no exhaustive inventory of the rallies that took place. However, as this map was primarily intended to coordinate the rallies, there was little incentive to declare false intentions.

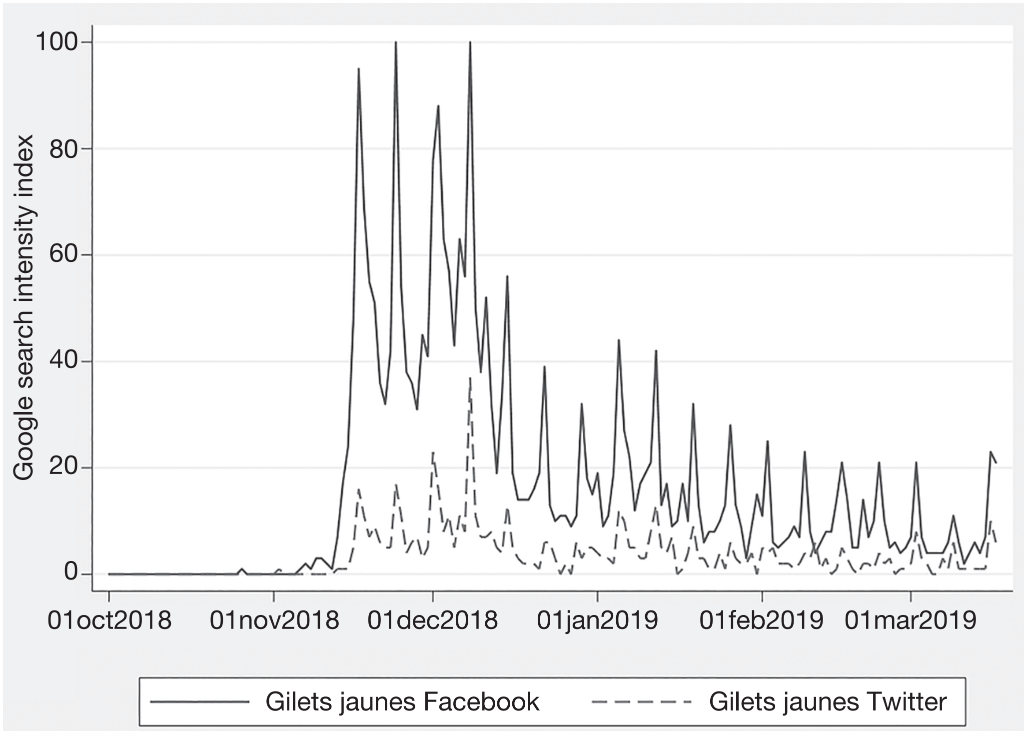
### 3.2.2 Online mobilization

The online activity of the Yellow Vests seems to have been concentrated on the social network Facebook. It was on Facebook that certain famous figures of the movement, such as Éric Drouet or Maxime Nicolle, made their names, and the websites connected to the movement (first www.blocage17novembre.fr, then www.gilets-jaunes.com and www.giletsjaunes-coordination.fr) coordinated the rallies by listing local Facebook groups.[31] The data provided by Google Trends, using the search terms "Yellow Vests Facebook" and "Yellow Vests Twitter," also indicate that Facebook was the dominant means of coordination (see 3.2).[32] When studying the coordination websites mentioned above, it seemed to us that the movement was organized more within Facebook groups than within Facebook pages. This may be partly explained by changes made to Facebook's algorithms in 2018, which then favored groups over pages.

We therefore decided to measure the online activity of the Yellow Vests by listing the Facebook groups as exhaustively as possible. In the absence of an open API, we performed this work manually between December 12 and 15, 2018. This collection therefore gives us a snapshot of Facebook groups

30 The address was: www.blocage17novembre.fr.
31 We also note the unequal penetration of Twitter across the country compared to Facebook. Twitter is mainly used in Paris, but even there it has a low number of users. Facebook has stronger penetration in the population and is better distributed across the country.
32 Google Trends is a commonly used tool in many research fields. For a detailed discussion, see Stephens-Davidowitz (2014) and Stephens-Davidowitz and Varian (2014).

**Figure 3.2:** Change in interest in the topics “Gilets jaunes” between Facebook and Twitter.




*Note:* Google search corresponds to “gilets jaunes Facebook” (continuous curve) and “gilets jaunes Twitter” (discontinuous curve). *Sources:* Google.

as of mid-December 2018. We used this list of groups to analyze the preparation of the movement when it was launched in October, as well as in its first month of existence.

To carry out this inventory of Facebook groups, following the methodol-ogy of Caren and Gaby (2011), we performed search queries on Facebook using a series of keywords related to the movement and associated, or not, with geographical indicators.[33] For each group thus identified, we retrieved the name of the group, its creation date, the number of members, and the number of messages posted. While discounting groups with fewer than 100 members,[34] this method allowed us to identify 1,548 different groups. In Figure 3.3, we show the change in the number of Facebook groups over time. The two intense phases of group creation correspond to the chronology mentioned above.

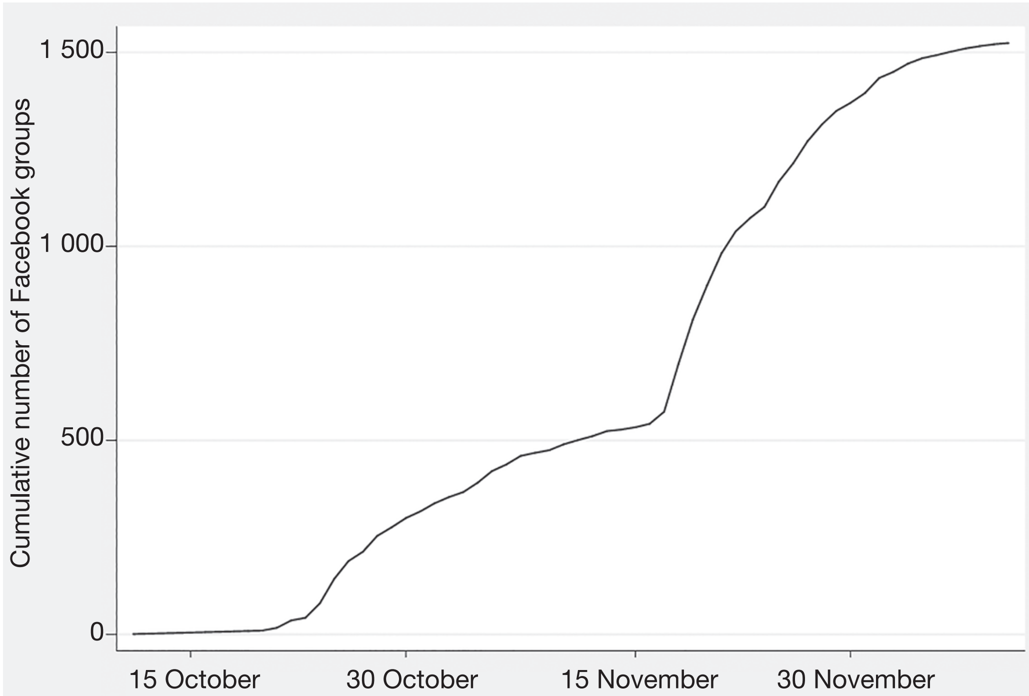
These different groups were then associated with an identifiable geographical level: national, regional, departmental, or infradepartmental (on the scale of a city or an urban area), depending on the explicit references present in the name of the group (for example, “Les Yellow Vests de Savoie,” “Gilet Jaune 74,” “Mobilisation Yellow Vests Senlis”). As detailed in Table 3.1, more than half of the groups analyzed were associated with a city, a small group of cities, or a “pays” (small region), and only a little more than 25% of the groups focused on a scale larger than that of the department.[35]

33 Some of the keywords used include “Yellow Vests Rennes,” “blocage,” “blocage Ain,” “colère,” “17 novembre Hauts-de-France,” “hausse carburant,” etc.

34 This decision was made to allow us to identify only truly active groups. The excluded groups represent only 1.1% of the total number of members.

35 It is important to note that members were not identified and thus may belong to more than one group.

**Figure 3.3:** Change in the number of Facebook groups over time.




*Sources:* Facebook.

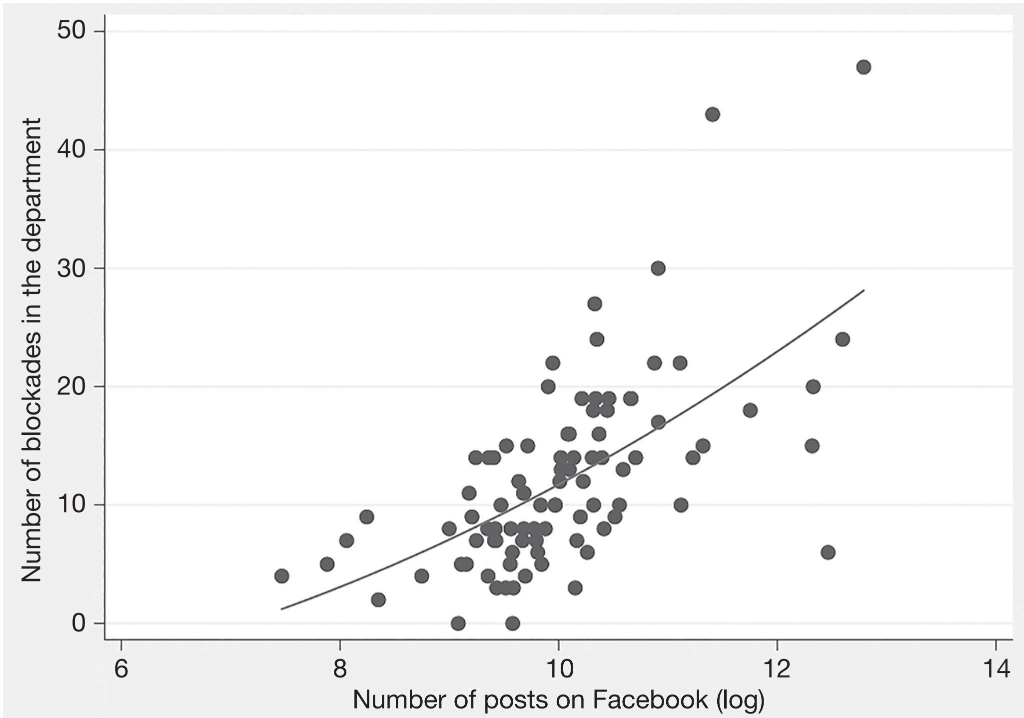
Moreover, almost half of the posts were made on local groups (infradepartmental scale), while just under 20% of total posts were on national groups. These observations reflect the local and decentralized nature of the movement. We used textual analysis methods (described in Appendix 3.B) to associate each group with a geographical entity (region, department, commune, or group of communes).

Despite the large number of searches carried out, we cannot guarantee the exhaustiveness of our database. Moreover, the use of data from social networks involves systematic selection biases. For example, in France, women, young people, people in lower socioprofessional categories, and voters of populist parties are overrepresented on Facebook. By contrast, members of intermediate professions, inhabitants of the Paris region, and voters of Emmanuel Macron (during

**Table 3.1:** Characteristics of the Facebook groups analyzed

| Scale of the group | Number of groups | Number of members | Number of posts before December 14, 2018 |
|---|---|---|---|
| National | 287 | 2,454,585 | 281,365 |
| Regional | 113 | 254,068 | 138,739 |
| Departmental | 317 | 529,412 | 323,217 |
| Local | 834 | 1,025,628 | 730,295 |
| Total | 1,548 | 4,263,693 | 1,473,616 |

**Figure 3.4:** Link between online and offline mobilization.




*Note:* Link between online and offline mobilization based on the location of these two modes of action. Each dot corresponds to a department. *Sources:* Gilet jaunes and Facebook.

the presidential election) are underrepresented.[36] Online mobilization also took place on other social networks (such as Twitter and WhatsApp), but the comparison of online behavior on different social networks during the movement is beyond the scope of this study.[37]

### 3.2.3 Mobilization indicators selected

With these two sources of compiled data, reflecting both online and offline mobilization, we constructed the following three indicators:

- *Number of rallies* planned by geographical area;
- *Number of members* of Facebook groups associated with each geographical area;
- *Number of posts* on Facebook groups associated with each geographical area.

The online and offline mobilization indicators are positively correlated, as shown in Figure 3.4. It seems that the relationship between these two modes of mobilization is not necessarily linear. In particular, we observe a greater range in the number of posts when this number is high, and its predictive capacity for the number of blockades then seems to be low. This is partly a result of the right-censoring of the number of posts per group at 10,000 imposed by Facebook. On the other hand, the predictive capacity of the number of groups, and even more so, the number of members, is high.

To further investigate the relationship between these mobilization indicators, Table 2 reports

36 Based on the IFOP survey *La confiance des Français dans les réseaux sociaux après l'affaire Cambridge Analytica* (FD/FK/JPD no. 115394) of March 2018.

37 This would require greater accessibility and interoperability of data from different platforms.

**Table 3.2:** Correlations between online and offline mobilization variables (departmental level)

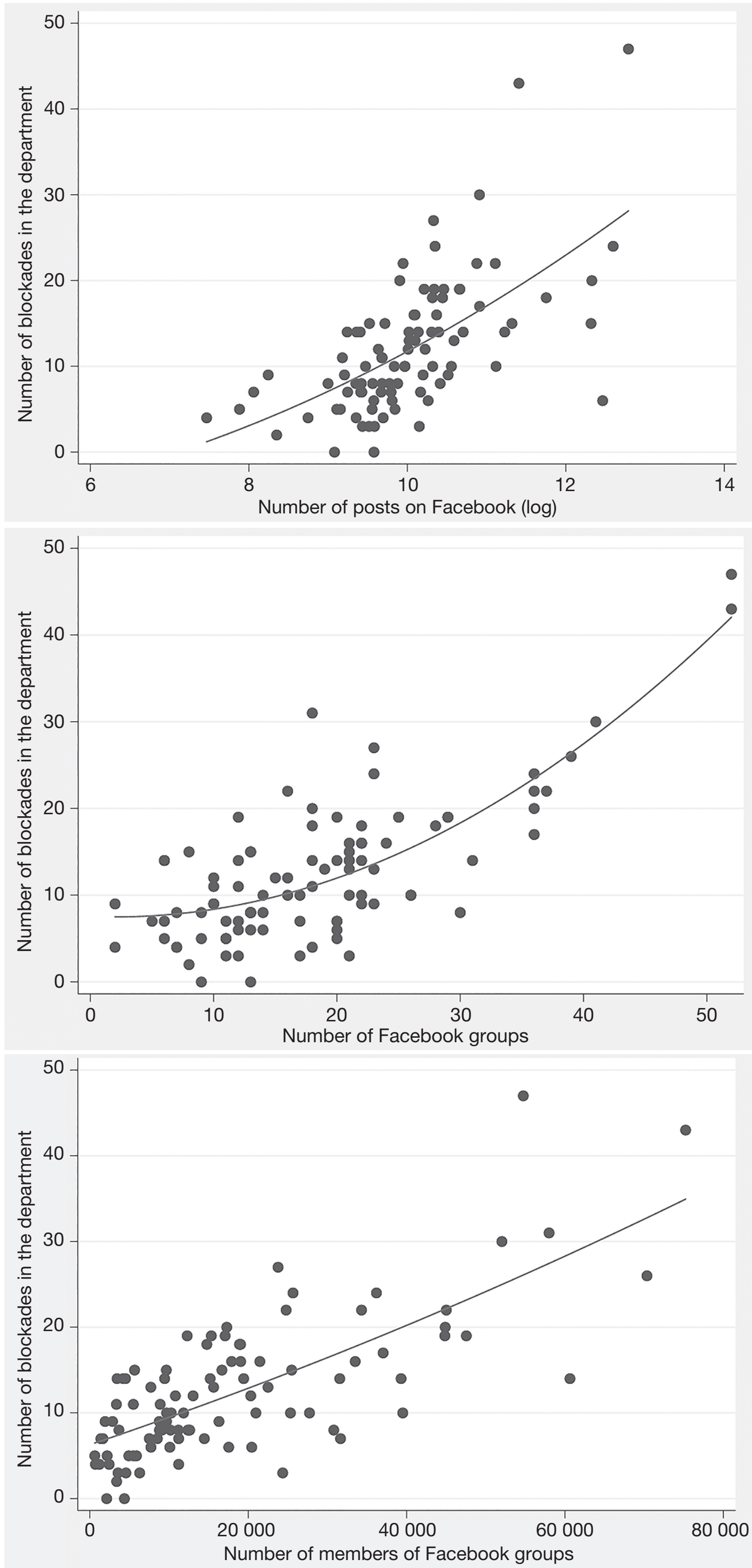
| | Blockades | Groups | Members |
|---|---|---|---|
| Groups | 0.74 | | |
| Members | 0.71 | 0.86 | |
| Log (Posts) | 0.62 | 0.74 | 0.69 |

*Note:* 1) "Members" and "Blockades" refer to the number of members of Facebook groups and the number of blockades planned for November 17, respectively, per $km^2$; 2) All variables are centered and reduced.

their level of correlation at the departmental level. There is a strong correlation between the number of members of Facebook groups and the number of blockades, at 71%. The relationship between the number of posts and the number of blockades is less strong, at 62%, which suggests the lower quality of this variable, due to the 10,000-post limit for groups. The correlation is, of course, strongest with the number of groups (74%), but we chose to focus on the number of members in the rest of our analysis, as it gives an idea of the scale of mobilization for each given event, which our blockade variable does not enable.

These very strong correlations testify to the close connection between the coordination of the movement via Facebook groups and physical action points. These results echo Enikolopov et al.'s analysis (2017) of the links between the use of social networks and the structuring of political protests. For the rest of the study, we focus, for the online aspect of the movement, on the number of members of Facebook groups. The offline variable, meanwhile, is measured by the number of declared blockades of roundabouts.

**Figure 3.5:** Link between online and offline mobilization.




*Note:* Link between online and offline mobilization based on the location of these two modes of action. Each dot corresponds to a department. *Sources:* Gilet jaunes and Facebook.

## 3.3 Mapping the Mobilization

### 3.3.1 Methodology

The Yellow Vests mobilization involved different territorial layers, from local rallies (blockades, roundabout occupations) to protests against government decisions on the national scale. We focus on two geographical partitions of France: the departments and the employment zones. The scale of the department is useful for the study of Facebook groups, as many of these groups identify themselves with this administrative and historical nomenclature (as shown by the large number of Facebook groups with the name or number of a department in their title). The relative homogeneity of departments in terms of their size (excluding Paris, its inner ring, and the Territoire de Belfort) and their spatial composition (a central hub—the prefecture or departmental capital—and various sub-hubs) makes them a relevant unit of study for spatial comparisons. Of the 96 departments in France, we retained only those of comparable size for the empirical study, which gave us 89 usable observations.[38] This nomenclature is particularly useful for the study of the online movement, as it corresponds in a significant number of cases to the scale of identification used by the agents themselves.

The employment zone is a geographical space defined by the Institut national de la statistique et des études économiques (INSEE) (National Institute of Statistics and Economic Studies) based on the analysis of commuting patterns: most individuals grouped in the same employment zone work and reside within this geographical area. It therefore seemed appropriate to use this scale for the study of blockades connected to the Yellow Vests mobilization. There are 296 such zones in mainland France. It is interesting to note that departments and employment zones form two non-overlapping divisions of space.[39] We describe this choice of dual nomenclature in more detail in Appendix 3.A.

We observe an average of 13 blockades per department.[40] At the level of employment zones, the average is 2.3 blockades per zone. Some employment zones have an unusually high intensity of blockades.[41] At the level of Facebook groups, we observe an average of 19,300 members per department and 3,347 members per employment zone.[42] The maximum values, which are very high, highlight the very strong heterogeneity of online mobilization at the level of employment zones.

### 3.3.2 Spatial analysis

We present different maps of the spatial reach of the mobilization.[43] A first series of maps (Figure 3.6) represents the number of members of Facebook groups per department, as an absolute value (left) and per capita (right). These maps show a high intensity of online mobilization in peripheral territories: the entire Atlantic coast shows high levels of mobilization, as do the Mediterranean arc,

38 We did not take into account overseas departments and regions, mobilizations abroad, or the two departments in Corsica, in order to maintain a continuous geographical set.

39 For example, the Alençon employment zone covers three departments (Orne, Mayenne, and Sarthe), while the department of Orne covers five employment zones.

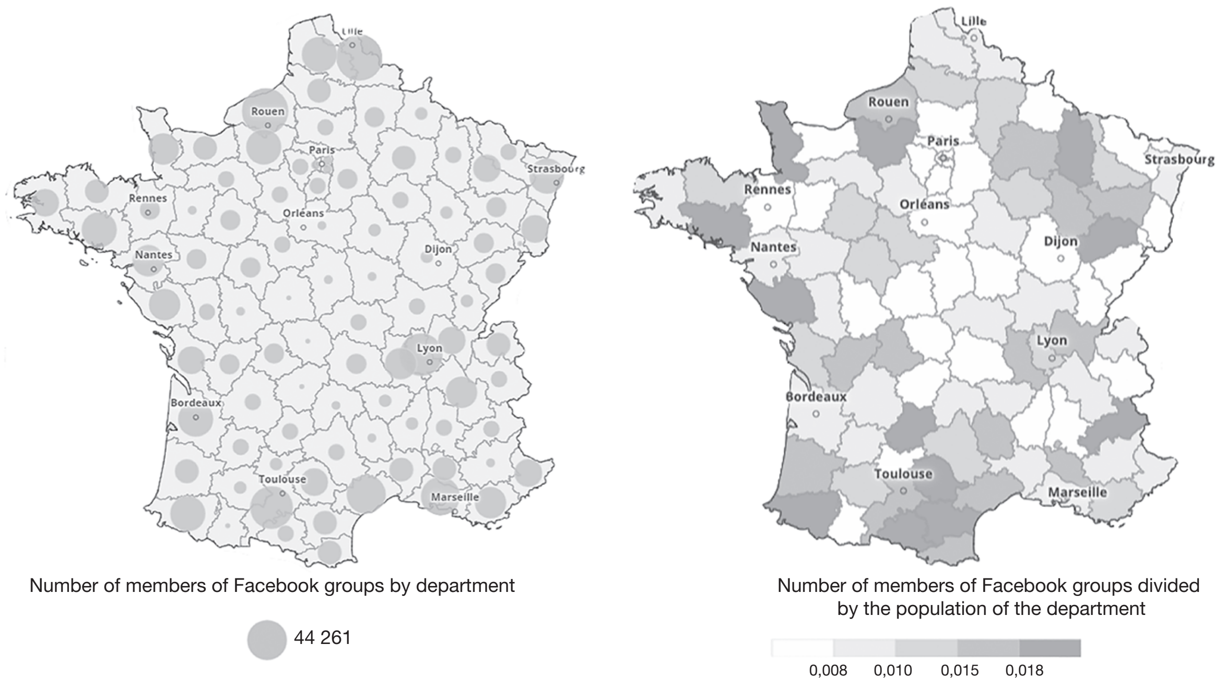
40 The highest values are located in the Bouches-du-Rhône, Nord, and Rhône departments, which all have high population densities. It should be noted that the departments in the inner ring of Paris had a low density of blockades.

41 The employment zones with a very high density of blockades were Troyes, Roubaix- Tourcoing, Lens-Hénin, and, to a lesser extent, Istres-Martigues and La Rochelle.

42 The highest values were observed for the employment zones Saintes–Saint-Jeand'Angély, Istres-Martigues, Neufchâteau, Troyes, Lens-Hénin, Douai, Argentan, Lorient, and Cherbourg-en-Cotentin.

43 Additional maps are presented in Appendix 3.C

**Figure 3.6:** Online mobilization by department.

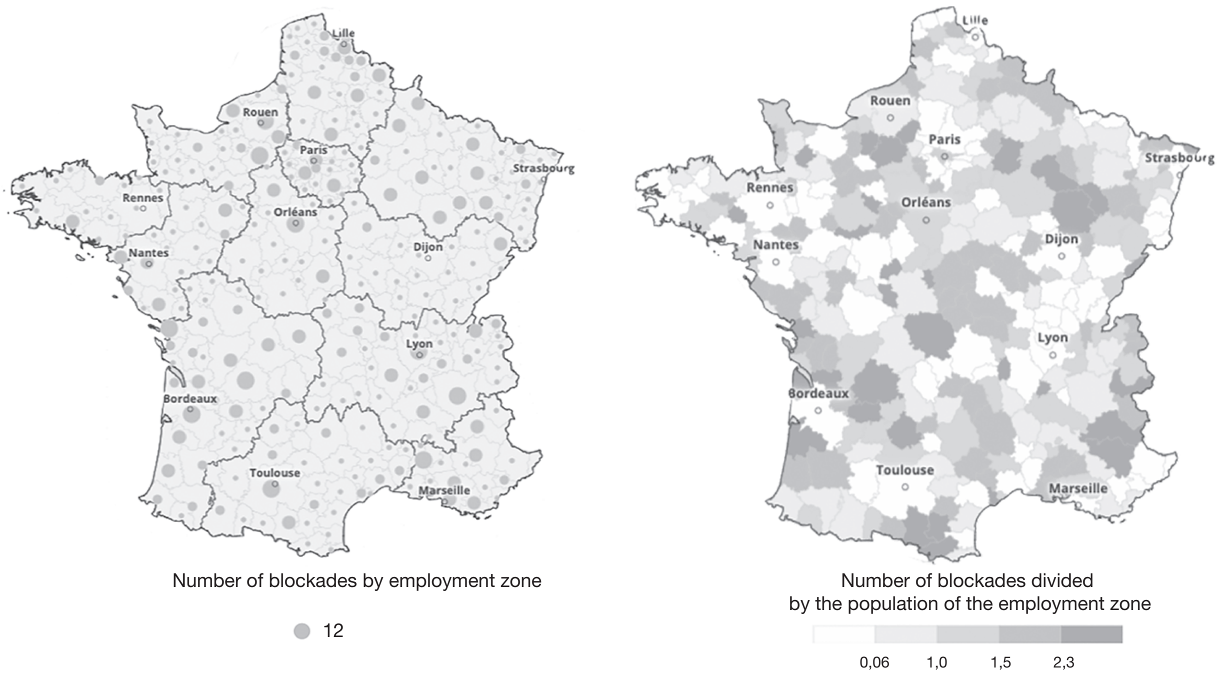



*Note:* These two maps represent the number of members of Yellow Vests Facebook groups by department, among those we were able to identify. The values are given as absolute numbers (left) and per capita (right). Orders of magnitude: On average, there are 18,593 members of Facebook groups associated with the Yellow Vests movement per department (left). Counting the number of members per capita, the average across departments is 263 members per 10,000 inhabitants (right). *Sources:* Data collected on Yellow Vests Facebook groups, between December 12 and 15, 2018. Map projection made using the online tool: `https://statistiques-locales.insee.fr/`.

the north, and Alsace. The spaces associated with the "empty diagonal" area of France (an area of relatively sparse population running from the northeast to the southwest) are among those with low activity on Facebook. However, some low-density departments are in fact highly mobilized if we relate the measure to the number of inhabitants, such as the Lot, Charente, or Hautes-Alpes departments.

A second series of maps (Figure 3.7) represents the intensity of blockades, based on the location of intended points of assembly at the employment zone scale: in absolute terms (left) and per capita (right). This series of maps differs from the previous one. In particular, we observe a much lower physical mobilization in Brittany compared with its online mobilization—the same is also true for Alsace. A Paris–Clermont-Ferrand axis, which is absent from the online mobilization, cuts through the empty diagonal. The nomenclature of employment zones allows us to identify a strong intradepartmental heterogeneity, for example in the Cher and Marne departments.

**Figure 3.7:** Offline mobilization by employment zone.




*Note:* These two maps represent the number of blockades by employment zone (left) and the number of blockades relative to the population of the employment zone (right). *Orders of magnitude:* On average, there are 2.63 blockades per employment zone (left). Counting the number of blockades per capita, the average across employment zones is 2.04 blockades per 100,000 inhabitants, with 21 employment zones having zero blockades (right). *Sources:* Data collected on the evening of November 16, 2018, from `www.blocage17novembre.fr`. Map projection made using the online tool: `https://statistiques-locales.insee.fr/`.

## 3.4 Four Dimensions of Mobilization

The determinants of mobilization are linked to many different issues. Four dimensions seem essential for highlighting the factors underlying the movement: electoral results, political decisions, the socioeconomic characteristics of populations, and the geographical characteristics of territories. We describe the variables selected below, and in Table 3.3 we present the correlations of the main variables selected with our mobilization indicators.

**1. Vote in the first round of the 2017 presidential election** . We collected electoral data on the first round of the 2017 presidential election, namely the abstention rate and the share of the vote obtained by each of the top five candidates: Emmanuel Macron, Marine Le Pen, François Fillon, Jean-Luc Mélenchon, and Benoît Hamon.

At both the department and employment zone scales, mobilization is negatively correlated with the vote for Emmanuel Macron, François Fillon, and Benoît Hamon, while it is positively correlated with the vote for Marine Le Pen and Jean-Luc Mélenchon, and to abstention. The correlation between abstention and mobilization is positive and quite significant. This is less the case for the voting variables, which are moderately correlated with our indicators: this first step seems to indicate that the movement does not amount to a notional third round of the 2017 presidential election, and that it has more to do with a rejection of the electoral process.

**Table 3.3:** Correlations between mobilization variables per km2 and explanatory variables

| | Department | | | Employment Zone | | |
|---|---|---|---|---|---|---|
| | Online | | Offline | Online | | Offline |
| | Memb. | Post. | Block. | Memb. | Post. | Block. |
| Share of diesel vehicles | – 0.36* | – 0.18 | – 0.31* | – 0.01 | – 0.05 | – 0.28* |
| Roads reduced to 80 km/h | 0.49* | 0.15 | 0.44* | 0.15* | 0.16* | 0.50* |
| Abstention | 0.21* | 0.08 | 0.38* | 0.11 | 0.12* | 0.28* |
| Voted for Macron 1st round | – 0.01 | – 0.06 | – 0.12 | – 0.12* | – 0.18* | – 0.11* |
| Average commute | 0.03 | 0.13 | – 0.12 | 0.15* | 0.15* | – 0.16* |
| Unemployment rate | 0.22* | 0.24* | 0.27* | 0.19* | 0.25* | 0.21* |
| Wage inequality | 0.49* | 0.30* | 0.52* | 0.12* | 0.14* | 0.31* |
| Average age | – 0.54* | – 0.27* | – 0.47* | – 0.04 | – 0.02 | – 0.32* |
| Population density | 0.53* | 0.34* | 0.66* | 0.11 | 0.12* | 0.80* |
| N | 89 | 89 | 89 | 296 | 296 | 296 |

*Note:* 1) Memb. is the number of members, Post. is the number of posts, and Block. is the number of blockades; 2) For variables marked with an asterisk, the hypothesis of non-correlation can be rejected with a 95% confidence level.

For the econometric analysis, we used the abstention rate and the vote for Emmanuel Macron as explanatory variables. These are the two variables most strongly correlated with our mobilization indicators. The abstention rate is interesting as a measure of level of engagement in political life and attachment to institutions (extensive margin). In cases where individuals did vote, the vote for Emmanuel Macron allowed us to analyze the extent to which the movement was constructed in opposition to the current government (intensive margin).

**2. The government's political decisions.** Beyond the increase in fuel prices, two political decisions seem to be strongly linked to the protest: the decision to increase taxes on diesel (carbon tax) and the reduction of the speed limit from 90 km/h to 80 km/h on secondary roads. The former policy was present in Macron's presidential manifesto,[44] while the latter, applied from July 2018, was not. The increase in fuel prices is difficult to measure at the local level, owing to spatial disparities in prices at the pump.

Based on administrative data, we were able to determine the share of registered diesel vehicles per commune. Surprisingly, this variable is negatively correlated with our mobilization indicators. However, this variable captures many other dimensions, such as social inequalities and mobility issues, which we also measured. In addition, it is worth noting that the increase in taxes on diesel, which was reaffirmed in September 2018, was scheduled for January 2019 (before being cancelled), and thus had not yet occurred at the start of the protest.

Using OpenStreetMap, we calculated the number of kilometers of roads affected by the

44 Objective 4 of his manifesto on ecological transition reads: "In order to massively reduce pollution linked to fine particulates, the taxation of diesel will be aligned with that of gasoline during the presidential term" (https://en-marche.fr/emmanuel-macron/le-programme/ environnement-et-transition-ecologique).

introduction of the 80 km/h speed limit.[45] In the context of our territorial study, we calculated the length of affected roads (km) in proportion to the surface area of the zone studied ($km^2$). This variable therefore indicates, for a representative square kilometer, the number of kilometers of roads affected by the reform.[46] There is a strong positive correlation between this variable and the mobilization indicators. Unlike the increase in diesel prices, this reform had just been implemented (July 1, 2018), and its effects were therefore beginning to be felt by commuters.[47]

**3. Socioeconomic factors.** Mobilization may also reflect economic disparities in France. To characterize this dimension, we used the unemployment rate as a measure of labor market integration (extensive margin). Within these local labor markets, we used the déclaration annuelle des données sociales (DADS) (annual declaration of social data) to calculate the 90/10 ratio in order to capture the level of wage equality (intensive margin).[48] These variables allowed us to characterize the populations present in the mobilized territories. It is important to take these variables into account in order to control for variations associated with political variables (voting and government measures), since announced political changes do not have the same impact on the various segments of the population.

In departments and employment zones, the variables of unemployment rate and inequality are strongly and positively correlated with mobilization, thus confirming the socioeconomic dimension of the movement. This indicates that the mobilized territories are those with higher-than-average unemployment rates and inequality levels, although this may be associated with other variables for which we control in the econometric study.

**4. Geographical constraints.** To account for mobility constraints, we calculated the average distance, as the crow flies, between employees' workplace and place of residence. This variable seems to us to be a good approximation of territorial constraints. It also indirectly reflects the impact of rising oil prices on territories.[49]

## 3.5 Econometric Analysis

As the above variables are sometimes highly correlated with each other, we conducted an econometric analysis to isolate the role of each one. Table 3.4 describes the estimation results of an ordinary least squares regression of our two main mobilization variables: the number of blockades on November 17 and the number of members of Facebook groups, both adjusted for the size of the area. We distinguish between an analysis at the departmental scale (columns 1 and 2) and at the employment zone scale (columns 3 and 4). There are nine explanatory variables, including two

45 We obtain a total length of affected roads of 390,000 km, similar to the estimates mentioned in the media (e.g., AFP, "Vitesse limitée à 80 km/h: ce qui va changer," *Le Point.fr*, June 27, 2018).

46 The variable that we constructed here may resemble a density, and thus capture the intensity of the reform across territories.

47 Commuters are defined as people making a daily journey to their place of work.

48 Formally, the 90/10 ratio is the ratio between the ninth and first decile of wage distribution.

49 According to the INSEE mobility survey (2008), excluding Paris, the share of public transport in commuting is very low. From 45% in Paris (which is excluded from our database), it falls to about 15% for Lyon, Lille, Grenoble, and Strasbourg, and is well below 15% for all other cities in the country. On average, three-quarters of all commutes, two-thirds of all journeys, and five-sixths of the total distance travelled are by private car. These figures have been relatively stable for several decades.

**Table 3.4:** Econometric results

| | | Department | | Employment zone | |
|---|---|---|---|---|---|
| | | Members | Blockades | Members | Blockades |
| Vote | Macron | – 0.0624<br>(0.121) | – 0.0918<br>(0.119) | – 0.225**<br>(0.0934) | – 0.307***<br>(0.0553) |
| | Abstention | – 0.0569<br>(0.105) | 0.103<br>(0.103) | 0.105<br>(0.0878) | – 0.00436<br>(0.0520) |
| Political decisions | Speed limit reduction | 0.295***<br>(0.105) | 0.117<br>(0.103) | 0.240***<br>(0.0863) | 0.134***<br>(0.0511) |
| | Share of diesel vehicles | – 0.0454<br>(0.110) | 0.123<br>(0.108) | 0.0587<br>(0.0886) | 0.00766<br>(0.0525) |
| Economy and society | Inequality | 0.256**<br>(0.126) | 0.125<br>(0.124) | 0.175*<br>(0.0894) | 0.0772<br>(0.0530) |
| | Unemployment | 0.122<br>(0.110) | 0.134<br>(0.108) | – 0.0683<br>(0.0918) | – 0.0677<br>(0.0544) |
| Geography | Distance | 0.248***<br>(0.0894) | 0.0891<br>(0.0879) | 0.127*<br>(0.0671) | 0.0234<br>(0.0398) |
| Additional controls | Age | – 0.366***<br>(0.101) | – 0.210**<br>(0.0988) | – 0.0114<br>(0.0763) | 0.0117<br>(0.0452) |
| | Density | 0.100<br>(0.123) | 0.488***<br>(0.121) | 0.0488<br>(0.0819) | 0.781***<br>(0.0485) |
| Regional fixed effects | | No | No | Yes | Yes |
| N | | 89 | 89 | 305 | 305 |
| Adjusted $R^2$ | | 0.513 | 0.529 | 0.104 | 0.687 |

*Note:* 1) "Members" and "Blockades" refer, respectively, to the number of members of Facebook groups and the number of blockades planned for November 17, both per km2; 2) All variables are centered and reduced; 3) Standard deviations are given in parentheses; $^{***}$, $^{**}$, and $^{*}$ denote the significance of results at 99.9%, 99%, and 95%, respectively.

additional controls: resident population density and the average age of residents. Population density controls for the mechanical correlation effect between population density and the probability of observing an event in a geographical area, a correlation that is indeed very significant, as shown in Table 3.[50] As for average age, it is the simplest and most transparent variable for controlling for local socio-demographic differences. Older territories were less mobilized. All variables are centered and reduced in order to facilitate the interpretation of the estimated coefficients.

We chose to carry out this statistical work both by department and by employment zone. Indeed, although the measurement of online mobilization is, for the reasons detailed above, more precise at the departmental level, the small number of observations it provides forces us to be very cautious in the selection of explanatory variables. In particular, it is not possible to ensure that an unobserved local characteristic correlated with the different variables in the database does not bias the results obtained. It is likely that people in certain localities are more likely to protest, for historical reasons for example, and historical legacies may have an impact on political preferences or the current socioeconomic context. On the other hand, the breakdown by employment zone does

[50] We could also have run the regressions directly on our indicators of mobilization relative to the number of inhabitants of each area. However, given that our object of study is the territory, we consider that it is more direct to use a spatial intensity measure of mobilization and to use population density as a control, as this also has its own explanatory power, highlighting differences between rural areas and large cities, for example.

not allow for such an exhaustive measurement of the intensity of the movement on Facebook, but the higher number of observations allows us to control for all the unobserved characteristics of the territory at a larger scale: as such, in the employment zone-level regressions, we included indicator variables corresponding to the 21 metropolitan regions that existed prior to the 2015 reform.[51] The results obtained through this fixed effects specification are interpreted with a view to their intraregional variation, which limits the risks of omitted variables. Table 4 thus presents our two preferred specifications: at the departmental level without regional controls, and at the employment zone level with regional controls.[52]

Analysis of the explanatory power of our model shows that we manage to explain about half of the spatial dispersion of our department-level measures of mobilization intensity (columns 1 and 2). The inclusion of regional fixed effects increases the $R^2$ to 69% for blockade density by employment zone (column 4). In contrast, online mobilization at the employment zone level is poorly explained, reflecting the measurement problems with this dimension, as discussed above. It is interesting to note that controlling for population density is, paradoxically, significant only for understanding the number of blockades. This result suggests that population density, beyond its mechanical contribution to the intensity of mobilization per km2, reveals fundamental differences regarding the type of territory (rural) that was most extremely mobilized through blockades.

With respect to our variables of interest, we observe first of all that taking into account the correlation between the different explanatory variables removes, for certain variables, any significant correlation with mobilization: this is the case for the abstention rate, the share of diesel vehicles, and the unemployment rate. In particular, the coefficients associated with the share of diesel vehicles are now much lower, indicating that the strong negative correlation, which we observed previously, and which was difficult to explain, reflected the correlation of this variable with other controls.

Of the remaining variables, two are positively correlated with online mobilization, but not with offline mobilization: wage inequality and commuting distance. The measured effect is large, since one additional standard deviation of these variables from their mean level is associated with an increase in protest density of one-eighth to one-quarter standard deviation. The contrast with the zero correlation of the unemployment rate is interesting: it may mean that the movement, at least on Facebook, is determined less by inequalities in access to employment than by inequalities among workers. This finding corroborates the conclusions of sociological studies on the “intermediate” character of the Yellow Vests, who are more often poorly paid workers than people in situations of extreme exclusion.

The positive coefficients associated with commuting distance confirm the importance of the issue of mobility for online mobilization. On the other hand, in our specifications, commuting distance is not related to the density of blockades. This result may be due to the fact that commuting time does not prevent mobilization on Facebook, but it may act as an obstacle to physical

51 We preferred to control for regional rather than departmental fixed effects because, as explained earlier, the geography of departments and employment zones does not consistently overlap. Conversely, only nine employment zones straddle two regions: Mont-de-Marsan, Alençon, Cosne-Clamecy, Mâcon, Nogent-le-Rotrou, La vallée de la Bresle, Roissy, Brive, and Avignon. In our regressions, these areas are duplicated so that they can be associated with their two regions, which increases the number of observations from 296 to 305.

52 By way of comparison, Table A1 in Appendix 3.D complements these results by including the measure of the number of posts as a variable of interest (columns 1 and 2) and by including our two preferred variables of interest in an employment zone specification without regional fixed effects (columns 3 and 4).

mobilization. Even more notably, the presence of a large quantity of secondary roads whose speed limit was reduced to 80 km/h is very strongly correlated with mobilization, both on Facebook and in blockades. This is the most salient result of our study. This seems to indicate that the places most affected by this reform represented a *latent potential for mobilization*, which, thanks to the emergence of a protest movement against the rise in fuel prices, could then be *actualized*.

## 3.6 Conclusion

By focusing on the beginning of the Yellow Vests mobilization (the blockades of November 17 and Facebook activity from October to mid-December), this chapter examines the factors that triggered it. To this end, we propose an original approach, adapted to the spatial dimension of the movement. We show that the territory of social mobilization, in the context of the Yellow Vests, expand beyond the urban core of the largest metropolis thanks to the support of online social media. In addition, our study shows that the reduction in the speed limit on secondary roads to 80 km/h played an important role in the emergence of the mobilization, which could explain why its motivation was poorly understood by a part of the population[53] and by political parties at the start. This results connects with chapter 2 that describes the making of the suburban metropolis and highlights the role of roads in the rise of large-scale commuting. It also show that public transports are complement to roads, rather than substitute, in shaping access to labor market opportunities within metropolis, such that suburb-to-suburb commuters are trapped in their dependence on car.

The results presented in this chapter are interesting in that they suggest a link between *mobilization* and *mobility*. The Yellow Vests movement calls into question our relationship to physical and digital territories. These synergies between new phenomena of online mobilization and more traditional protests involving the occupation of public space constitute a new framework for interpreting social movements in France. Unlike other recent mobilizations in France, such as *La Manif pour tous* or *Nuit debout*, the spatial dispersion of the action points of the Yellow Vests movement was remarkable, which gave it an unprecedented nationwide coverage from the first day of mobilization.

Our work nevertheless has several limitations. First, the territorial approach masks, by definition, certain dimensions of individual heterogeneity within the geographical areas considered. Second, the statistical results we discuss are not all valid if we consider indicators of mobilization intensity *per capita*, rather than *per square kilometer*.[54] Finally, the subsequent evolution of the movement seems to have gone far beyond the issue of mobility alone. Only a dynamic study of the mobilization would make it possible to explain the mutations that the movement underwent. The following chapter undertake such a study by gathering additional data and focusing the study on a finer geographic level, the municipalities. It also display a quantitative case study of the topics shared on the Yellow Vests' Facebook pages.

---

[53] See, for example, "SansMoiLe17," `https://www.youtube.com/watch?v=P1MuWx9FR_A`.

[54] We refer here to Appendix 3.D, which replicates Table 3.3 for mobilization variables related to the number of inhabitants in the area. Although we think that our approach in terms of territorial intensity is more relevant in this case, an intuition corroborated by the higher explanatory power of the associated regressions, tests on other mobilization variables will need to be conducted to determine whether these differences reflect a statistical artifact related to the small number of observations at our disposal. As shown in Figure 3.5, mobilization per capita is much less correlated with absolute mobilization than is mobilization per $km^2$.

# Bibliography


**Acemoglu, Daron, Tarek A Hassan, and Ahmed Tahoun**, "The power of the street: Evidence from egypte's arab spring," *Review of Financial Studies*, 2017, *31* (1), 1–42. ↑139

**Aidt, Toke S and Raphaël Franck**, "Democratization under the threat of revolution: Evidence from the Great Reform Act of 1832," *Econometrica*, 2015, *83* (2), 505–547. ↑141

**Algan, Yann, Elizabeth Beasley, Daniel Cohen, Martial Foucault, and M Péron**, "Qui sont les Gilets jaunes et leurs soutiens," Technical Report, CEPREMAP et CEVIPOF 2019. ↑140

**Barbera, Salvador and Matthew O Jackson**, "A model of protests, revolution, and information," *Matthew O., A Model of Protests, Revolution, and Information (March 1, 2018)*, 2018. ↑139

**Bastos, Marco T, Dan Mercea, and Arthur Charpentier**, "Tents, tweets, and events: The interplay between ongoing protests and social media," *Journal of Communication*, 2015, *65* (2), 320–350. ↑140

**Bennani, Hamza, Pauline GandrÃ©, and Benjamin Monnery**, "Les déterminants locaux de la participation numérique au Grand débat national : une analyse économétrique," *Revue Èconomique*, 2019. ↑140

**Caren, Neal and Sarah Gaby**, "Occupy online: Facebook and the spread of Occupy Wall Street," *Available at SSRN 1943168*, 2011. ↑140

**Delalande, Nicolas**, *Les batailles de l'impôt. Consentement et résistances de 1789 à nos jours*, Le Seuil, 2011. ↑139

_ , "Le retour des révoltes fiscales?," *Pouvoirs*, 2014, (4), 15–25. ↑139

**Edmond, Chris**, "Information manipulation, coordination, and regime change," *Review of Economic Studies*, 2013, *80* (4), 1422–1458. ↑139

**Enikolopov, Ruben, Alexey Makarin, and Maria Petrova**, "Social media and protest participation: Evidence from Russia," *Available at SSRN 2696236*, 2017. ↑140, ↑150

**Little, Andrew T**, "Communication technology and protest," *Journal of Politics*, 2016, *78* (1), 152–166. ↑139

**Madestam, Andreas, Daniel Shoag, Stan Veuger, and David Yanagizawa-Drott**, "Do political protests matter? evidence from the tea party movement," *Quarterly Journal of Economics*, 2013, *128* (4), 1633–1685. ↑141

**Neveu, Erik**, *Sociologie des mouvements sociaux*, La Découverte (Repères), 1996. ↑139

**Ostrom, Elinor**, *Governing the commons*, Cambridge University Press, 2015. ↑139

**Parkin, Frank**, *Middle Class Radicalism*, Melbourne University Press, 1968. ↑141

**Passarelli, Francesco and Guido Tabellini**, "Emotions and political unrest," *Journal of Political Economy*, 2017, *125* (3), 903–946. ↑141

**Porta, Donatella Della, Massimiliano Andretta, Angel Calle, Helene Combes, Nina Eggert, Marco G Giugni, Jennifer Hadden, Manuel Jimenez, and Raffaele Marchetti**, *Global justice movement: Cross-national and transnational perspectives*, Routledge, 2015. ↑141

**Qin, Bei, David Strömberg, and Yanhui Wu**, "Why does China allow freer social media? Protests versus surveillance and propaganda," *Journal of Economic Perspectives*, 2017, *31* (1), 117–40. ↑140

**Russell, Hardin**, "Collective action," *RFF press*, 1982, *150*, 2000. ↑139

**Sandler, Todd and Lawrence E Blume**, *Collective action: Theory and applications*, Vol. 4, University of Michigan Press Ann Arbor, 1992. ↑139

**Sebbah, Brigitte, Natacha Souillard, Laurent Thiong-Kay, and Nikos Smyrnaios**, "Les Gilets Jaunes, des cadrages médiatiques aux paroles citoyennes," *Rapport de recherche préliminaire*, 2018. ↑140

**Steinert-Threlkeld, Zachary C, Delia Mocanu, Alessandro Vespignani, and James Fowler**, "Online social networks and offline protest," *EPJ Data Science*, 2015, *4* (1), 19. ↑139

**Stephens-Davidowitz, Seth**, "The cost of racial animus on a black candidate: Evidence using Google search data," *Journal of Public Economics*, 2014, *118*, 26–40. ↑146

— **and Hal Varian**, "A hands-on guide to Google data," *Mountain View, CA*, 2014. ↑146

**Touraine, Alain**, *Le mouvement de mai ou Le communisme utopique*, Edition du Seuil, 1968. ↑141

— , *La fin des sociétés*, Edition du Seuil, 2013. ↑141

**Vasi, Ion Bogdan and Chan S Suh**, "Protest in the internet age: Public attention, social media, and the spread of Occupy protests in the United States," in "Politics and Protest workshop," Vol. 13 2013. ↑140

**Wieviorka, Michel**, *Neuf lecons de sociologie*, Robert Laffont et EHESS, 2008. ↑141

## 3.A Territories and Geographical Division

**For a geographical approach.** The aggregation of socioeconomic and political data at the geographical level has many advantages. It makes it possible to take into account interactions between agents that occur at the territorial level (local job market, exchange of information), the effects of localized public policies (local civil service, reduction in the speed limit from 90 km/h to 80 km/h), and certain modalities of action (blockades of roundabouts). Moreover, it makes it possible to merge information from different sources, for example to merge average income with voting behavior, which may not be possible at the individual level (unless a survey is conducted on a sample of the population).

**The modifiable areal unit problem.** Conducting studies with a spatial approach raises the question of the choice of the geographical unit studied. Indeed, aggregation at the spatial level carries the risk of producing results that are oriented by the choice of nomenclature, such as scale or zoning effects. This is the modifiable areal unit problem (MAUP), as documented by Openshaw and Taylor (1979). Highlighting the sensitivity of the study to the choice of spatial division is not strictly speaking a problem, but corresponds to the multi-scalar aspect of the object of study. The difference in results obtained from one nomenclature to another must therefore be seen as a contribution in itself to the knowledge of the phenomenon considered.

**Commune.** The smallest administrative unit available is the commune. There are about 35,000 communes in France, but the territories they cover are too small for our analysis. Indeed, the Facebook groups and roundabout blockades identified themselves with larger spatial units, such as the department (for example, “Gilets Jaunes Gironde,” “Union Gilets Jaunes 84”) or the “ pays” (“Gilets Jaunes du Pays d’Auray,” “Gilets Jaunes Dinan et environs”).

**Department.** The first spatial unit chosen in our study corresponds to the administrative division of the department, established by a December 22 decree of the Constituent Assembly of 1789. Because of the age of this division and its discretionary nature, we consider this division to be “exogenous,” that is, we assume that its definition is external to the socioeconomic context discussed in this article. Moreover, the departments are fairly homogeneous and comparable in terms of size and organization. In general, they are composed of a central commune (the prefecture or departmental capital) and more or less populated departmental sub-hubs. This allows for a relatively consistent division of the territory. However, this disconnection of contemporary economic dynamics from the departmental division also presents disadvantages for a detailed study of the impact of localized public policies or social movements. For example, a department groups together heterogeneous types of housing and activities, and potentially competing hubs or sub-hubs, which obscures the variations.

**Employment zone.** A good “endogenous” definition of territory, based on an economic reality, is that provided by the employment zone, as defined by INSEE. It defines a geographical area within which most of the active population resides and works (at least 40offered (Jayet 1985). The division of the territory into employment zones is common in labor market studies (Malgouyres

2017). Grouping together the broad living areas where individuals work, reside, and consume, the employment zone also defines relevant territories for local diagnoses. In particular, it aims to guide the delimitation of territories for the implementation of territorial policies. Here we use the updated 2010 division. It should be noted, however, that an analysis based on this spatial unit can obscure the disparities that exist within zones that are polarized between the center and its periphery, the inner and outer suburbs. In addition, since they cover the entire country, employment zones are sometimes more artificial in sparsely populated geographical areas without urban centers (such as the employment zones of Morvan, Le Blanc, or L'Aigle).

**Constituency.** An intermediate territorial division between the department and the commune, which might seem relevant for our study, is that of the constituency orcircon- scription (the current division dates from 2011). However, it is often difficult to retrieve data at this scale, as it is disconnected from communal delimitations. Some communes are divided up by different constituencies, which makes it impossible to match data. It is therefore not possible for us to associate Facebook groups with a single constituency and thus build a satisfactory database.

In addition, most constituencies are constructed in such a way as to divide the terri - tory in a homogeneous manner: starting from a portion of the central city, they extend into the peri-urban and rural areas of the department, without following any economic or administrative logic.

## 3.B Geolocation of Facebook Groups

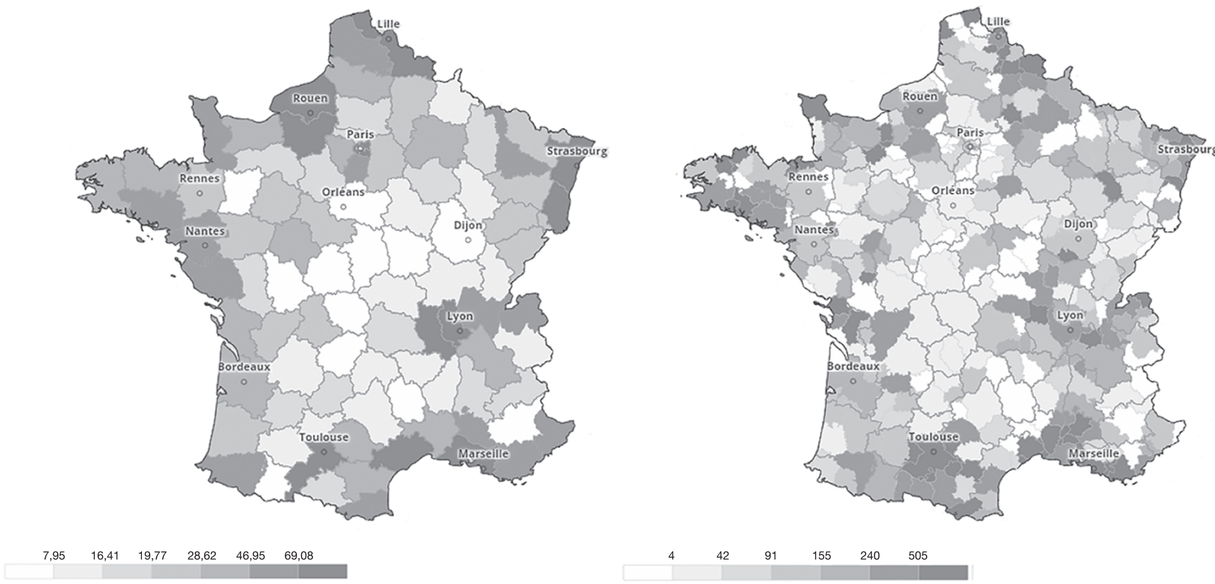
We performed a task of character matching, starting from the contents of two lists: one containing the names of all French cities and another containing all the names of the Facebook groups studied. We proceeded in three stages:

1. *Cleaning the text:* we removed accents, punctuation, and single-letter words, and transformed the whole text into lower case. We replaced all occurrences of "ste" with "sainte" and "st" with "saint."
   *Example 1:* " Les Gaulois de Calais ! MOBILISATION contre les taxes du gouvt Macron" became "les gaulois de calais mobilisation contre les taxes du gouvt macron."
2. *Matching:* for each Facebook group, we checked which cities appeared in the name.
   *Example 1:* "calais" appears in the name of the Facebook group. No other cities appear.
3. *Choosing the best candidate:* if several cities appear in the name of the Facebook group, we took the longest city name as the city with which the group is associated. This simple rule drastically reduced the number of false positives (cities associated with a group when the group does not correspond to that city).
   *Example 1:* We associated the Facebook group with the city of Calais.

## 3.C Other Mobilization Maps

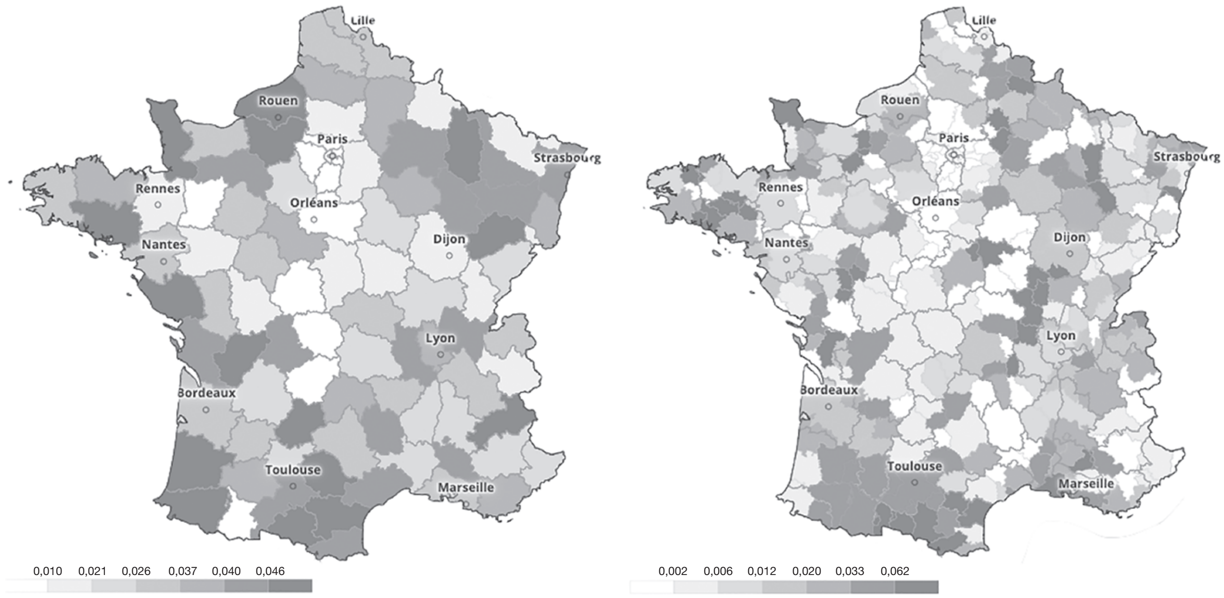
Here we present four additional series of maps, representing online mobilization per km2 (Figure A1) and per capita (Figure A2), as well as the equivalent maps for offline mobilization (Figures A3 and A4). The maps in Figures A1 and A3 represent the measures used in the empirical analysis. They closely resemble the population density maps, which justifies our use of this measure as a control in our regressions.

**Figure 3.8:** Online mobilization (number of members of Facebook groups) per km$^2$: Departments and employment zones.



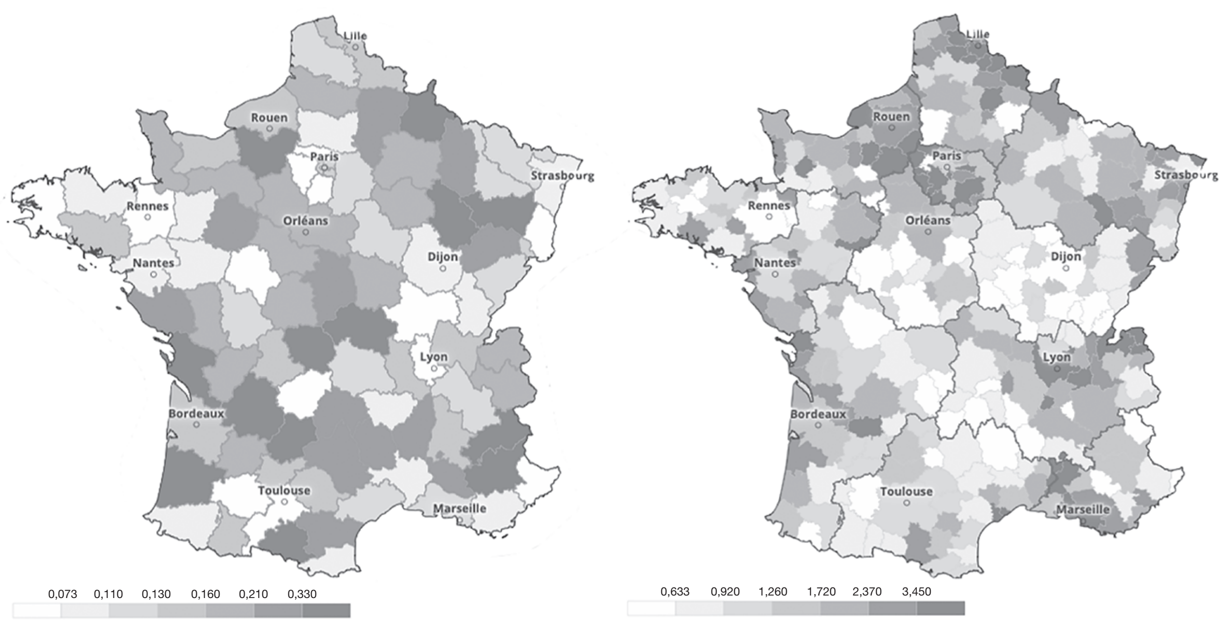

*Note:* These two maps represent the number of members of Facebook groups per km 2 localizable at the level of departments (left) and employment zones (right). The recording of this variable is more reliable at the departmental level and shows a strong correlation with population density. The employment zone level highlights small localities, where individuals identify more with their commune. *Sources:* Data collected manually on *Gilets jaunes* Facebook groups, between December 12 and 15, 2018. Map projection made using the online tool: `https://statistiques-locales.insee.fr/`.

**Figure 3.9:** Online mobilization (number of members of Facebook groups) per 10,000 inhabitants: Departments and employment zones.



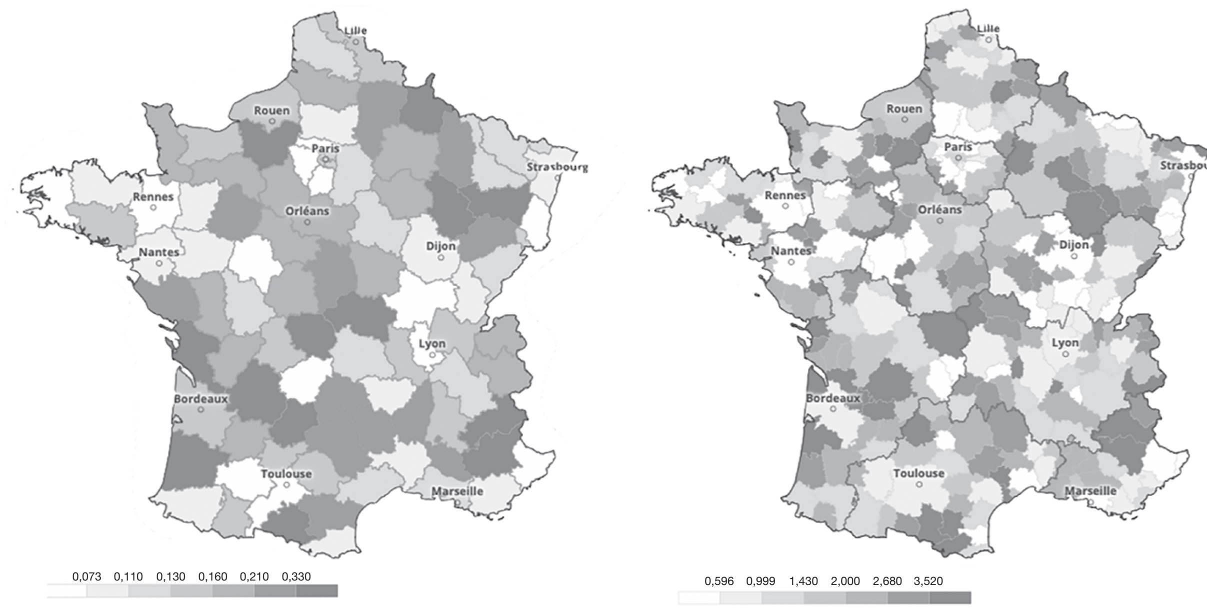

*Note:* These two maps represent the number of members of Facebook groups per 10,000 inhabitants localizable at the level of departments (left) and employment zones (right). The recording of this variable is more reliable at the departmental level and shows a strong correlation with population density. The employment zone level highlights small localities, where individuals identify more with their commune. *Sources:* Data collected manually on *Gilets jaunes* Facebook groups, between December 12 and 15, 2018. Map projection made using the online tool: `https://statistiques-locales.insee.fr/`.

**Figure 3.10:** Offline mobilization (number of roundabout blockades) per km$^2$: Departments and employment zones.




*Note:* These two maps represent the density of blockades per km2 at the level of departments (left) and employment zones (right). We can see that there is a high degree of heterogeneity and that the comparison of these different types of division reveals heterogeneity within departments, thanks to the employment zone nomenclature. *Sources:* Data collected on the evening of November 16 at `www.blocage17novembre.fr`. Map projection made using the online tool: `https://statistiques-locales.insee.fr/`.

**Figure 3.11:** Offline mobilization (number of roundabout blockades) per 10,000 inhabitants: Departments and employment zones.




*Note:* These two maps represent the density of blockades per 10,000 inhabitants at the level of departments (left) and employment zones (right). We can see that there is a high degree of heterogeneity and that the comparison of these different types of division reveals heterogeneity within departments, thanks to the employment zone nomenclature. *Sources:* Data collected on the evening of November 16 at `www.blocage17novembre.fr`. Map projection made using the online tool: `https://statistiques-locales.insee.fr/`.

## 3.D Additional Results

**Table 3.5:** Additional econometric results, variables of interest per km$^2$

| | | Department | | Employment zone | |
|---|---|---|---|---|---|
| | | Members | Blockades | Members | Blockades |
| Vote | Macron | – 0.0196 | – 0.316*** | – 0.0872 | – 0.209*** |
| | | (0.163) | (0.0908) | (0.0754) | (0.0453) |
| | Abstention | – 0.0962 | 0.0998 | 0.0295 | – 0.0369 |
| | | (0.141) | (0.0854) | (0.0754) | (0.0444) |
| Political decisions | Speed limit reduction | – 0.0449 | 0.268*** | 0.136* | 0.127*** |
| | | (0.147) | (0.0839) | (0.0722) | (0.0425) |
| | Share of diesel vehicles | 0.0257 | 0.0263 | 0.0859 | 0.0509 |
| | | (0.149) | (0.0861) | (0.0827) | (0.0487) |
| Economy and society | Inequality | 0.0941 | 0.194** | 0.112 | 0.0289 |
| | | (0.171) | (0.0869) | (0.0819) | (0.0482) |
| | Unemployment | 0.196 | – 0.0691 | 0.0600 | – 0.00999 |
| | | (0.148) | (0.0893) | (0.0770) | (0.0453) |
| Geography | Distance | 0.194 | 0.108* | 0.166** | 0.0341 |
| | | (0.121) | (0.0652) | (0.0665) | (0.0392) |
| Age and density | | Yes | Yes | Yes | Yes |
| Regional fixed effects | | No | Yes | No | No |
| N | | 89 | 305 | 296 | 296 |
| Adjusted R2 | | 0.151 | 0.158 | 0.059 | 0.674 |

*Note:* 1) “Members” and “Blockades” refer to the number of members of Facebook groups and the number of blockades planned for November 17, respectively, per km$^2$; 2) All variables are centered and reduced; 3) Standard deviations are in parentheses; $^{***}$, $^{**}$, and $^{*}$ denote the significance of results at 99.9%, 99%, and 95%, respectively.

**Table 3.6:** Additional econometric results, variables of interest per km$^2$

| | Department | | Employment zone | |
|---|---|---|---|---|
| | Members/capita | Blockades/capita | Members/capita | Blockades/capita |
| Macron | – 0.2838*<br>(0.1612) | – 0.1436<br>(0.1478) | – 0.1099<br>(0.0977) | – 0.3638***<br>(0.0901) |
| Abstention | – 0.3024**<br>(0.1406) | 0.1018<br>(0.1288) | – 0.0077<br>(0.0919) | – 0.0860<br>(0.0847) |
| Speed limit reduction | 0.1867<br>(0.1405) | – 0.0717<br>(0.1288) | 0.0936<br>(0.0902) | – 0.0872<br>(0.0832) |
| Share of diesel vehicles | 0.1474<br>(0.1463) | 0.2374*<br>(0.1341) | 0.0665<br>(0.0926) | 0.1684**<br>(0.0854) |
| Inequality | 0.1719<br>(0.1687) | – 0.0478<br>(0.1546) | 0.0271<br>(0.0935) | 0.0370<br>(0.0862) |
| Unemployment | – 0.0192<br>(0.1469) | 0.0098<br>(0.1347) | – 0.0714<br>(0.0961) | – 0.1707*<br>(0.0886) |
| Distance | 0.2018*<br>(0.1194) | – 0.0878<br>(0.1094) | 0.1884***<br>(0.0702) | – 0.0069<br>(0.0647) |
| Age | – 0.0321<br>(0.1342) | 0.3326***<br>(0.1231) | 0.1108<br>(0.0798) | 0.2742***<br>(0.0736) |
| Density | – 0.2019<br>(0.1648) | – 0.0490<br>(0.1210) | 0.0204<br>(0.0856) | 0.0438<br>(0.0789) |
| Fixed effects | No | No | Yes | Yes |
| N | 89 | 89 | 305 | 305 |
| Adjusted R2 | 0.130 | 0.269 | 0.015 | 0.173 |

*Note:* 1) "Members" and "Blockades" refer to the number of members of Facebook groups and the number of blockades planned for November 17, respectively, per km$^2$; 2) All variables are centered and reduced; 3) Standard deviations are in parentheses; ***, **, and * denote the significance of results at 99.9%, 99%, and 95%, respectively.

# Bibliography


**Jayet, Hubert**, "Les zones d'emploi et l'analyse locale des marchés du travail," 1985. ↑162

**Malgouyres, Clément**, "The impact of Chinese import competition on the local structure of employment and wages: Evidence from France," *Journal of Regional Science*, 2017, *57* (3), 411–441. ↑162

**Openshaw, Stan and Peter J. Taylor**, "A Million or so Correlation Coefficients: Three Experiments on the Modifiable Areal Unit Problem," *Statistical Applications in the Spatial Sciences*, 1979, p. 127–44. ↑162

# CHAPTER 4

## Mobilization without Consolidation

### *Social Media and the Gilets jaunes*

with Pierre BOYER, Germain GAUTHIER, Vincent ROLLET, and Benoit SCHMUTZ

**Abstract**

How do social media affect the way protest movements unfold? We study this question using the French Yellow Vests movement (i.e., les *Gilets jaunes*) as a case study. The Yellow Vests came into the spotlight on November 2018, with hundreds of road blockades across the country. We bring together data covering the different form of actions over the period from October 2018 to April 2019 on road blockades, petition signatories, and activity on Facebook pages and groups. We show how the road blockades that started the Yellow Vests movement were planned online and later reinforced online activism. However, online discussions progressively shifted away from practical demands and organizational concerns towards more violent and critical content. This evolution was almost equally driven by selective attrition of less antagonistic discussants and individual radicalization over time.

**Keywords**: Yellow Vests; Protests; Social Media; Online Mobilization.

**JEL Classification**: H54, N94, R12, R41.

Circulate under the following reference: Boyer, Pierre, Delemotte, Thomas, Gauthier, Germain, Rollet, Vincent, and Schmutz, Benoît. "*The Gilets Jaunes: Offline and Online*". CEPR Discussion Paper #14780 (2020).

For comments that have improved this article, we thank Micael Castanheira, Luisa Dörr, Allan Drazen, Elie Gerschel, Julien Grenet, Sophie Hatte, Fanny Henriet, Clément Imbert, Francis Kramarz, James Lo, Nolwenn Loisel, Andrew Lonsdale, Phoebe Mac Donald, Clément Malgouyres, Isabelle Méjean, George Melios, Pierre-Guillaume Méon, Vincent Pons, Audrey Rain, Anasuya Raj, Manon Revel, Ilan Tojerow, Clemence Tricaud, Thomas Van Casteren, Oliver van den Eynde, Lionel Wilner, and Ekaterina Zhuravskaya, as well as seminar and conference participants at CREST, "Dulbea Workshop on the Political Economy of Mass Demonstrations" at ULB Brussels, Boston University, PSE, Sciences Po, Harvard (Government), Paris 8 and IIPF 2020 for their comments.

The authors gratefully acknowledge the Investissements d'Avenir (ANR-11-IDEX-0003/Labex Ecodec/ANR-11-LABX-0047), ANR-19-CE41-0011-01 and ANR-20-CE41-0013-01 for financial support, the CASD (Centre d'Accès Sécurisé aux Données) and INSEE for the access to French administrative data, and Change.org for sharing their anonymized data.

The results presented in this chapter are the sole responsibility of the authors.

# Introduction

Internet has profoundly changed the way social movements emerge, spread and are organized (Loader, 2008; Earl and Kimport, 2011; Bennett and Segerberg, 2012; Castells, 2015). Social media, in particular, are often cited as essential signaling and coordination devices for demonstrations. As a result, over the past decade, the most salient protest movements worldwide have often been a combination of offline and online mobilization. Twitter, Facebook, and other platforms have been instrumental to the development of the Arab Spring, the Hirak movement in Algeria, and the 17 October Revolution in Lebanon, among others. However the mechanisms between online and offline actions remain unknown.

The irruption of social media in social unrest movements disrupt the way they unfold. On the one hand, new means of information sharing and coordination allows for more accessible movement, where anyone can initiate or rally a protest. On the other hand, a strong dependency on a leaderless social media infrastructure, where anyone can voice their opinion, may dampen the ability to structure long lasting and effective political campaign.

In this paper, we extend the study done in the previous chapter, with a focus on the dynamics of this hybrid social movement. We complete the previous data with additional sources: Facebook pages and Change.org. This allows us to explore further the interactions between the online and offline components of the movement, and to study how the movement grew and declined. This case study is of particular interest to described the role of social media in social movement dynamics. In addition, as described in the previous chapter, it went through the different stages of the social movement theorized by Blumer (1969) and Tilly (1977): emergence, coalescence, and decline,[1] for which of them we have been able to collect data.

**Emergence** The wide sharing of the Change.org petition against the rise in gasoline prices signaled a strong opposition to government policies throughout France, providing a social ferment for the Yellow Vests movement to launch itself. As the petition became viral, the movement started to coalesce and a large social media infrastructure was put in place to organize road blockades. We provide correlational evidence on the role of social media in the organization of these demonstrations, based on spatial regressions at the municipal level. Our analysis shows the extent to which a close monitoring of online activity may help anticipate where protests will occur. In particular, we show that a few variables describing the online activity of the Yellow Vests predict the occurrence of a local blockade as precisely as a wide set of geographic, demographic, economic, and political controls. This evidence supports previous work[2] showing that online mobilization facilitates the emergence of offline protests.

**Coalescence** Our analysis further explores the reverse direction of this link, and shows that offline demonstrations spur further online activity. Daily time series of Facebook group creations

[1] The traditional description of social movements includes a fourth stage between coalescence and decline: bureaucratization. In this phase, formal organizations are structured to support the movement – for instance, the Gay and Lesbian Alliance Against Discrimination (GLAAD) emerged in the bureaucratization phase of the LGBT rights movement. In the case of the Yellow Vest movement, such organizations failed to emerge, in large part because many Yellow Vests wished that their movement remain a grassroots movement.

[2] See Section 4.

show a massive peak of group creations right after the 11/17 blockades. On top of this descriptive fact, we provide causal evidence of the impact of the first wave of blockades on subsequent online mobilization, with an instrumental variable strategy based on the spatial dispersion of roundabouts, a peculiar feature of the French urban landscape that plausibly reflects idiosyncratic local urban planning preferences. Roundabouts were heavily targeted by protesters because they allow demonstrators to block several roads at a time and are easy to set camp on. We show that the presence of local blockade triggers the creation of 0.7 additional local Facebook groups (on average and conditional on a broad range of confounding factors). This finding substantiates qualitative evidence that many protesters sought to continue to exchange with fellow protesters whom they had met on the blockades. This considerably enriched the Yellow Vests' online infrastructure, with a tripling of the number of dedicated groups, which could have paved the way to a subsequent consolidation of the movement.

**Decline** In the months following the 11/17 blockades, the Yellow Vests organized weekly demonstrations, but the number of protesters quickly subsided. To understand the evolution of the movement during this decline phase, we extract a corpus of 2.8 million sentences posted by Yellow Vests on a large number of public Facebook pages. This allows us to build a quantitative case study based on different Natural Language Processing (NLP) methods. We train a topic model and show that the share of messages related to organizational concerns and practical demands decreased over time, contrary to messages with more antagonistic content such as insults or mentions of violence. Similarly, the share of messages with a negative sentiment or which are predicted to be written by affiliates of extreme parties increased over time. A focus on a subset of our text corpus where we can follow (de-identified) individual users over time allows us to distinguish between an extensive margin (changes in the composition of the population of discussants) and an intensive margin (an individual-level increased tendency to post antagonistic messages) of radicalization. With a simple linear decomposition, we show that both margins play an important role. These findings suggest that online discussions progressively departed from the movement's initial goals, as protests were subsiding in the streets and the various Yellow Vest leaders could not agree on a consensus regarding future political strategy.

## Related literature

The importance of social media in the development of large protest movements has been investigated in theoretical (e.g., Little, 2016; Barbera and Jackson, 2020) and empirical studies (e.g., Rane and Salem, 2012; Clarke and Kocak, 2020; Gaffney, 2010; Bakker and De Vreese, 2011; Borge-Holthoefer et al., 2011; González-Bailón et al., 2011; Bastos et al., 2015; Acemoglu et al., 2018; Jost et al., 2018; Qin et al., 2017; Bursztyn et al., 2021; Enikolopov et al., 2020).

Our contribution with respect to this literature is twofold. First, we complement previous studies which showed that social media contributes to the emergence of offline protests by showing that offline demonstrations lead to further online action. Together with the previous literature, our study shows how social media can play as a force keeping protest movements alive. Second, we are able to analyze the radicalization which accompanies the decline phase of many social movements (see, e.g., Christiansen, 2009). Using a panel of online protesters, we are able to disentangle which part of the radicalization comes from a composition bias (less radicalized contributors dropping out) or an individual-level increased tendency to post antagonistic messages.

On a more distant note, we relate to the literature that studies the impact of social media and the Internet on trust in government – e.g. Guriev et al. (2019) and Algan et al. (2019) – turnout – e.g. Falck et al. (2014), Campante et al. (2018), Gavazza et al. (2019) and polarization of the electorate – see, e.g., Halberstam and Knight (2016), Boxell et al. (2017), Barrera et al. (2020), Henry et al. (2020), Manacorda and Tesei (2020), Pan and Siegel (2020).[3]

Turning to the literature studying the Yellow Vests specifically, this paper is the first one to bring together comprehensive data on signatories of the most important petition related to the movement, the dynamics on the movement on Facebook pages and groups, and administrative data at the local level. A recent review of the literature is provided in Bendali and Rubert (2020). Previous chapter investigates the determinants of the origins of Gilets jaunes, focusing on the first months of the movement. Algan et al. (2019), Algan et al. (2020) and Davoine et al. (2020) enriched the analysis of the movement by correlating Yellow Vest activity with well-being and social capital indicators. Using textual analysis methods, Sebbah et al. (2018) and Cointet et al. (2021) are early attempts to identify the topics discussed by the Yellow Vests.

The remaining of the paper proceeds as follows. In Section 4.1, we describe the context of our study and the data sources that we use to analyze the Yellow Vest movement. We then describe the two contributions of the paper, each corresponding to a separate phase of the movement. In Section 4.2, we explore the interaction between online and offline mobilization as the movement coalesced. In Section 4.3, we then turn to our quantitative case study of the decline of the Yellow Vest movement. Section 4.4 concludes.

## 4.1 Context and data

### 4.1.1 Brief history

The success of the Yellow Vests and the 11/17 blockades results from the combination of chance and the social media ecosystem. The previous chapter provide a full timeline of the events, but several milestones stand out that we recall in this section.

A social movement against increases in gasoline prices started gaining traction when a petition against high gas prices ("Pour une Baisse des Prix du Carburant à la Pompe !") was picked up by a local journalist on October 12, 2018. The petition was initially created in May 2018 by a local motorist (Priscilla Ludosky) on the Change.org platform and had garnered fewer than 1,000 signatures by the time of the article. The day following the article's publication, the number of signatories in tripled in Ms. Ludovsky's county (Seine-et-Marne). The article caught the eye of the wife of a truck driver who had been planning a blockade on Facebook with fellow angry car users. She linked the petition on Facebook, which extended its reach. Nine days and thousands of local signatures later, a national newspaper published the two stories and signatures skyrocketed all over the country. On October 24, the yellow road security jacket, which every car owner is compelled by law to have in her trunk, was

[3] See Zhuravskaya et al. (2020), Allcott and Gentzkow (2017) and Guriev and Papaioannou (2020) for reviews.

proposed as a rallying sign for angry motorists. The first organizers of the movement heavily relied on Facebook to diffuse information, and as 11/17 was approaching, several dedicated websites[4] were created to provide a list of local relevant Facebook groups. On 11/17, hundreds of blockades took place across the country, organized by hundreds of thousands of protesters.

As many of these blockades were quickly evacuated, the movement then resorted to more conventional weekly demonstrations in the main cities. A climax of violence was reached on December 1 in Paris. The following Saturday, police tanks were mobilized and 2,000 people were arrested. On December 10, President Macron presented a 10-billion-euro plan that significantly bent the government's budgetary policy, and asked for a compilation of lists of grievances (*Cahiers de doléances*, as took place during the French Revolution in 1789) from across the country, to be followed by hundreds of self-organized town halls meant to give anyone the opportunity to voice their concerns through a "Great National Debate" (*Grand Débat National*).

This response was not inconsequential, but it did not tip the balance of power toward peripheral cities and rural areas in the way that the Yellow Vests were hoping for. Some blockades turned into permanent campsites and weekly demonstrations continued for months, but the number of demonstrators soon became negligible (except in Paris where some large scale demonstrations still took place until March 2019). At the same time, polls indicate that public support decreased in the French population, and the Yellow Vests ultimately failed to present a united front for the upcoming elections. The movement was still active online in Spring 2022, with sporadic protests (e.g., against sanitary lockdowns) where yellow vests were worn as a badge of honor. As such, this simple piece of garment became a persistent and divisive icon in the French political landscape.

### 4.1.2 Measuring the intensity of mobilization

**Petition signatories signal a willingness to protest.** To understand the launch of the movement, we collected anonymized data on petition signatories from Change.org. The data includes city of residence and associated ZIP code. As of October 16, 2019, the petition garnered 1,247,816 signatures in total, including 1,043,337 with a valid ZIP code. As shown in Appendix Figure 4.8, the petition quickly spread throughout France during the month preceding the 11/17 blockades.

**Self-declared blockades proxy the offline mobilization.** Following a call for a national blockade of roads and the organization of several local demonstrations, a website was created to coordinate the actions planned for 11/17. It provided a map of the organized blockades, updated in real time, which we collected on the evening of November 16. This map documented 788 announced blockades in metropolitan France, which all pointed to precise road infrastructures and included specific descriptions of the planned events.[5]

Many places were chosen for their potential to block traffic and economic activity. To analyze the extent to which France was affected by the blockades, we use the partition of the country in "Living Zones" (hereafter, LZ). They are administrative units defined as the smallest territorial units

[4] `blocage17novembre.fr`, then `gilets-jaunes.com`, and `giletsjaunes-coordination.fr`.

[5] Note that these are declarations of an intent to demonstrate. Yet, as the map was created to coordinate the blockades, there was little incentive to falsely declare an intent to demonstrate.

in which residents have access to basic infrastructure and services and conduct a large part of their daily lives. We observe that 551 out of the 1,632 LZs in France experienced a blockade. They correspond to half of the country's population and to a sizable fraction of the French territory as shown with the geography of the movement in chapter 3.

**Facebook groups and pages proxy the online mobilization.** To document online mobilization, we searched for all public Facebook groups related to the movement. Using the methodology of Caren and Gaby (2011), we compiled a list of the Facebook groups that were still active one month after 11/17 by performing search requests using a large set of keywords linked to the movement. For each group, we recorded its name, creation date, number of members, and number of publications. As shown in Table 4.1, we identified 3,033 groups, with over four million members in total. Over two-thirds of the groups were associated with a geographical area and more than 40% of the total number of members belonged to these localized groups. Moreover, only 20% of the posts emanated from national groups, which suggests that localized groups were the most active ones.

**Table 4.1:** Characteristics of Facebook groups

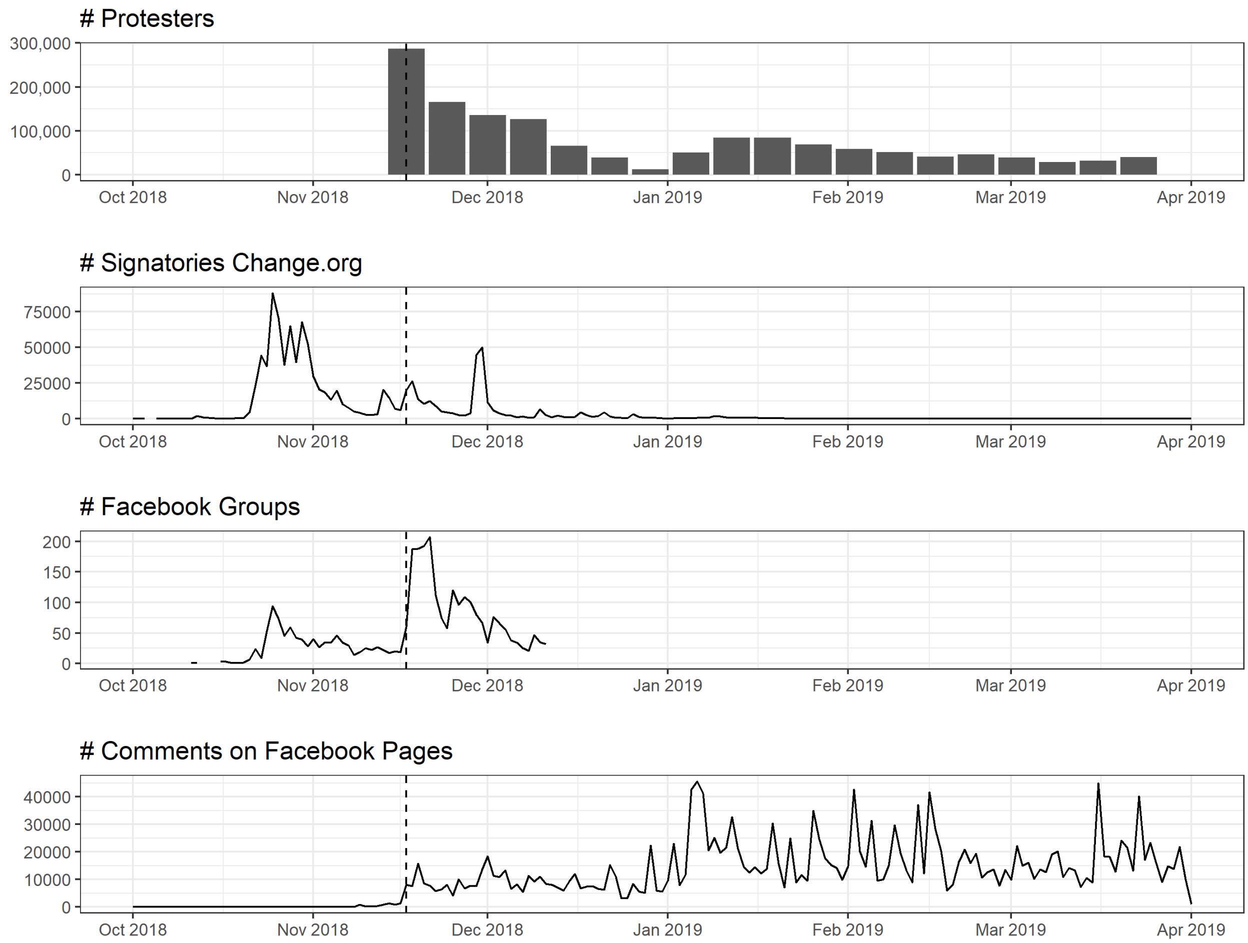
| Targeted Audience | Groups | Members | Publications |
|---|---|---|---|
| **National** | 520 (64%) | 2,381,562 | 264,034 |
| **Regional** | 171 (81%) | 249,516 | 138,367 |
| **Département** | 724 (81%) | 528,500 | 336,437 |
| **Municipal** | 1,608 (65%) | 949,342 | 714,882 |
| **Total** | 3,033 (70%) | 4,109,325 | 1,453,878 |

Notes: In the first column of this table, we show the number of Facebook groups for each geographic focus. In parentheses, we indicate the share of the number of groups created after 11/17. Other columns show total number of members and total number of publications (this number is right-censored by Facebook at 10,000 publications per group). The last line ("Total") includes 10 "foreign" groups, 9 of which created after 11/17, including 405 members and associated with 158 publications.

Using a similar method, We also identified 617 Facebook pages and used Netvizz as well as a custom webscrapper to retrieve their content (Rieder, 2013): posts, comments and interactions (such as likes and shares). This corpus features over 121,000 posts, 2.1 million comments, 2.8 million sentences and 21 million interactions.

**Stylized facts.** The number of petition signatories, the number of Facebook group creations, and the number of comments on Facebook pages are depicted as daily time series in Figure 4.1. While the petition was mostly signed before 11/17, there were two distinct episodes of group creation: a small one in the weeks prior to the blockades and a large one shortly afterward. This pattern suggests that some Facebook groups were used as a means to organize the blockades, but many of them served as virtual meeting places that allowed the movement to carry on after an initial mobilization in the streets. This hypothesis is corroborated by the evolution of the intensity of the discussions that took place on dedicated Facebook pages. Discussions gained in importance in January 2019 and, contrary to the weekly number of protesters, remained strongly active during the following months.

**Figure 4.1:** Evolution of online and offline mobilization




Notes: In the top panel, we show the number of demonstrators reported each week by the Ministry of the Interior. In the second panel, we plot the daily number of petition signatures. In the third panel, we plot the daily number of new Facebook groups created. Finally, in the bottom panel, we plot the daily number of comments posted on Facebook pages. The vertical dashed line in all panels corresponds to 11/17.

### 4.1.3 Extracting meaningful insights from Facebook discussions

To analyze discussions on Facebook pages, we rely on various NLP methods (see Grimmer and Stewart, 2013 and Gentzkow et al., 2019 for an overview).

**Topic model.** First, to identify the topics discussed online by the Yellow Vests, we rely on a topic model tailored to the analysis of short text snippets (Demszky et al., 2019).[6] In our main specification,

[6] Standard models, based on Latent Dirichlet Allocation (LDA), allow texts to span over multiple topics. Though this is often a reasonable assumption, it is implausible in the case of short text snippets (such as sentences) which often refer to only one topic (Yan et al., 2013). Our algorithm proceeds in several steps. First, we produce word embeddings for the corpus and represent each sentence as a vector in the embedding space. We train a Word2Vec model using Gensim's implementation, with moving windows of eight tokens and ten iterations of training. Then, we build sentence embeddings as the weighted average of the constituent word vectors, where the weights are smoothed inverse term frequencies (to assign higher weights to rare/distinctive words) (Arora et al., 2017). The resulting embedding space allows for a low-dimensional representation of text, in which phrases which appear in similar contexts are located close to one another. After this transformation, we group sentence vectors together into a small set of clusters using the K-

we allow for 15 different topics, but qualitatively similar results are found with alternative numbers of clusters. We display the topics that we obtain in Figure 4.2.[7] These topics may be grouped into different categories, such as protest organization (a and b), socialization (c and g), and online mobilization (d). Other topics reflect both the reasons behind the protests and the political goals the Yellow Vests were trying to achieve (e and f). Finally, five topics refer to antagonistic messages (h, i, j, k, and l) and reflect the protesters' anger towards government officials and recent policies.[8]

**Sentiment analysis.** To measure emotional content in messages, we use a dictionary-based approach that assigns to each sentence a sentiment score ranging from -1 (very negative) to 1 (very positive).[9] Figure 4.2 distinguishes the 15 topics depending on whether the average sentiment of messages in the topic is above or below zero. The five topics which we classify as antagonistic are associated with negative sentiment.

**Political classification.** Finally, to understand the political stance of messages, we build a measure of partisanship using a supervised learning model based on tweets from parliamentarians (Peterson and Spirling, 2018; Widmer et al., 2020).[10] The political leanings that we consider are those of the five largest parties in France at the time of the events: the Rassemblement National (RN), Les Républicains (LR), La République en Marche (LREM), the Parti Socialiste (PS), and La France Insoumise (LFI).[11]

Figure 4.3 plots the predicted distribution of political partisanship for users in the Facebook corpus. We find that the political preferences of online activists differed from the political preferences in the general French population (as measured by votes in the first round of the 2017 presidential elections), with an over-representation of the far-left (LFI).

Armed with these different sources of data, we now turn to the description of the interaction between online and offline mobilization of the Yellow Vests.

Means algorithm. Each cluster that we obtain is interpreted as a topic. We train the algorithm on 100,000 randomly drawn sentences and predict clusters for the rest of the corpus. Finally, we use the ten most salient words for each cluster to manually label topics.

7 Examples of sentences associated to each topic are provided in the appendix.

8 The remaining three topics are more difficult to interpret.

9 For each sentence, the sentiment score is obtained as the average of the sentiment scores of its constituent words. We rely on the VADER (Valence Aware Dictionary for Sentiment Reasoning) library for our main results. In the Appendix, we show that our results are robust to two alternative measures of sentiment.

10 We train our model using a penalized multinomial logistic regression on the tweets of politicians in the French Parliament. The prediction is solely driven by textual similarity between the words used in the snippet and the words used by politicians in each party. (see, e.g., Friedman et al., 2001). Given the large size of the vocabulary, we further penalize the regression with the L1-norm (Lasso) to force some coefficients to zero. As some unigrams are not informative of political partisanship, the penalization mitigates over-fitting of the training set by shrinking coefficients.

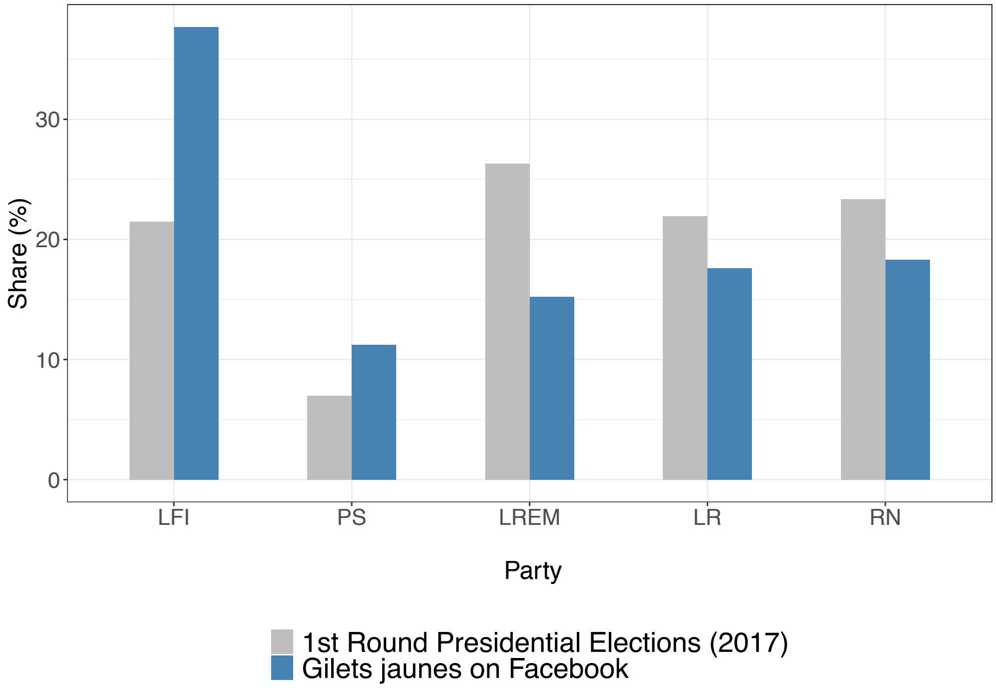
11 To assess the quality of the model, we shuffle the corpus and split it into a 80% training data subsample and a 20% test data subsample. We build the classifier in the training set and evaluate its performance in the test set. We obtain an accuracy score of 55.5%, when a random guess would correctly infer the author's party 20% of the time. Our model thus assigns the correct party to a text snippet between two and three times more often than a guess at random would. For comparison, Peterson and Spirling (2018) predict party affiliation with an accuracy between 60 and 80% for two parties. Additional validation metrics and details of the implementation can be found in the Appendix.

**Figure 4.2:** Results of the topic model


(a) Places [2.9%]
(b) Towns and hours [4.4%]
(c) Support [3.9%]
(d) Diffusion [4.8%]
(e) Economic concerns [5.9%]
(f) Political institutions [7.6%]
(g) Food and objects [6%]
(h) Critiques [6%]
(i) Insults [4.5%]
(j) Violence [5.9%]
(k) Conspiracy [5.9%]
(l) Actions [7%]
(m) Names [6.6%]
(n) Foreign Languages [8.4%]
(o) Other [20%]

Notes: This figure shows wordclouds associated to the fifteen topics we identify in our corpus. The size of words is determined by a term frequency-inverse document frequency (TF-IDF) metric, where each document is the entire collection of sentences associated to a given topic. This metric is meant to give higher scores to words which are (i) more frequent in the corpus and (ii) particularly meaningful for each topic. Wordclouds are boxed inside a rectangle when the average sentiment of messages in the topic is negative. Squared brackets indicate topic frequency (as the share of total messages in the corpus).

**Figure 4.3:** Predicted Political Leaning of the Yellow Vests




Notes: This figure compares the predicted political leaning of the active Yellow Vest users on Facebook (in blue) to the scores of each party at the first round of the presidential elections (in light gray). Vote shares at the elections are modified so as to sum up to a hundred (there were other smaller parties that we exclude from the analysis). We assign a political leaning to each Facebook user in our corpus based on the average probability of her sentences being pronounced by a given party according to our supervised learning model.

## 4.2 The Online and Offline Activity of the Yellow Vests

### 4.2.1 Online to offline

We now describe the coalescence of the movement which culminated in the first large action of the Yellow Vests: the 11/17 blockades. The view that social media were widely used to organize and coordinate these blockades is substantiated by the evidence presented in the previous section. To move beyond these qualitative elements, we perform a spatial regression analysis to explain why some parts of France had some Yellow Vest activity, while others did not.

In this spatial regression, we study the mobilization at the municipality level. There are more than 35 thousand municipalities in France,[12] and their boundaries date back to the French Revolution. They are the lowest level of government, which allows us to gather data on a wide range of characteristics along the following dimensions: demography, geography, commuting and mobility, labor market, votes, and Facebook penetration.[13]

12 Excluding French overseas territories and Corsica from our sample leaves us with a sample of 34,434 municipalities.

13 This variable is measured as the sum of Facebook accounts where users declare either to live in or to come from the municipality over the municipal population

**The low predictive power of traditional characteristics**

We first ask the following question: was it possible to predict which locations were going to experience a blockade at the beginning of the movement? To that end, we construct measures of mobilization at the municipal level, and consider two periods: until the 11/17 blockades (hereafter, "pre-11/17") and after them (hereafter, "post-11/17"). Early mobilization is defined by a variable $M_m^{\text{early}}$ that can either be: $B_m^{11/17}$, a dummy variable equal to 1 if there was a blockade in the municipality $m$ on 11/17; $S_m^{\text{pre-11/17}}$, the signature rate in the municipality before 11/17; $NG_m^{\text{pre-11/17}}$ the number of local Facebook groups before 11/17[14]; and $LG_m^{\text{pre-11/17}}$, a dummy equal to 1 if there was a municipal Facebook group before 11/17. We use OLS to estimate equation (4.1):

$$M_m^{\text{early}} = X_m\gamma + \delta_{LZ(m)} + \varepsilon_m \tag{4.1}$$

where $X_m$ is a large set of economic, geographic, demographic, and political controls (see the full list in Table 4.5) and $\delta_{LZ(m)}$ is a LZ fixed effect to account for fixed unobserved heterogeneity at a higher spatial level. Rather than interpreting each estimate separately, we focus on the explanatory power of such a descriptive model. To that end, we perform a variance decomposition exercise using the method of Shorrocks (1982).

Table 4.5 in Appendix displays the results of this decomposition. We find that most municipal characteristics have very little explanatory power: even if we do not control for LZ fixed effects, the full set of controls never accounts for more than 10% of the spatial dispersion in the intensity of the mobilization. This finding suggests that the mobilization was very difficult to anticipate. We further find that regional heterogeneity explains most of the variation in signature rates, but hardly matters to explain the spatial dispersion in blockades or Yellow Vest Facebook activity. This finding confirms the very local nature of the mobilization.

**The high predictive power of early measures of online mobilization.**

Since local characteristics do not predict accurately the presence of a blockade, we then assess whether variables of online mobilization before the 11/17 blockades have some predictive power. To that end, we focus on $M_m^{\text{early}} = B_m^{\text{11-17}}$ and estimate an augmented version of equation (4.1):

$$B_m^{11/17} = O_m^{\text{pre-11/17}}\beta + X_m\gamma + \delta_{LZ(m)} + \varepsilon_m \tag{4.2}$$

where $O_m^{\text{pre-11/17}} = \left(S_m^{\text{pre-11/17}}, NG_m^{\text{pre-11/17}}, LG_m^{\text{pre-11/17}}\right)$ is a vector of our three measures of early online mobilization.

OLS estimates of this regression are displayed in Table 4.2. On top of estimates for $\beta$, we show estimates for the coefficients on three control variables: the Facebook penetration rate and two variables related to roundabouts, the discussion of which is deferred to Section 4.2.2. Column (1) shows the baseline regression results associated with Column (8) in Table 4.5. The adjusted R-squared is equal to 20%. Then, Columns (2) to (4) include each measure of early online mobilization separately. Columns (2) shows that the petition signature rate is positively correlated with the blockades, which we interpret as signatures being a signal for mobilization potential. Columns

[14] We construct a weighted measure of the number of local Facebook groups at the municipal level, where all groups of different levels are included, apportioned by the population of the municipality in the département or the region.

(3) and (4) show that, as expected, early mobilization on Facebook is strongly correlated with the occurrence of a blockade. In particular, municipalities with a dedicated Facebook group are 32 p.p. more likely to experience a blockade. Controlling for the existence of a municipal group increases the prediction accuracy, as measured by the adjusted R-squared, by 25%.

Finally, in Column (5), we control for the three measures simultaneously. The coefficient associated with the signature rate is quite stable compared to Column (2), which suggests that both types of online mobilization are not substitutes for one another. On the contrary, the coefficient associated with the number of local Facebook groups (in the municipality, but also at a higher geographical level) drops to a precise zero when we control for the existence of a municipal group. This result suggests a close spatial mapping between early mobilization on Facebook and the organization of the blockades.

The estimates associated with the Facebook penetration rate are also interesting. Columns (1) and (2) show that there is a positive correlation between Facebook penetration and the blockades, in line with a vast body of evidence on the role of social media in social unrest movements. However, as shown in columns (3) to (6), this correlation disappears once we control for local measures of Facebook use that are specific to the Yellow Vests. This finding suggests that in a context where most people use social media[15], it is less access to social media than specific use of social media that can impact social movements.

[15] There are over 30 million accounts referenced in our data, out of less than 70 million inhabitants.

**Table 4.2:** Predictors of a blockade in a municipality

| | (1) | (2) | (3) | (4) | (5) | (6) |
|---|---|---|---|---|---|---|
| Signatures (pre-17/11) | | $0.012^{***}$ | | | $0.010^{***}$ | $0.010^{***}$ |
| | | (0.002) | | | (0.002) | (0.002) |
| Nb. Groups (pre-17/11) | | | $0.036^{***}$ | | 0.003 | 0.003 |
| | | | (0.003) | | (0.009) | (0.009) |
| Local Group (pre-17/11) | | | | $0.325^{***}$ | $0.301^{***}$ | $0.288^{***}$ |
| | | | | (0.024) | (0.073) | (0.068) |
| Roundabout Density (Municipality) | | | | | | $0.010^{***}$ |
| | | | | | | (0.003) |
| Roundabout Density (Other Munic. in L.Z.) | | | | | | $-0.086^{***}$ |
| | | | | | | (0.016) |
| Facebook Penetration | $0.465^{*}$ | $0.480^{*}$ | -0.069 | -0.305 | -0.280 | -0.250 |
| | (0.241) | (0.248) | (0.229) | (0.320) | (0.309) | (0.291) |
| Controls | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| Fixed Effects | LZ | LZ | LZ | LZ | LZ | LZ |
| Observations | 34,434 | 34,434 | 34,434 | 34,434 | 34,434 | 34,434 |
| Adjusted R-squared | 0.200 | 0.203 | 0.240 | 0.250 | 0.252 | 0.262 |
| Within R-squared | 0.185 | 0.188 | 0.225 | 0.235 | 0.238 | 0.248 |

Notes: This table shows OLS estimates for a linear probability model predicting whether a municipality experienced a blockade or not, as formalized in equation (4.2). "Signatures (pre-11/17)" is the municipality's signature rate of the Change.org petition prior to 11/17; "Nb. Groups (pre-11/17)" is the apportioned number of Facebook groups (from all geographical levels) created before 11/17. These two variables are standardized. "Local Group (pre-11/17)" is a dummy variable for the existence of a specific municipal group prior to 11/17. We measure Facebook penetration in a municipality as the number of Facebook users who declare either to live or to come from that municipality, divided by the municipal population. This variable is standardized and multiplied by 1000. The last column represents the first-stage estimates of equation (4.4), associated with the second-stage results of Table 4.3. The two instruments that we use are the number of roundabouts per squared kilometer in the municipality and the corresponding average for all other municipalities in the LZ. Both variables are standardized. Standard errors are clustered at the LZ level. $^{*}$: $p < 0.1$, $^{**}$: $p < 0.05$, $^{***}$: $p < 0.01$.

### 4.2.2 Offline to Online

Our analysis of the Yellow Vest movement supports previous findings in the literature showing that online mobilization can spur offline protests. A complementary question is the following: do protests on the streets help to keep online movements online? We answer this question in this section, evaluating the effects of the 11/17 blockades on further online action. To measure a causal effect, we use an instrumental variables strategy which relies on the local supply of easy-to-block locations: roundabouts.

#### Empirical strategy

The rationale for the instrument stems from the fact that calls for demonstrations urged protesters to block roundabouts. By design, they allow to block several roads at a time and are equipped with a central median strip on which it is convenient to set camp. The history of roundabouts makes it likely that the conditional distribution of local roundabout density reflects local idiosyncrasies. Roundabouts are partly a French architectural fad, invented in 1906 by the French urban planner Eugène Hénard.[16] While there are plausible reasons related to road safety to support them, roundabouts can almost always be replaced with traffic lights. A map of the prediction error of roundabout density at the municipality level, after an OLS regression including our controls, shows a seemingly random distribution – see Appendix Figure 4.15.

Appendix Table 4.6 explores the determinants of the spatial distribution of roundabouts in France using the variance decomposition method of Shorrocks (1982). As one can expect, the distribution of roundabouts in France is closely related to the population distribution: more than 40% of the variance in roundabout density is explained by the distribution of population. LZ fixed-effects only have little explanatory power, indicating low levels of spatial autocorrelation in roundabout density. Importantly, after controlling for population density, other controls only explain a residual fraction of the variation in roundabout density/

Column (7) in Table 4.2 shows that roundabouts played an important role in the organization of the protests: we find that increasing the density of roundabouts in a municipality by one standard deviation increases the probability of a blockade by 1.2 p.p. However, this is not the whole story: since organizing a blockade requires significant manpower, protesters had to coordinate to choose blockade locations. This spatial coordination problem suggests a second instrument, which is the mirror image of the first: the density of roundabouts in the other municipalities of the LZ. As shown in Column (7), an increase of one standard deviation in this variable decreases the blockade probability of a municipality by 9.6 p.p.

Our econometric specification is the following – our first stage predicts the probability of a blockade in a municipality:

$$B_m^{11/17} = \alpha_1 + O_m^{\text{pre}-11/17}\beta_1 + X_m\gamma_1 + \delta_1^{LZ(m)} + \zeta_1 R_m + \zeta_2 R_{LZ(-m)} + \varepsilon_m \tag{4.3}$$

Where $R_m$ is the density of roundabouts (number of roundabouts per square kilometer) in municipality $m$ and $R_{LZ(-m)}$ is the density of roundabouts in all municipalities of the life zone $LZ$

16 There are over sixty-thousand roundabouts in France, which is four times more than the United Kingdom. One third of French municipalities have at least one.

of municipality $m$, excluding municipality $m$.

The second stage regresses a measures of online mobilization after 11/17, $Y_m^{\text{post}}$, on the predicted blockade probability from equation 4.3:

$$Y_m^{\text{post}} = \alpha_2 + O_m^{\text{pre}}\beta_2 + X_m\gamma_2 + \delta_2^{LZ(m)} + \eta\widehat{B_m^{11/17}} + \epsilon_m, \tag{4.4}$$

We primarily consider four measures of post-11/17 online mobilization: the signature rate of the Change.org petition after 11/17, the creation of a local group post-11/17, the number of members (per inhabitant) of these new local groups and the number of messages (per inhabitant) of these new local groups.[17] Our main coefficient of interest is $\eta$, which provides under our identifying assumptions the causal effect of a 11/17 blockade on subsequent online mobilization.

**Results**

Results for the second stage are shown in Table 4.3. Column (1) shows that even though the bulk of petition signatures occurred prior to 11/17, having a blockade increases the post-11/17 signature rate by 1.3 standard deviations. This result suggests that protests helped spread information about the Yellow Vests' demands, at a period (until the end of 2018) where support for the movement was still quite high in the population. Previous signatory rate is also correlated with subsequent signatory dynamics; Facebook penetration seems to play no role in this dynamics, while the existence of the local group seems negatively correlated with an increase in signatories after the first offline protest.

We also find a strong positive impact of blockades on subsequent Facebook activity related to the movement: a blockade in a municipality triggers the creation of 0.70 new local Facebook groups (column 2), which can be associated with a higher increase in Facebook Yellow Vests members, by 0.032 standard deviation, which corresponds to 3.9 additional members for 1,000 inhabitants (column 3), as well as a higher online activity with around 7 additional posts per 1,000 inhabitants compare to municipalities that were not blocked during the first Act (column 4).

As shown in Appendix Table 4.7, these results are robust to several specification changes.[18] In particular, the estimates are very stable if we do not control for municipal characteristics (Panel A) or if we use only one roundabout instrument instead of two (Panels B and C). These three tests are reassuring regarding the validity of the exclusion restriction. Effects are also fairly similar if we define location fixed effects and the instrument at the commuting zone ($N = 297$) rather than at the LZ ($N = 1596$) level (Panel D).

Overall, our results show that the 11/17 blockades triggered a new wave of online activity, especially on Facebook. It seems that the social network was used by protesters who had met in the streets to pursue their discussions in the following months. We can rule out an interpretation of these results as a relabeling artifact, whereby new groups would only result from the transfer of old groups into new groups, created in honor of the 11/17 blockades, because the existence of early groups is uncorrelated with the creation of new groups after the first blockade.

While most blockades were rapidly lifted following 11/17, this rebound effect online kept the movement alive. However, was this new online activity still mobilization per se, that is, meant to

[17] In Table **??**, we show that the results that we obtain are robust to considering logs instead of rates.

[18] Interestingly, the 2SLS estimates are much higher than the OLS estimates, which suggests that political sentiment or political traditions prompting more blockades also translate into lower online mobilization, and vice-versa.

**Table 4.3:** Effects of a blockade on further mobilization

| | Signatures (post-11/17) (1) | Local Group (post-11/17) (2) | Members in Local Groups (3) | Messages (post-11/17) (4) |
|---|---|---|---|---|
| Local Blockade | 1.267*** | 0.701*** | 0.032*** | 0.036** |
| | (0.257) | (0.118) | (0.012) | (0.016) |
| Signatures (pre-17/11) | 0.479*** | -0.000 | 0.000 | -0.000 |
| | (0.049) | (0.001) | (0.001) | (0.002) |
| Nb. Groups (pre-17/11) | -0.018 | -0.008 | 0.000 | 0.001 |
| | (0.015) | (0.005) | (0.001) | (0.001) |
| Local Group (pre-17/11) | -0.160 | 0.012 | -0.005 | -0.007 |
| | (0.119) | (0.060) | (0.009) | (0.014) |
| Facebook Penetration | 2.900 | 10.500*** | 5.875** | 5.747* |
| | (3.327) | (2.712) | (2.913) | (2.972) |
| Controls | ✓ | ✓ | ✓ | ✓ |
| Fixed-Effects | L.Z. | L.Z. | L.Z. | L.Z. |
| Observations | 34,434 | 34,434 | 34,434 | 34,434 |
| Kleibergen-Paap F-stat | 31.061 | 31.061 | 31.061 | 31.061 |
| p-value Hansen | 0.446 | 0.866 | 0.186 | 0.136 |

Notes: This table shows 2SLS estimates of the impact of a municipal blockade on four measures of online mobilization after 11/17: the signature rate of the Change.org petition after 11/17 (column 1), the creation of a local group post-11/17 (column 2), the number of members per inhabitant, with s.d.(members post /pop) = 0.1218 (column 3) and posts per inhabitant, with s.d.(publications post /pop) = 0.1961 (column 4) in these newly created local groups. Signature rates, Number of groups, Number of members and posts per inhabitants are standardized. Standard errors are clustered at the LZ level. *: $p < 0.1$, **: $p < 0.05$, ***: $p < 0.01$.

organize further mobilization? Or only used for socialization, and/or as a place to voice discontent? By progressively morphing into a purely online movement with no organizational purpose, the movement and its members changed in focus, as we show in the next section.

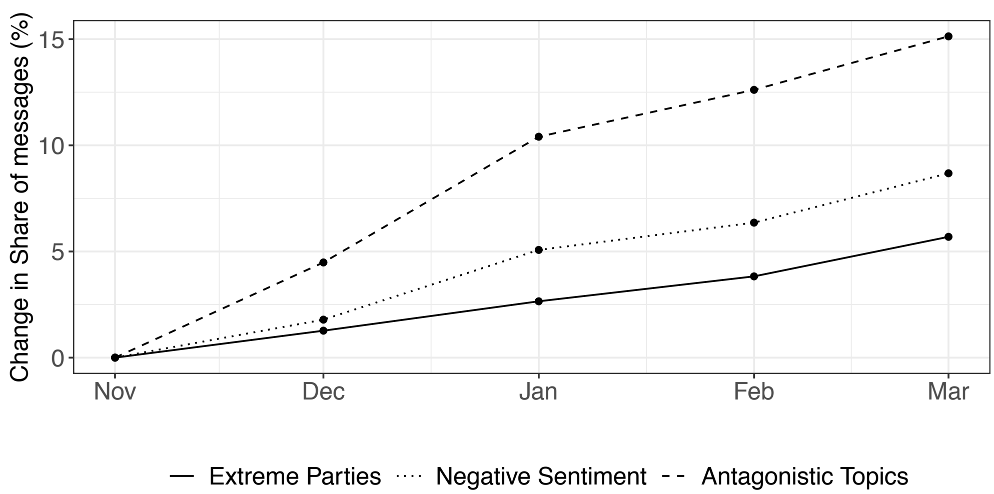
## 4.3 The Rise of Online Antagonism and Decline of the Yellow Vests

The 11/17 blockades were both the first and most intense day of offline demonstrations for the Yellow Vests. In the weeks that followed 11/17, the number of offline protesters sharply declined, and the movement overall gradually lost strength. In this section, we document the decline of the movement, following the online discussions of the Yellow Vests over a six-month period between the end of October 2018 and the beginning of April 2019. To do so, we rely on the textual analysis of Facebook pages described in Section 4.1.3.

### 4.3.1 Evolution of online discussions

Figure 4.4 shows how the online discussions of the Yellow Vests radicalized over time. The share of messages associated to antagonistic topics increased by 15 p.p. between November 2018 and March 2019. Figure 4.5 shows how the importance of each of our 15 topics evolved over the study period: messages associated to political concerns became relatively rarer over time, while messages associated with violence, conspiracy theories, or containing insults became more widespread. A similar trend towards antagonism is captured by our other classifications of Facebook messages: the share of messages classified as negative (resp., associated to an extreme party – on the far-right or the far-left) increased by 8 p.p. (resp. 6 p.p.) over the study period.

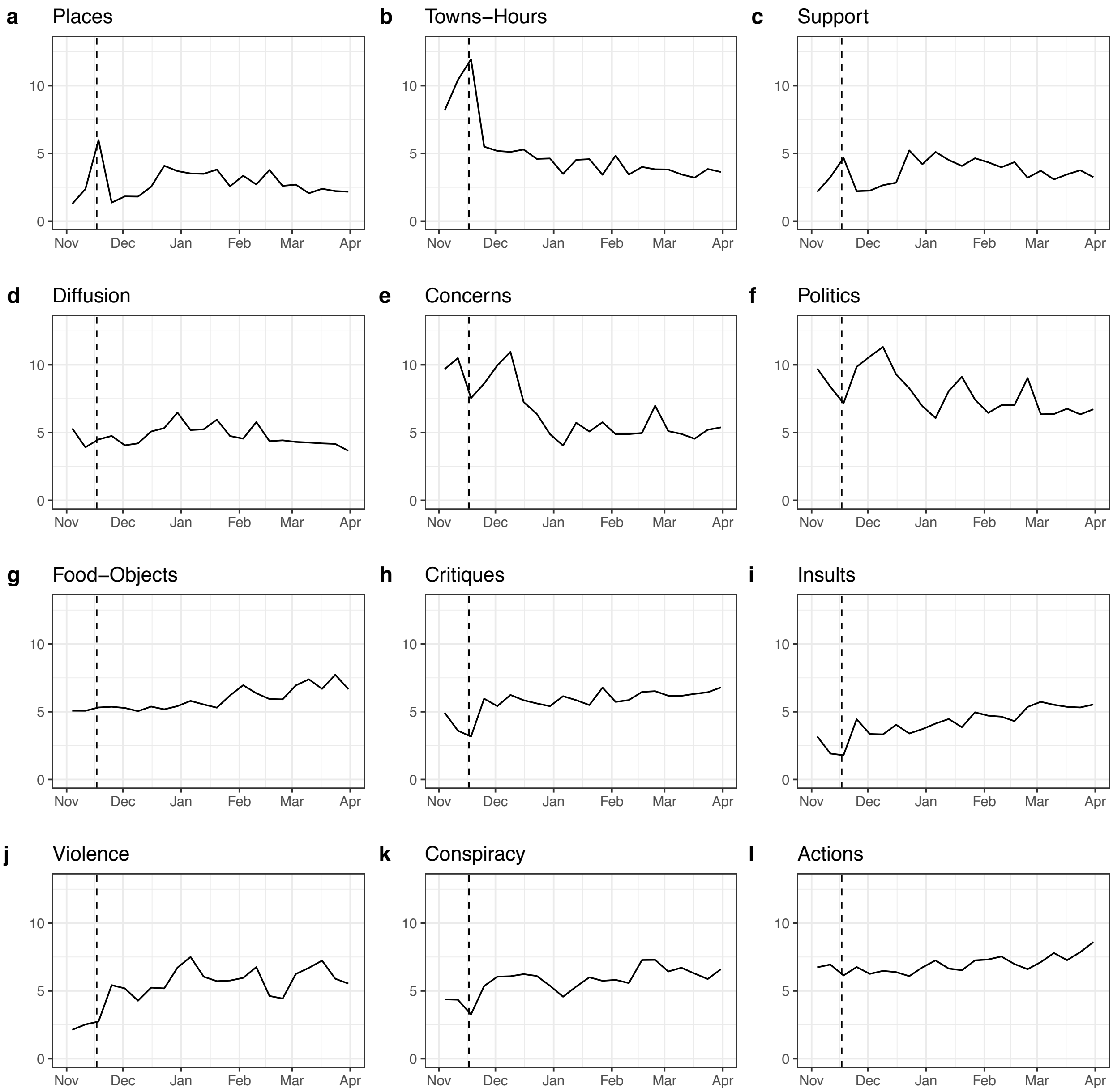
**Figure 4.4:** Evolution of Partisanship, Sentiment, and Antagonism Over Time




Notes: This figure shows changes in the monthly share of messages which fall in the following three categories, defined in Section 4.1.3 : (i) messages associated with the far-left or the far-right political; (ii) messages associated with a negative sentiment; (iii) messages associated with antagonistic topics. We use November as the reference month.

This trend could be explained by two mechanisms: first, it could be explained by a shift of the base of active users towards more antagonistic ones: we refer to such a shift as the extensive margin of radicalization. Second, it could be explained by the increased propensity of any individual user to post an antagonistic message as time passes: we refer to such changes as the intensive margin of radicalization.

**Figure 4.5:** Topic Shares in Facebook Discussions Over Time




Notes: This figure shows weekly shares of associated with the twelve topics of interest – see Figure 4.2. For all topics, the vertical dashed line corresponds to 11/17. The share of messages associated to violence is below 2.5% in early November and is consistently above 5% after December 10.

### 4.3.2 Linear decomposition

In order to assess whether the trends that we observe are more likely to reflect shifts in the intensive or extensive margin of radicalization, we exploit the panel dimension of the data, and the fact that we are able to follow individual Facebook users over time.

Suppose that the probability $Y_{i,t}$ that a message posted by user $i$ during month $t$ is classified as antagonistic is determined by the following linear probability model:

$$Y_{i,t} = \delta_i + \gamma_t + \varepsilon_{i,t} \tag{4.5}$$

Where $\delta_i$ is a user-specific propensity to post a conflictual message, $\gamma_t$ accounts for the additional propensity of users to post conflictual messages during month $t$, and $\varepsilon_{i,t}$ is a residual noise. The share of messages that are violent during a given month $t$ is then given by:

$$\bar{Y}_t = \underbrace{\sum_i s_{i,t}\hat{\delta}_i}_{\text{Extensive margin}} + \underbrace{\hat{\gamma}_t}_{\text{Intensive margin}} \tag{4.6}$$

Where $s_{i,t}$ is the share of messages posted during month $t$ that originated from user $i$ (it is a weighted average of individual fixed effects). The first term of expression (4.6) corresponds to the average propensity to post antagonistic messages for users active during month $t$. An increase of this term over time means that as time passes, the share of messages posted by more antagonistic users increases. It therefore captures shifts in the extensive margin of radicalization. An increase in the second term of expression (4.6) corresponds to an increase in the propensity of any given user to post an antagonistic message at a given time: it therefore captures the intensive margin of radicalization.

### 4.3.3 Results

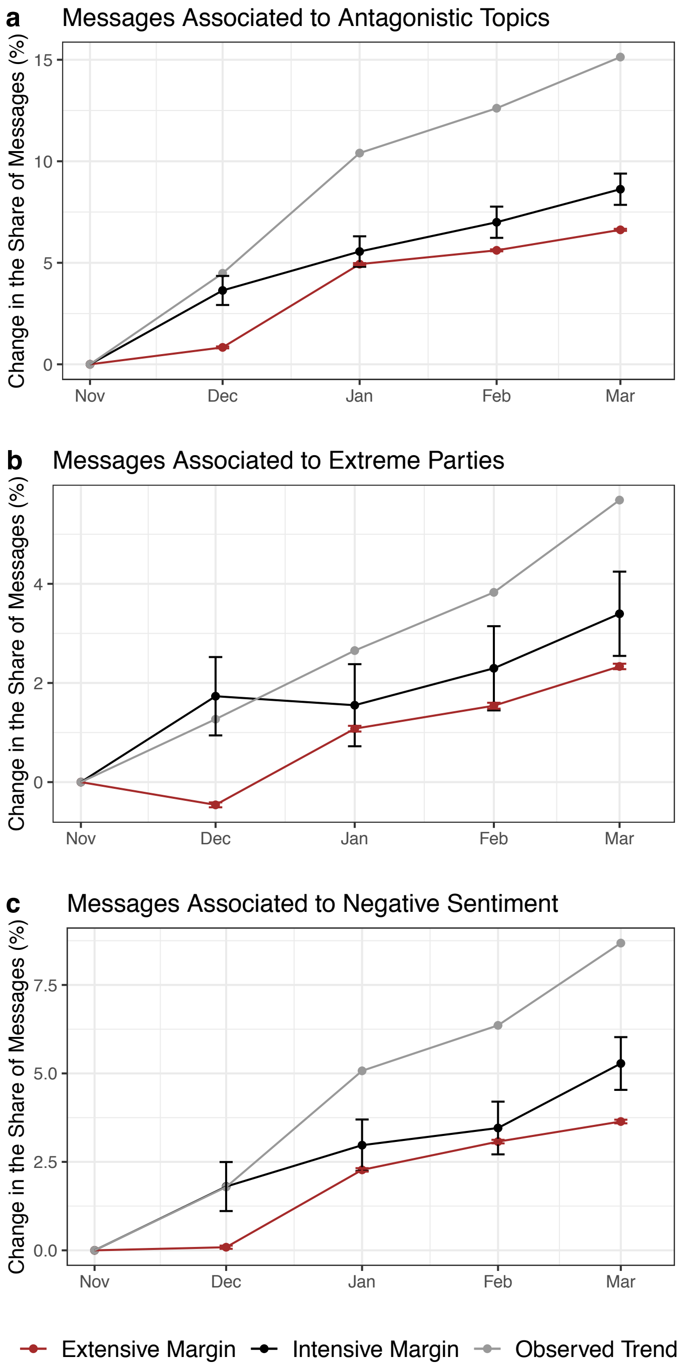
We estimate equation (4.6) using our three measures of radicalization, using as the outcome variable $Y_{i,t}$ dummies indicating whether a message was associated with an antagonistic topic, whether a message was associated with an extreme political party, whether the sentiment score associated to a message was below 0 (indicating negative sentiment). Results are presented in Figure 4.6.[19]

At first sight, the graphs of the decomposition of radicalization display a similar contribution of the two margins of radicalization. It yields that both a selective attrition of less radical members (individual fixed effects) and the radicalization of remaining members occurred over the period (time fixed effects).

Nevertheless, the intensive margin always dominate the extensive one, which implies that the radicalization of protesters has contributed more to the radicalization of the movement than selective attrition. Interestingly, the selective attrition was nonexistent at the beginning of the protest. Looking at the probability of writing a sentence associated with a politically extreme party, it seems that the movement enlarge in terms of users (with a negative incidence of individual fixed effects), while users radicalized strongly during this phase of coalescence. The major phase of selective attrition is between December and January after the President has made

19 For binary dependent variables, results from the same exercise with a logistic regression may be found in the appendix.

**Figure 4.6:** Extensive and Intensive Margins of Radicalization




Notes: This figure decomposes the increase in online antagonism presented in Figure 4.4 using the method outlined in Section 4.3.2. Panel (a) presents estimates for the probability of posting a message associated with an antagonistic topic. Panel (b) presents estimates for the probability of writing a sentence associated with a politically extreme party (i.e., on the far left or the far right). Panel (c) presents estimates for the probability of posting a message associated with negative sentiment. Standard errors for the time fixed effects are computed with the variance-covariance matrix. Standard errors for the individual fixed effects are computed via bootstrap with 200 iterations. Confidence intervals are plotted at the 95% confidence level.

significant concessions to the protesters and while the movement is struggling to consolidate. These observations suggests that the selective attrition lag behind the radicalization of messages. However, we cannot state potential causal relationship between the two with our data, such investigation offers promising way for further research.

Finally, the decline of the movement accelerate after January with the simultaneous contribution of higher attrition and higher radicalization of the bunch of members that remains. This crowds pages with antagonistic posts, preventing further consolidation of the movement.

## 4.4 Conclusion

Large protest movements have become a combination of online and offline events. Many have hypothesized that social media would favor the emergence and spread of protests, lowering coordination costs and making it easier to signal discontent. A large literature has shown that social media and online protest can bring more demonstrators to the streets. Using the Yellow Vest movement as a case study, we empirically show that there is a two-way street between online and offline mobilization. When the Yellow Vests protested in a location, this increased the subsequent online activity of the protesters. This finding suggests that social media can also help keep protest movements alive.

The Yellow Vest movement did stay alive much longer than commentators initially imagined. Instead of remaining a single day of road blockades, the movement expanded and major cities in France witnessed weekly Yellow Vest demonstrations for over a year. However, the movement fragmented over time and Yellow Vest candidates at the 2019 European elections only won a residual share of the vote. Our analysis shows that as time passed, the movement radicalized, first driven by individuals being more antagonist and then combined with the attrition of moderate protesters.

Overall, we believe this evidence help better understand an important tension of hybrid social movements. On the one hand, new means of information sharing and coordination allow for more accessible movements where anyone can initiate or organize a protest. On the other hand, a strong dependency on a leaderless social media infrastructure, where anyone can voice their individual opinion, may dampen the ability to structure long-lasting, effective political campaigns. As summarized by della Porta (2013), mobilization triggered by social media can be “very successful in terms of number, but tends to be more volatile and intermittent than in the past.”

# Bibliography


**Acemoglu, Daron, Tarek A Hassan, and Ahmed Tahoun**, "The power of the street: Evidence from Egypt's arab spring," *Review of Financial Studies*, 2018, *31* (1), 1–42. ↑172

**Algan, Yann, Clément Malgouyres, and Claudia Senik**, "Territoires, bien-être et politiques publiques," *Note CAE*, 2020, (55). ↑173

_ , **Elizabeth Beasley, Daniel Cohen, and Martial Foucault**, *Les origines du Populisme*, La République des Idées. Seuil, 2019. ↑173

**Allcott, Hunt and Matthew Gentzkow**, "Social media and fake news in the 2016 election," *Journal of economic perspectives*, 2017, *31* (2), 211–36. ↑173

**Arora, Sanjeev, Yingyu Liang, and Tengyu Ma**, "A simple but tough-to-beat baseline for sentence embeddings," *Conference paper at ICLR 2017*, 2017. ↑176

**Bakker, Tom P and Claes H De Vreese**, "Good news for the future? Young people, Internet use, and political participation," *Communication research*, 2011, *38* (4), 451–470. ↑172

**Barbera, Salvador and Matthew O Jackson**, "A model of protests, revolution, and information," *Quarterly Journal of Political Science*, 2020, *15* (3), 297–335. ↑172

**Barrera, Oscar, Sergei Guriev, Emeric Henry, and Ekaterina Zhuravskaya**, "Facts, alternative facts, and fact checking in times of post-truth politics," *Journal of Public Economics*, 2020, *182*, 104–123. ↑173

**Bastos, Marco T., Dan Mercea, and Arthur Charpentier**, "Tents, Tweets, and Events: The Interplay Between Ongoing Protests and Social Media," *Journal of Communication*, 2015, *65* (2), 320–350. ↑172

**Bendali, Zakaria and Aldo Rubert**, "Les sciences sociales en gilet jaune," *Politix*, 2020, *4*, 177–215. ↑173

**Bennett, W Lance and Alexandra Segerberg**, "The logic of connective action: Digital media and the personalization of contentious politics," *Information, communication & society*, 2012, *15* (5), 739–768. ↑171

**Blumer, Herbert**, *Principles of sociology* number 26, Barnes & Noble, 1969. ↑171

**Borge-Holthoefer, Javier, Alejandro Rivero, Iñigo García, Elisa Cauhé, Alfredo Ferrer, Darío Ferrer, David Francos, David Iniguez, María Pilar Pérez, Gonzalo Ruiz et al.**, "Structural and dynamical patterns on online social networks: the spanish may 15th movement as a case study," *PloS one*, 2011, *6* (8), e23883. ↑172

**Boxell, Levi, Matthew Gentzkow, and Jesse M. Shapiro**, "Greater Internet use is not associated with faster growth in political polarization among US demographic groups," *Proceedings of the National Academy of Sciences*, 2017, *114* (40), 10612–10617. ↑173

**Boyer, Pierre C., Thomas Delemotte, Germain Gauthier, Vincent Rollet, and Benoît Schmutz**, "Les déterminants de la mobilisation des Gilets jaunes," *Revue Économique*, 2020, *71* (1), 109–138.

**Bursztyn, Leonardo, Davide Cantoni, David Yang, Noam Yuchtman, and Jane Zhang**, "Persistent political engagement: Social interactions and the dynamics of protest movements," *American Economic Review: Insights*, 2021, *3* (2), 233–50. ↑172

**Campante, Filipe, Ruben Durante, and Francesco Sobbrio**, "Politics 2.0: The multifaceted effect of broadband internet on political participation," *Journal of the European Economic Association*, 2018, *16* (4), 1094–1136. ↑173

**Caren, Neal and Sarah Gaby**, "Occupy online: Facebook and the spread of Occupy Wall Street," *Available at SSRN 1943168*, 2011. ↑175

**Castells, Manuel**, *Networks of Outrage and Hope: Social Movements in the Internet Age*, Wiley, 2015. ↑171

**Christiansen, Jonathan**, "Four stages of social movements," *Sociology Reference Guide*, 2009. ↑172

**Clarke, Killian and Korhan Kocak**, "Launching revolution: Social media and the Egyptian uprising's first movers," *British Journal of Political Science*, 2020, *50* (3), 1025–1045. ↑172

**Cointet, Jean-Philippe, Pedro Ramaciotti Morales, Dominique Cardon, Caterina Froio, Andrei Mogoutov, Benjamin Ooghe, and Guillaume Plique**, "De quelle(s) couleur(s) sont les Gilets jaunes? Plonger des posts Facebook dans un espace idéologique latent," *Statistique et Société*, 2021, *9* (1-2), 79–107. ↑173

**Davoine, Eva, Etienne Fize, and Clément Malgouyres**, "Les déterminants locaux du mécontentement : analyse statistique au niveau communal," *Focus CAE*, 2020, (39). ↑173

**della Porta, Donatella**, *Can democracy be saved?*, Polity Press, 2013. ↑190

**Demszky, Dorottya, Nikhil Garg, Rob Voigt, James Zou, Matthew Gentzkow, Jesse Shapiro, and Dan Jurafsky**, "Analyzing polarization in social media: Method and application to tweets on 21 mass shootings," *arXiv preprint arXiv:1904.01596*, 2019. ↑176

**Earl, Jennifer and Katrina Kimport**, *Digitally enabled social change: Activism in the internet age*, Mit Press, 2011. ↑171

**Enikolopov, Ruben, Alexey Makarin, and Maria Petrova**, "Social media and protest participation: Evidence from Russia," *Econometrica*, 2020, *88* (4), 1479–1514. ↑172

**Falck, Oliver, Robert Gold, and Stephan Heblich**, "E-lections: Voting Behavior and the Internet," *American Economic Review*, 2014, *104* (7), 2238–65. ↑173

**Friedman, Jerome, Trevor Hastie, Robert Tibshirani et al.**, *The elements of statistical learning*, Vol. 1, Springer series in statistics New York, 2001. ↑177

**Gaffney, Devin**, "iranElection: Quantifying online activism," in "In Proceedings of the Web Science Conference (WebSci10)" Citeseer 2010. ↑172

**Gavazza, Alessandro, Mattia Nardotto, and Tommaso Valletti**, "Internet and politics: evidence from UK local elections and local government policies," *Review of Economic Studies*, 2019, *86*, 2092–2135. ↑173

**Gentzkow, Matthew, Bryan Kelly, and Matt Taddy**, "Text as data," *Journal of Economic Literature*, 2019, *57* (3), 535–74. ↑176

**González-Bailón, Sandra, Javier Borge-Holthoefer, Alejandro Rivero, and Yamir Moreno**, "The dynamics of protest recruitment through an online network," *Scientific reports*, 2011, *1* (1), 1–7. ↑172

**Grimmer, Justin and Brandon M Stewart**, "Text as data: The promise and pitfalls of automatic content analysis methods for political texts," *Political analysis*, 2013, *21* (3), 267–297. ↑176

**Guriev, Sergei and Elias Papaioannou**, "The political economy of populism," *CEPR Discussion Paper*, 2020, (14433). ↑173

_ **, Nikita Melnikov, and Ekaterina Zhuravskaya**, "3G Internet and Confidence in Government," *Available at SSRN 3456747*, 2019. ↑173

**Halberstam, Yosh and Brian Knight**, "Homophily, group size, and the diffusion of political information in social networks: Evidence from Twitter," *Journal of Public Economics*, 2016, *143*, 73–88. ↑173

**Henry, Emeric, Ekaterina Zhuravskaya, and Sergei Guriev**, "Checking and sharing alt-facts," *CEPR Discussion Paper DP14738*, 2020. ↑173

**Jost, John T, Pablo Barberá, Richard Bonneau, Melanie Langer, Megan Metzger, Jonathan Nagler, Joanna Sterling, and Joshua A Tucker**, "How social media facilitates political protest: Information, motivation, and social networks," *Political psychology*, 2018, *39*, 85–118. ↑172

**Little, Andrew T**, "Communication technology and protest," *Journal of Politics*, 2016, *78* (1), 152–166. ↑172

**Loader, Brian D**, "Social movements and new media," *Sociology Compass*, 2008, *2* (6), 1920–1933. ↑171

**Manacorda, Marco and Andrea Tesei**, "Liberation technology: mobile phones and political mobilization in Africa," *Econometrica*, 2020, *88* (2), 533–567. ↑173

**Pan, Jennifer and Alexandra A Siegel**, "How Saudi crackdowns fail to silence online dissent," *American Political Science Review*, 2020, *114* (1), 109–125. ↑173

**Peterson, Andrew and Arthur Spirling**, "Classification accuracy as a substantive quantity of interest: Measuring polarization in westminster systems," *Political Analysis*, 2018, *26* (1), 120–128. ↑177

**Qin, Bei, David Strömberg, and Yanhui Wu**, "Why does China allow freer social media? Protests versus surveillance and propaganda," *Journal of Economic Perspectives*, 2017, *31* (1), 117–40. ↑172

**Rane, Halim and Sumra Salem**, "Social media, social movements and the diffusion of ideas in the Arab uprisings," *Journal of international communication*, 2012, *18* (1), 97–111. ↑172

**Rieder, Bernhard**, "Studying Facebook via data extraction: the Netvizz application," in "Proceedings of the 5th annual ACM web science conference" ACM 2013, pp. 346–355. ↑175

**Sebbah, Brigitte, Natacha Souillard, Laurent Thiong-Kay, and Nikos Smyrnaios**, "Les Gilets Jaunes, des cadrages médiatiques aux paroles citoyennes," *Rapport de recherche préliminaire*, 2018. ↑173

**Shorrocks, Anthony F**, "Inequality decomposition by factor components," *Econometrica*, 1982, pp. 193–211. ↑180, ↑183

**Tilly, Charles**, "From mobilization to revolution," *CRSO Working Paper*, 1977, (#156). ↑171

**Widmer, Philine, Sergio Galletta, and Elliott Ash**, "Media slant is contagious," *Center for Law & Economics Working Paper Series*, 2020, *14.* ↑177

**Yan, Xiaohui, Jiafeng Guo, Yanyan Lan, and Xueqi Cheng**, "A biterm topic model for short texts," in "Proceedings of the 22nd international conference on World Wide Web" 2013, pp. 1445–1456. ↑176

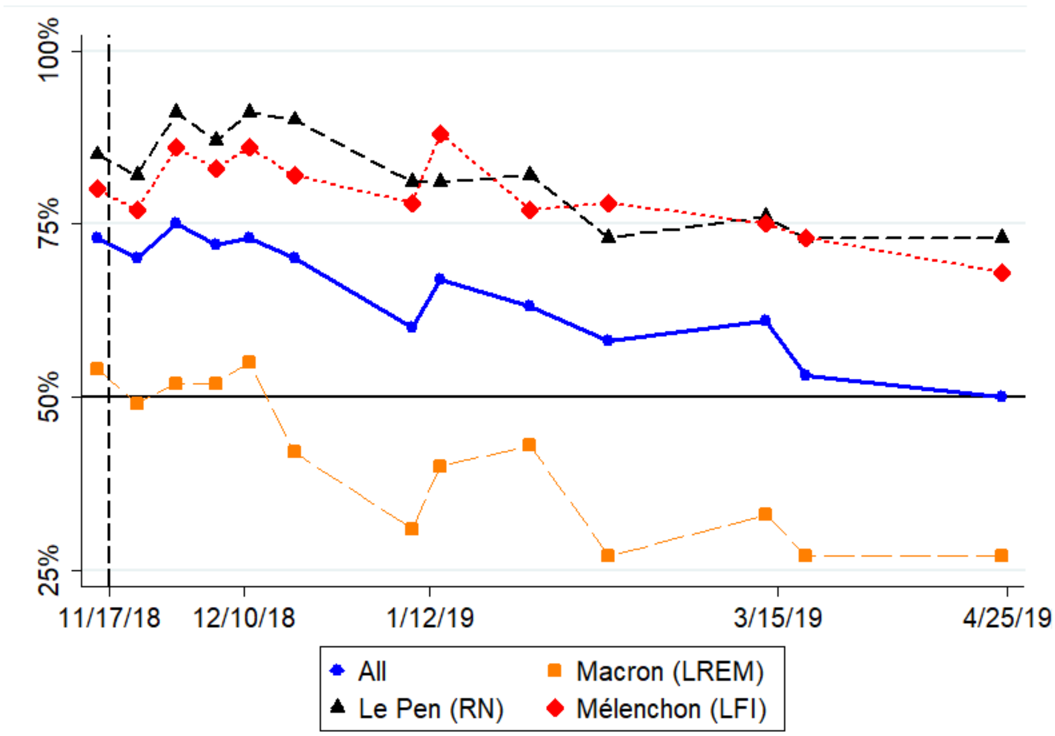
**Zhuravskaya, Ekaterina, Maria Petrova, and Ruben Enikolopov**, "Political Effects of the Internet and Social Media," *Annual Review of Economics*, 2020, *12*, 415–438. ↑173

## 4.A Data

### 4.A.1 Polls

Figure 4.7 reports polling results on the evolution of public support for the Yellow Vests movement.[20]

**Figure 4.7:** Evolution of the support for the Yellow Vests




Notes: This figure shows how the share of the population which declared they were supportive or sympathetic to the Yellow Vests movement evolved over time. We show poll results for the entire population, and split according to the candidate supported in the first round of the 2017 presidential election. Survey sample sizes are between 980 and 1,010 for the full sample and around 200 for each of the three sub-samples. The vertical dashed line corresponds to 11/17. ELABE, the survey institute from which we collected data, conducted polls on 11/14/2018, 11/21/2018, 11/28/2018, 12/5/2018, 12/11/2018, 12/19/2018, 1/9/2019, 1/14/2019, 2/13/2019, 3/13/2019, 3/20/2019, and 4/24/2019.

### 4.A.2 Online Petition

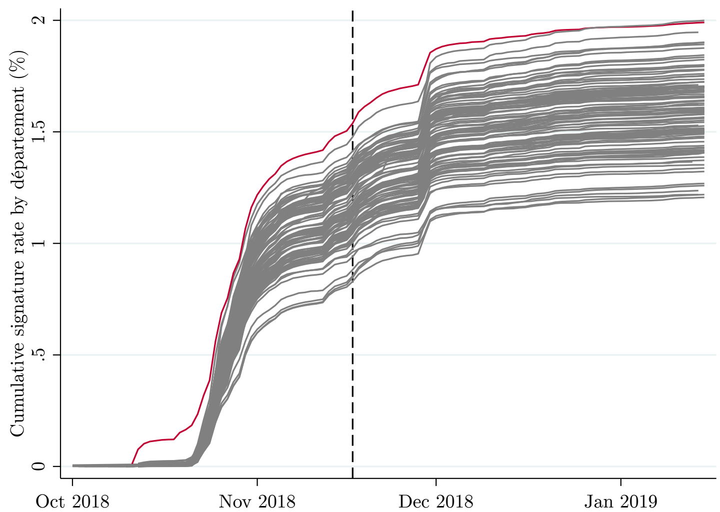
### 4.A.3 Facebook

The main websites that were associated with the organization of the movement coordinated the demonstrations by listing local dedicated Facebook groups.[21] To document online mobilization, we looked for public Facebook groups and pages related to the movement.

[20] Our data comes from ELABE, a polling institute which conducted several surveys between November 2018 and April 2019 for the news Channel BFM TV. Other institutes, such as ODOXA, IFOP or OPINIONWAY also conducted polls, with similar results, as can be seen on Wikipedia page dedicated to the movement (link).

[21] First `blocage17novembre.fr`, then `gilets-jaunes.com` and `giletsjaunes-coordination.fr`.

**Figure 4.8:** Cumulative signature rate of the Change.org petition by département




Notes: This figure shows the evolution of the cumulative number of signatories of Change.org petition which was instrumental to the launch of the movement. We plot one line per département, and highlight in red the first département where the petition took off (Seine-et-Marne). It is in this département that the petition initiator lived and where a local newspaper wrote a first article about the petition on 10/12. The first national newspaper which reported the story did so on 10/21. The vertical dashed line corresponds to 11/17.

### Gilets jaunes – Groups

Using the methodology of (**?**), we compiled a list of the Facebook groups that were still active one month after 11/17 by performing numerous search requests using a set of keywords linked to the movement – see the appendix. For each group, we recorded the name of the group, creation date, number of members, and number of publications. We eventually identified 3,033 groups in total, with over four million members. Over two-thirds of the groups were associated with a geographical area and more than 40% of the total number of members belonged to these localized groups. Moreover, only 20% of the posts emanated from national groups, which suggests that localized groups were the most active type.

### Gilets jaunes – Pages

We also identified 617 Facebook pages and used Netvizz to retrieve their content (**?**): posts, comments and interactions (such as likes and shares).[22] This corpus features over 121,000 posts, 2.1 million comments and 21 million interactions. Netvizz did not provide user ids associated to scraped content. We therefore scraped Facebook a second time in January 2022 and collect additional basic user information. To protect users' privacy, all user ids were anonymized. Approximately 30% of pages had been deleted by January 2022. On the remaining pages, we were able to retrieve 46% of the original posts and 18% of the original comments for this second data retrieval (see Table 4.4). Both datasets appear similar both in terms of their distribution of political preferences and in terms of the topics discussed (see Figures 4.9 and 4.10). They also display qualitatively similar trends in our

22 Netvizz is no longer available since the 21st August, 2019.

three main outcome variables in Section 4.3.3, though the second dataset generally displays larger increases in antagonistic attitudes (see Figure 4.11).

**Table 4.4:** Comparison between the Two Data Collections on Facebook Pages

| Data Collection | Pages | Posts | Comments | Sentences | Users |
|---|---|---|---|---|---|
| First | 617.00 | 120242 | 1936921 | 2860427 | NA |
| Second | 411.00 | 56062 | 352733 | 706182 | 120463 |

**Figure 4.9:** Political Attitudes for Each Data Collection

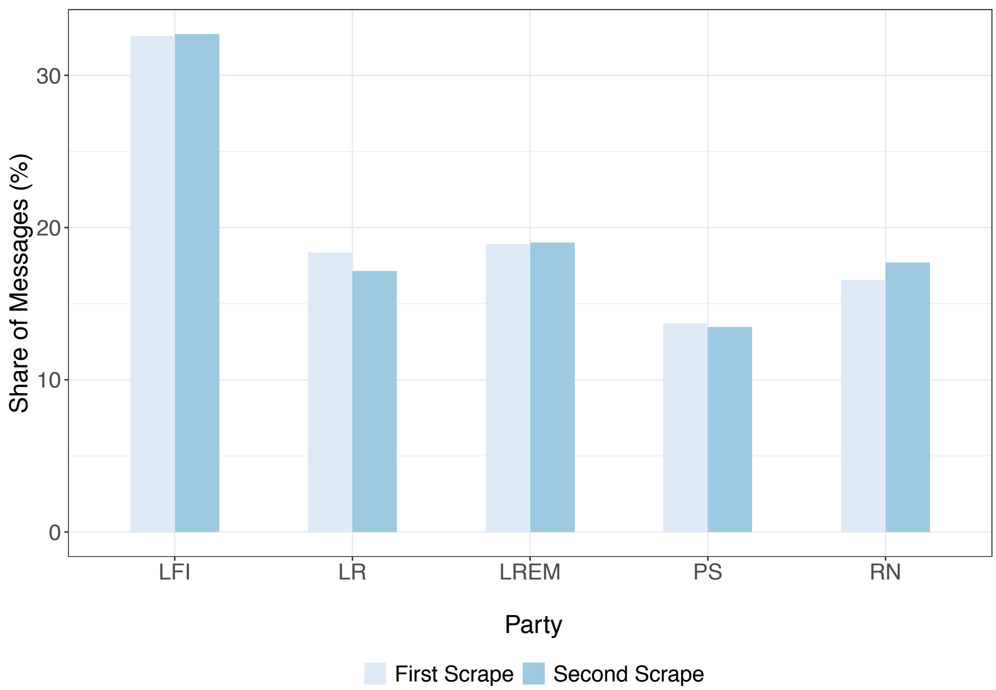


**Figure 4.10:** Topic Shares for Each Data Collection

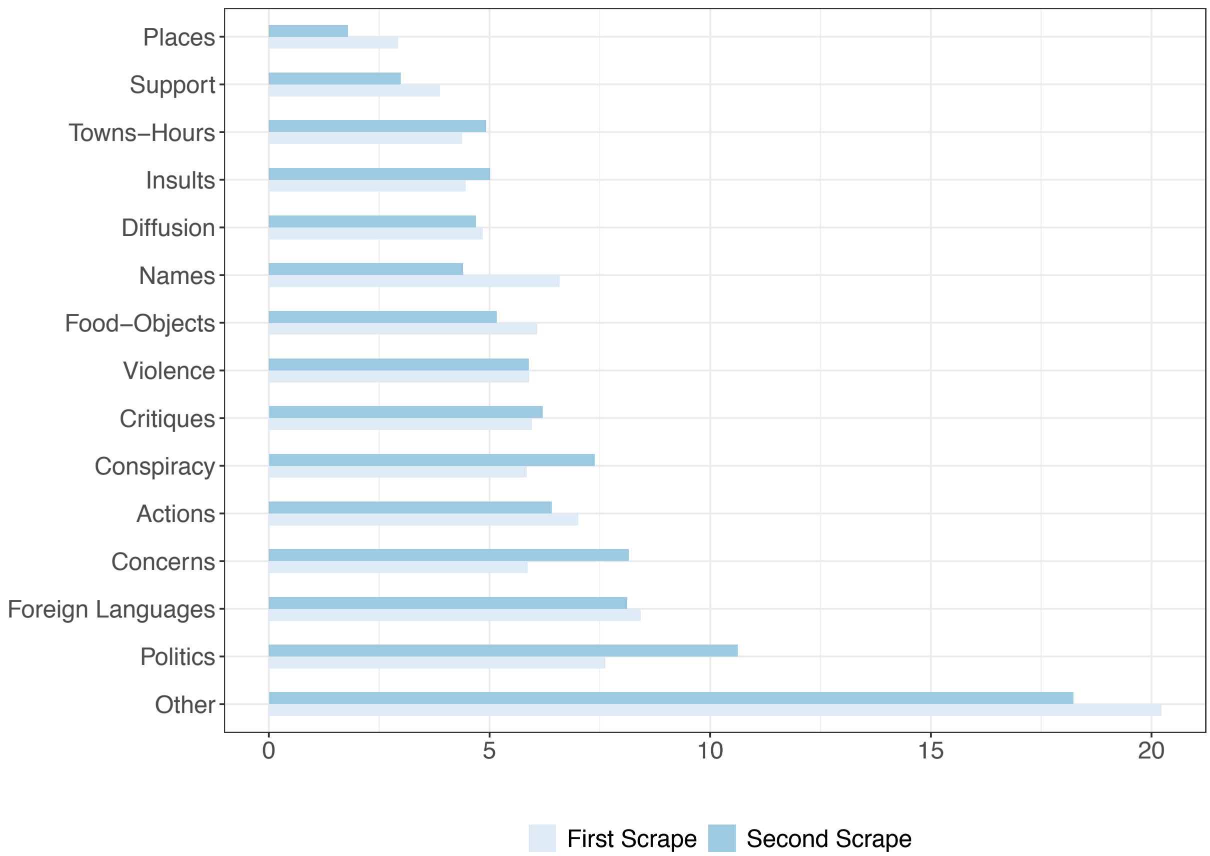


**Figure 4.11:** Evolution of Outcomes for Each Data Collection

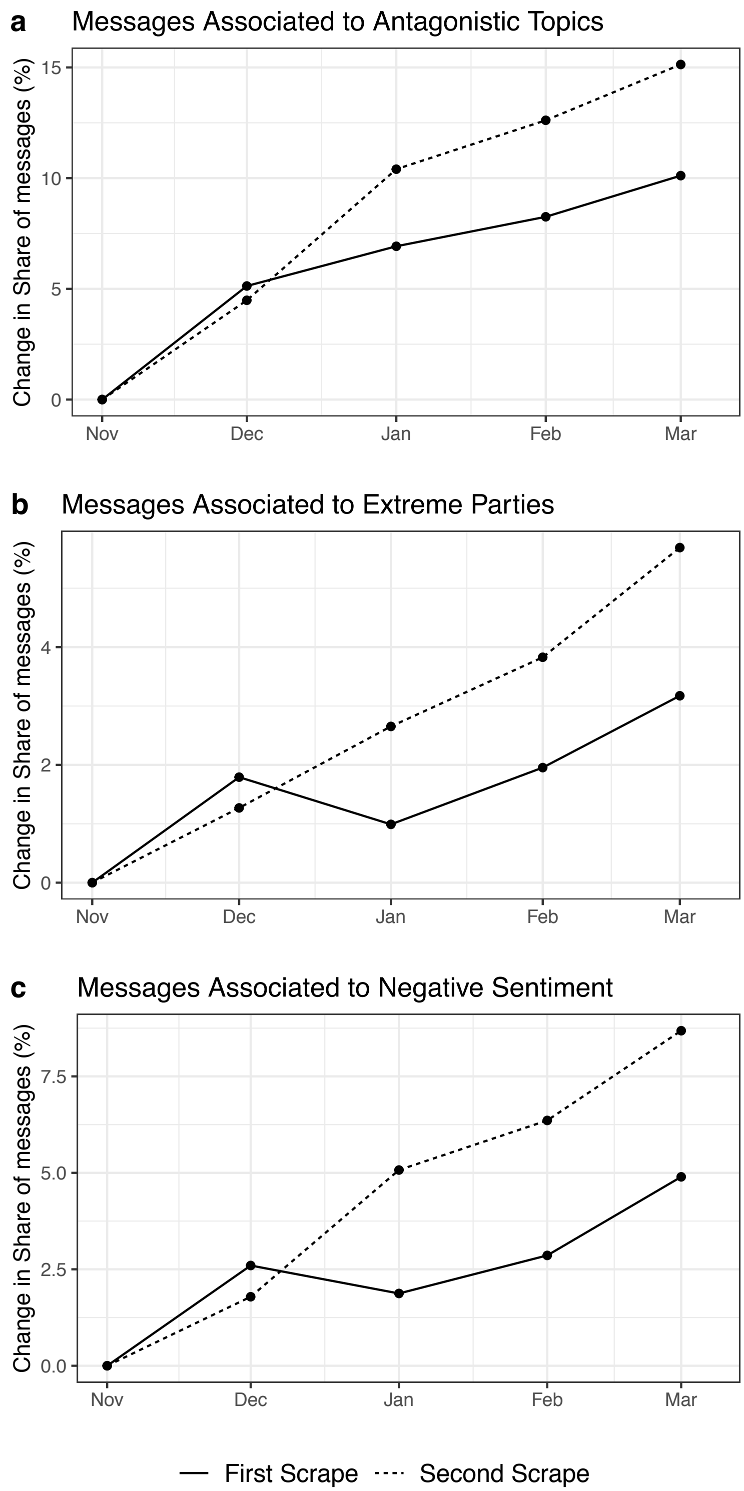


**Figure 4.12:** Kernel Density of Observed Days per User

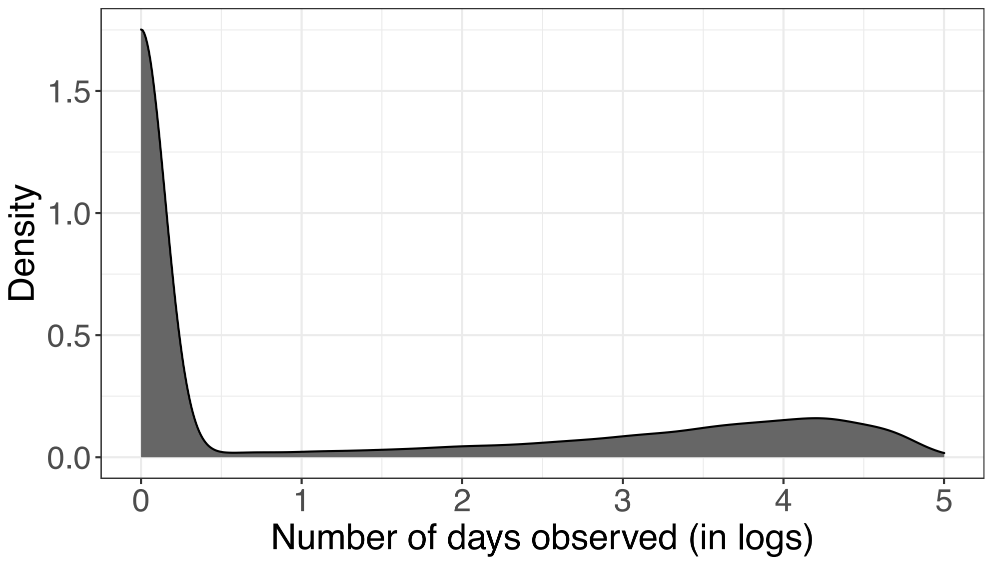
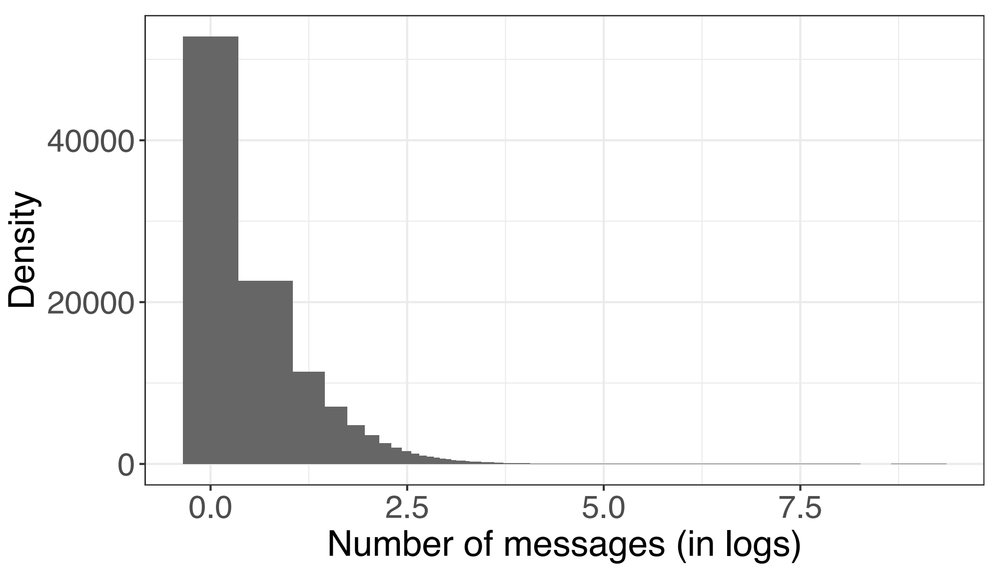



**Figure 4.13:** Histogram Density of Messages Per User

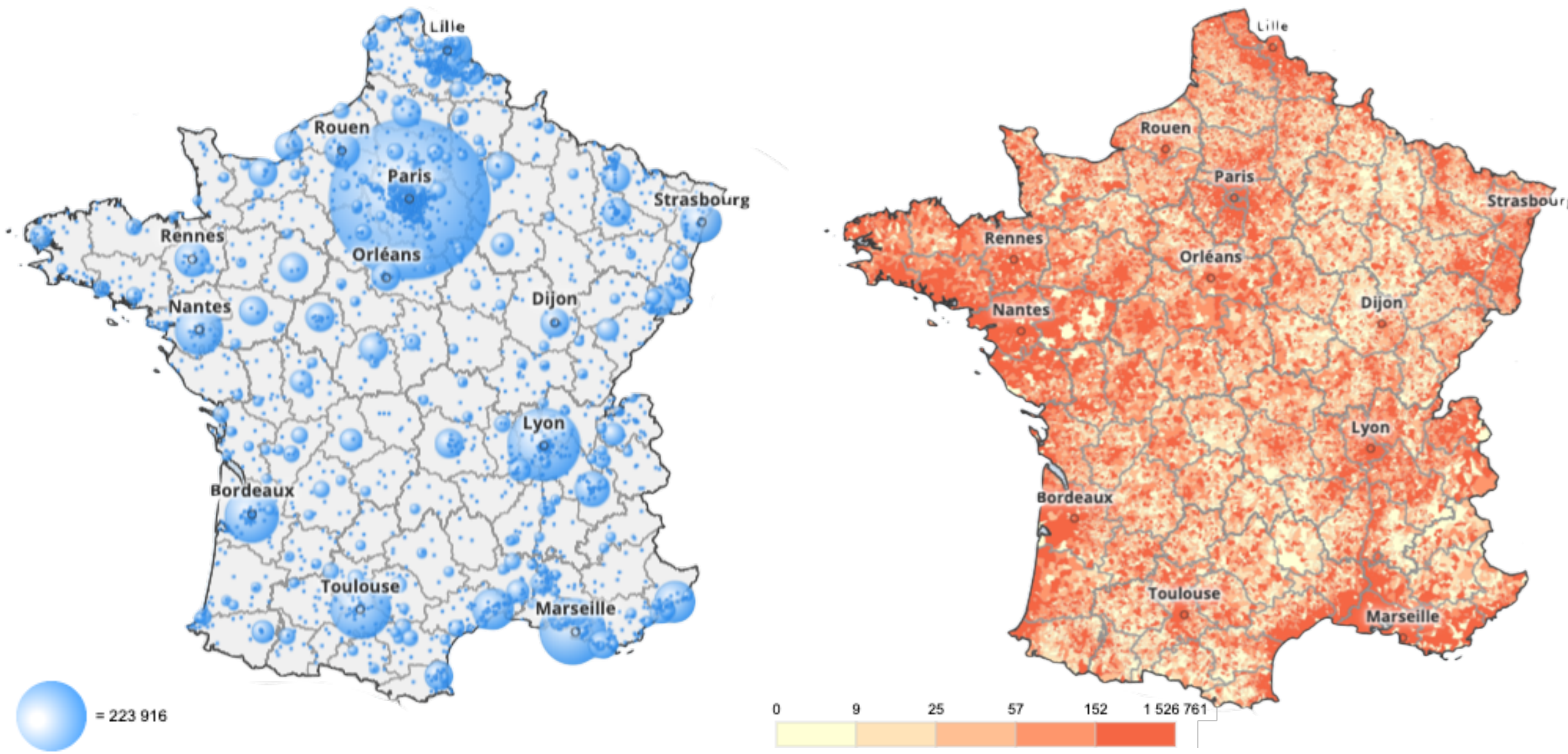


**Facebook Penetration Rates**

To measure Facebook penetration within France, we leverage one of the largest data leaks in the history of Facebook. In 2019, a massive dataset of Facebook users was made publicly available.[23] The dataset contains the name, account creation date, marriage status, self-declared location and phone number of half a billion Facebook users, including 30 million French accounts. The hackers used Facebook ID numbers in ascending order to get to Facebook public profiles. For this reason, this is a priori a close-to-comprehensive dataset of Facebook users within each country before 2019. It is to this date the largest publicly available dataset of Facebook users with location information. This makes it particularly attractive to compute Facebook penetration rates before the onset of the Gilets jaunes protests. We use character-matching techniques to link self-declared locations of users to French administrative data.

[23] *Source:* `https://www.bbc.com/news/technology-56772772`

**Figure 4.14:** Spatial distribution of Facebook users




Notes: These two maps display the spatial distribution of Facebook users (who declare either to live in or to come from the location) in absolute values at the municipality level. The map on the left-hand side displays the values in proportion of circle radius, while the map on the right-hand side displays the value in terms of color intensity.

### 4.A.4 Twitter

We build a dataset of tweets by politicians in the lower chamber of the French Parliament (the Assemblée Nationale). We consider the five largest French political parties: Rassemblement National (RN), Les Républicains (LR), La République en Marche (LREM), le Parti Socialiste (PS) and La France Insoumise (LFI). Politicians use Twitter to directly speak to their constituents. Thus, tweets are closer to daily social media messages than parliamentary speeches. They provide a natural, labeled dataset to train a machine learning classifier of party affiliation based on written text. We then use our classifier to infer online protesters' political partisanship based on their Facebook messages.

The complete list of politicians at the Assemblée Nationale is available here. The dataset of French politicians on Twitter is available here. We retrieve the 3200 last tweets of each politician via the Twitter API.

### 4.A.5 Administrative French Data

We construct a wide set of local controls, partly based on individual-level data.[24] The set of municipal controls included our regressions may be grouped as follows:

- **Geography** includes the population of the municipality, its density, the distance to the closest city with over 20,000 inhabitants and 100,000 inhabitants, whether the municipality was classified as urban in 2015 and whether it switched from rural to urban between 1999 and 2015.
  *Source: Census (RP, complementary exploitation), 2016, INSEE.*
- **Transport** includes the shares of the employed population commuting by car and public transportation, as well as the median commuting distance.
  *Source: Census 2016, INSEE. Déclarations Annuelles de Données Sociales (DADS), 2015, INSEE.*
- **Economy** include the local unemployment rate, the fraction of employees with a non-permanent contract, mean income, and population immigrant share.
  *Source: Census 2016, INSEE. DADS, 2015, INSEE.*
- **Occupation** includes the share of the different catégories socio-professionelles defined by INSEE: executive, independent, middle-management, employee, manual worker and agriculture.
  *Source: Census 2016, INSEE.*
- **Age** includes the shares of the population in the following groups: 18-24 y.o.; 25-39 y.o.; 40-64 y.o.; over 65 y.o.
  *Source: Census 2016, INSEE.*
- **Education** includes the shares of the population without the high-school diploma, and with a university degree.
  *Source: Census 2016, INSEE.*
- **Vote** includes the vote share for the five major candidates in the 2017 presidential election (Macron, Le Pen, Fillon, Mélenchon, Hamon), as well as the share of abstention.
  *Source: Ministry of the Interior.*
- **Signature** is the local signature rate of the Change.org petition before 11/17.
  *Source: Change.org.*
- **LZ** is a set of 1,596 dummies for Life Zones.
  *Source: INSEE.*

[24] Some of the data is available on the INSEE website, but individual-level data is only available through CASD (Centre d'Accès Sécurisé aux Données).

### 4.A.6 Variance Decomposition

**Table 4.5:** Variance decomposition: Yellow Vests movement (pre-17/11)

| Variables | Signatures | | Nb. group | | Local group | | Blockade | |
|---|---|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) |
| $\hat{Y}$ | **11.24** | **49.41** | **36.37** | **39.82** | **11.15** | **16.22** | **18.21** | **23.78** |
| **Fixed-Effects (LZ)** | | **43.51** | | **1.28** | | **2.43** | | **1.60** |
| Density | 1.17 | 0.74 | 22.34 | 24.62 | 5.67 | 1.47 | 0.74 | 5.63 |
| Population | 0.11 | 0.15 | 3.97 | 7.65 | 1.53 | 1.44 | 1.59 | 0.91 |
| Squared Population | 0.29 | 0.23 | 0.79 | 1.36 | 0.45 | 0.33 | 0.86 | 0.41 |
| Pop.spline: median | 0.12 | 0.11 | 0.02 | 0.01 | 0.03 | 0.02 | 0.12 | 0.04 |
| Pop.spline: $75^{th}$percentile | 1.30 | 0.84 | 8.24 | 2.60 | 0.88 | 7.64 | 8.38 | 10.86 |
| **Population measures** | **2.99** | **2.07** | **35.36** | **36.25** | **8.58** | **10.93** | **11.69** | **17.85** |
| Urbanization | 0.15 | 0.32 | 0.23 | 0.54 | 0.80 | 0.86 | 2.25 | 1.41 |
| Urbanized (1999-2015) | 0.01 | 0.04 | 0.01 | 0.02 | 0.03 | 0.03 | 0.04 | 0.03 |
| Min dist. to closest mid-size city | 0.22 | 0.09 | 0.02 | 0.04 | 0.01 | 0.02 | 0.00 | 0.06 |
| Min dist. to large city | 0.25 | 0.25 | 0.01 | 0.06 | 0.01 | 0.03 | 0.02 | 0.02 |
| **Geography** | **0.65** | **0.69** | **0.27** | **0.66** | **0.85** | **0.93** | **2.31** | **1.52** |
| Car commuters | 0.14 | 0.20 | 0.04 | 0.09 | 0.07 | 0.08 | 0.14 | 0.10 |
| Public transp. Commuters | 0.36 | 0.12 | 0.17 | 0.28 | 0.20 | 0.15 | 0.37 | 0.18 |
| Median commuting dist. | 0.13 | 0.02 | 0.00 | 0.00 | 0.01 | 0.01 | 0.06 | 0.03 |
| **Commuting** | **0.63** | **0.34** | **0.21** | **0.36** | **0.28** | **0.24** | **0.57** | **0.31** |
| Mean permanent labor earning | 0.87 | 0.02 | 0.00 | 0.00 | 0.00 | 0.00 | 0.01 | 0.00 |
| Share of non-tenure workers | 0.10 | 0.04 | 0.01 | 0.03 | 0.05 | 0.04 | 0.15 | 0.10 |
| Share of unemployment | 0.12 | 0.03 | 0.11 | 0.23 | 0.32 | 0.33 | 0.61 | 0.33 |
| Share retail workers | 0.01 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.01 | 0.01 |
| Share executives | 0.05 | 0.01 | 0.00 | 0.01 | 0.01 | 0.02 | 0.04 | 0.02 |
| Share intermediates | 0.17 | 0.08 | 0.01 | 0.03 | 0.02 | 0.02 | 0.04 | 0.01 |
| Share clerical workers | 0.11 | 0.06 | 0.00 | 0.01 | 0.01 | 0.01 | 0.05 | 0.00 |
| Share blue collar | 0.09 | 0.05 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| **Labor Market** | **1.51** | **0.29** | **0.15** | **0.31** | **0.41** | **0.42** | **0.91** | **0.49** |
| Share 18 to 24 y.o. | 0.05 | 0.03 | 0.13 | 0.28 | 0.27 | 0.28 | 0.59 | 0.39 |
| Share 25 to 39 y.o. | 0.12 | 0.06 | 0.01 | 0.03 | 0.01 | 0.02 | 0.08 | 0.05 |
| Share 40 to 64 y.o. | 0.02 | 0.01 | 0.01 | 0.02 | 0.08 | 0.05 | 0.03 | 0.03 |
| Share over 65 y.o. | 1.43 | 0.71 | 0.00 | 0.00 | 0.00 | 0.00 | 0.01 | 0.01 |
| **Age groups** | **1.63** | **0.81** | **0.15** | **0.33** | **0.36** | **0.36** | **0.71** | **0.47** |
| Share high-school drop-out | 0.44 | 0.12 | 0.00 | 0.00 | 0.01 | 0.00 | 0.01 | 0.01 |
| Share graduates (post-bac) | 0.15 | 0.09 | 0.00 | 0.02 | 0.00 | 0.00 | 0.01 | 0.01 |
| **Education** | **0.60** | **0.21** | **0.00** | **0.02** | **0.01** | **0.00** | **0.03** | **0.02** |
| **Share of immigrants** | **0.39** | **0.38** | **0.06** | **0.10** | **0.06** | **0.02** | **0.05** | **0.10** |
| EM (center) | 0.03 | 0.06 | 0.00 | 0.01 | 0.01 | 0.01 | 0.04 | 0.05 |
| LFI (far-left) | 0.08 | 0.03 | 0.01 | 0.00 | 0.00 | 0.01 | 0.00 | 0.02 |
| UMP (right) | 0.11 | 0.18 | 0.00 | 0.00 | 0.01 | 0.00 | 0.01 | 0.01 |
| PS (left) | 0.02 | 0.04 | 0.01 | 0.01 | 0.01 | 0.02 | 0.02 | 0.02 |
| FN (far-right) | 2.30 | 0.62 | 0.00 | 0.03 | 0.01 | 0.03 | 0.01 | 0.02 |
| Abstention | 0.01 | 0.00 | 0.05 | 0.13 | 0.17 | 0.20 | 0.62 | 0.38 |
| **Votes 2017** | **2.55** | **0.93** | **0.07** | **0.18** | **0.20** | **0.27** | **0.72** | **0.51** |
| Cars per inhabitant | 0.02 | 0.02 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| Share diesel vehicles | 0.19 | 0.07 | 0.04 | 0.20 | 0.06 | 0.15 | 0.18 | 0.19 |
| Roads length | 0.01 | 0.02 | 0.02 | 0.01 | 0.09 | 0.16 | 0.27 | 0.33 |
| Reduced speed | 0.06 | 0.07 | 0.04 | 0.11 | 0.24 | 0.29 | 0.76 | 0.40 |
| **Motorists** | **0.29** | **0.17** | **0.10** | **0.33** | **0.39** | **0.60** | **1.22** | **0.92** |
| **Facebook Penetration** | **0.00** | **0.00** | **0.00** | **0.01** | **0.02** | **0.02** | **0.00** | **0.00** |

Notes: This table shows a post-OLS total variance decomposition constructed using the method of **?**. Coefficients correspond to the the percent share of the outcome variables (columns) explained by the selected variables (rows). In this decomposition, we use one observation per municipality.

**Table 4.6:** Variance decomposition: Roundabouts

| Variables | Roundabouts (Munic.) | | Roundabouts (LZ) | |
|---|---|---|---|---|
| | (1) | (2) | (3) | (4) |
| $\hat{Y}$ | **45,18** | **51,96** | **46,48** | **98,63** |
| **Fixed-Effects (L.Z.)** | | **2,86** | | **90,80** |
| Density | 18,85 | 18,77 | 20,24 | 3,01 |
| Population | 1,56 | 2,07 | 0,83 | 0,14 |
| Squared Population | 0,03 | 0,05 | 0,00 | 0,10 |
| Pop.spline: median | 0,07 | 0,05 | 0,01 | 0,00 |
| Pop.spline: $75^{th}$percentile | 19,86 | 20,06 | 20,45 | 2,39 |
| **Demography** | **40,37** | **41,01** | **41,53** | **5,66** |
| Urbanization | 2,01 | 3,13 | 0,49 | 0,90 |
| Urbanized (1999-2015) | 0,02 | 0,04 | 0,01 | 0,08 |
| Min dist. to closest mid-size city | 0,30 | 0,83 | 0,80 | 0,14 |
| Min dist. to large city | 0,10 | 0,11 | 0,37 | 0,12 |
| **Geography** | **2,43** | **4,11** | **1,66** | **1,25** |
| Car commuters | 0,03 | 0,05 | 0,01 | 0,01 |
| Public transp. Commuters | 0,24 | 0,46 | 0,63 | 0,05 |
| Median commuting dist. | 0,05 | 0,05 | 0,03 | 0,01 |
| **Commuting** | **0,32** | **0,56** | **0,67** | **0,07** |
| Mean permanent labor earning | 0,02 | 0,10 | 0,52 | 0,05 |
| Share of non-tenure workers | 0,26 | 0,37 | 0,07 | 0,07 |
| Share of unemployment | 0,06 | 0,09 | 0,00 | 0,01 |
| Share retail workers | 0,03 | 0,04 | 0,00 | 0,01 |
| Sahre executives | 0,03 | 0,03 | 0,04 | 0,02 |
| Share intermediates | 0,03 | 0,03 | 0,01 | 0,00 |
| Share clerical workers | 0,05 | 0,05 | 0,01 | 0,01 |
| Share blue collar | 0,01 | 0,01 | 0,01 | 0,00 |
| **Labor Market** | **0,49** | **0,72** | **0,66** | **0,17** |
| Share 18 to 24 y.o. | 0,31 | 0,46 | 0,13 | 0,08 |
| Sahre 25 to 39 y.o. | 0,03 | 0,07 | 0,01 | 0,02 |
| Share 40 to 64 y.o. | 0,10 | 0,12 | 0,01 | 0,01 |
| Share over 65 y.o. | 0,01 | 0,03 | 0,12 | 0,09 |
| **Age groups** | **0,44** | **0,69** | **0,27** | **0,20** |
| Share high-school drop-out | 0,00 | 0,00 | 0,22 | 0,06 |
| Share graduates (post-bac) | 0,00 | 0,08 | 0,42 | 0,03 |
| **Education** | **0,00** | **0,08** | **0,64** | **0,09** |
| **Share of migrants** | **0,16** | **0,49** | **0,12** | **0,03** |
| EM (center) | 0,22 | 0,19 | 0,23 | 0,05 |
| LFI (far-left) | 0,05 | 0,03 | 0,03 | 0,00 |
| UMP (right) | 0,01 | 0,00 | 0,01 | 0,01 |
| PS (left) | 0,03 | 0,03 | 0,02 | 0,01 |
| FN (far-right) | 0,05 | 0,05 | 0,00 | 0,03 |
| Abstention | 0,08 | 0,19 | 0,00 | 0,00 |
| **Votes 2017** | **0,43** | **0,49** | **0,29** | **0,09** |
| Cars per inhabitant | 0,01 | 0,01 | 0,02 | 0,01 |
| Share diesel vehicles | 0,40 | 0,70 | 0,60 | 0,23 |
| Roads length | 0,06 | 0,13 | 0,00 | 0,01 |
| Reduced speed | 0,06 | 0,11 | 0,01 | 0,02 |
| **Motorists** | **0,53** | **0,95** | **0,64** | **0,26** |
| **Facebook Penetration** | **0,00** | **0,00** | **0,00** | **0,00** |

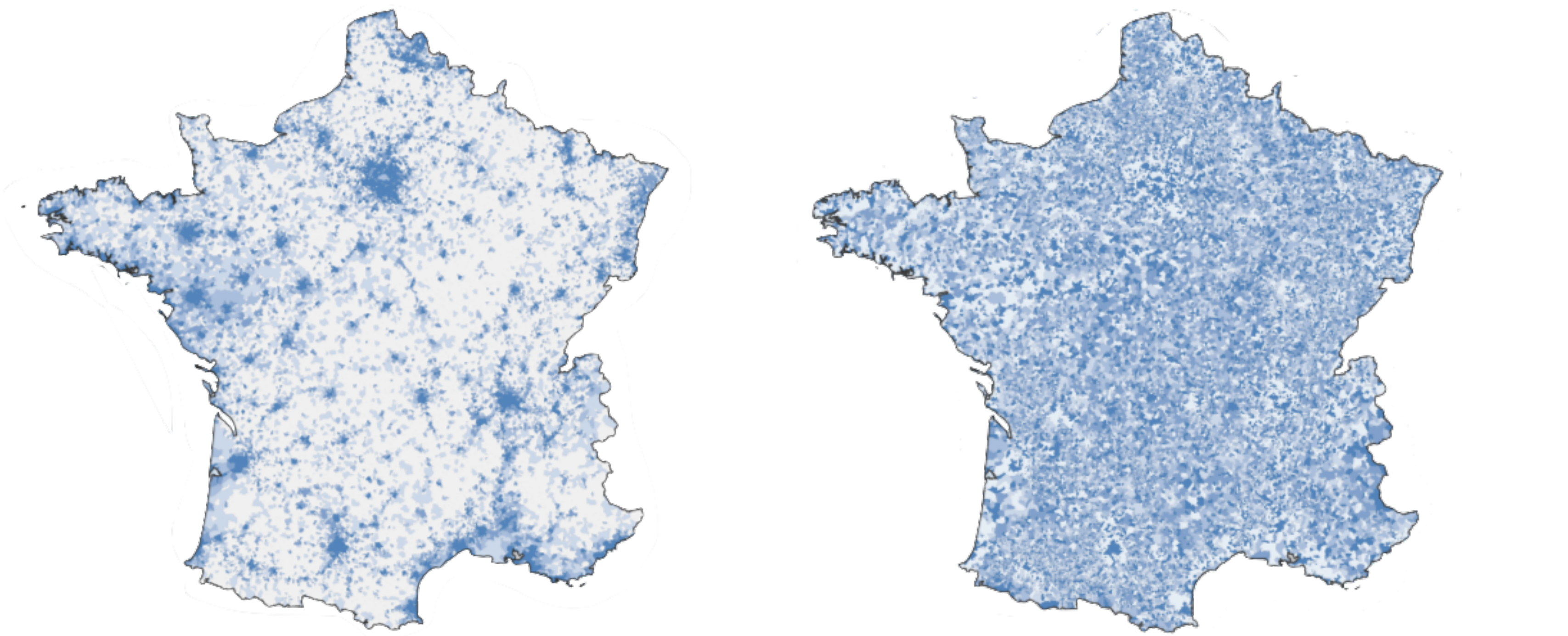
Notes: This table shows a variance decomposition similar to that of Table 4.5, using the two roundabout instruments as outcomes. "Roundabouts (Munic.)" corresponds to the density of roundabouts in a municipality, and "Roundabouts (LZ)" corresponds to the average roundabout density in the other municipalities of the LZ.

## 4.B Offline to Online – Details and Robustness

**Figure 4.15:** Roundabout density

**(a)** Roundabouts by sq. km

**(b)** Residuals



Notes: Panel (a) shows the density of roundabouts in mainland France, with darker colors corresponding to higher density. Panel (b) shows the residual density of roundabouts after controlling for the set of controls describe in Table 4.6. Color intensity corresponds to quantile thresholds.

**Table 4.7:** Impact of blockades on further online mobilization: Alternative specifications

| | Signature Rate (post-17/11) | Local Group (post-17/11) | Nb. Members in Local G. (post-17/11) | Nb. Posts in Local G. (post-17/11) |
|---|---|---|---|---|
| | | Ordinary Least Square | | |
| | | A. Baseline Specification | | |
| | (1) | (2) | (3) | (4) |
| Blockade | 0.078** | 0.164*** | 0.008*** | 0.012*** |
| | (0.035) | (0.016) | (0.002) | (0.003) |
| | | B. Without municipal controls | | |
| | (1) | (2) | (3) | (4) |
| Blockade | 0.434*** | 0.307*** | 0.011*** | 0.014*** |
| | (0.017) | (0.018) | (0.002) | (0.003) |
| | | 2 Stage Least Square | | |
| | | C. Without municipal controls | | |
| | (1) | (2) | (3) | (4) |
| Blockade | 1.631*** | 0.774*** | 0.027*** | 0.030*** |
| | (0.215) | (0.063) | (0.005) | (0.007) |
| Kleibergen-Paap F-stat | 59.433 | 59.433 | 59.433 | 59.433 |
| p-value Hansen | 0.082 | 0.548 | 0.240 | 0.186 |
| | | D. Only municipal instrument | | |
| | (1) | (2) | (3) | (4) |
| Blockade | 1.487*** | 0.721*** | 0.046** | 0.055** |
| | (0.407) | (0.176) | (0.020) | (0.026) |
| Kleibergen-Paap F-stat | 14.669 | 14.669 | 14.669 | 14.669 |
| | | E. Only LZ instrument | | |
| | (1) | (2) | (3) | (4) |
| Local Blockade | 1.135*** | 0.689*** | 0.024** | 0.025* |
| | (0.295) | (0.132) | (0.009) | (0.013) |
| Kleibergen-Paap F-stat | 45.271 | 45.271 | 45.271 | 45.271 |
| | | F. Surroundings defined at the employment area level | | |
| | (1) | (2) | (3) | (4) |
| Blockade | 0.986*** | 0.862*** | 0.042*** | 0.045** |
| | (0.289) | (0.183) | (0.013) | (0.019) |
| Kleibergen-Paap F-stat | 12.050 | 12.050 | 12.050 | 12.050 |
| p-value Hansen | 0.280 | 0.704 | 0.126 | 0.154 |

Notes: This table shows estimates corresponding to variations of the regressions of Table 4.3. In Panel A, we present OLS estimates of the effect of experiencing a blockade on 11/17 on online mobilization after 11/17, using the same set of controls as in Table 4.3. Panel B shows similar estimates, but without controls. Panel C shows results for the 2SLS estimation of Table 4.3 when we do not include a set of controls. In Panel D (resp. E), we show 2SLS results using the roundabout density of the municipality as an instrument (resp., the density of roundabouts in other municipalities of the LZ) only. In Panel F, instead of considering the density of roundabouts in other municipalities of the LZ as a second instrument, we use density of roundabouts in other municipalities of the employment zone. In all regressions, we cluster standard errors at the LZ level (except in Panel F, where we cluster standard errors at the employment zone level). *: $p < 0.01$, **: $p<0.05$, ***: $p<0.1$.

**Table 4.8:** Impact of blockades on further online mobilization: placebo (traffic light)

| | Signatures (post-11/17) (1) | Local Group (post-11/17) (2) | Members in Local Groups (3) | Messages (post-11/17) (4) |
|---|---|---|---|---|
| Local Blockade | 0.951 | 0.763*** | 0.015 | -0.002 |
| | (0.595) | (0.125) | (0.013) | (0.017) |
| Signatures (pre-17/11) | 0.483*** | -0.001 | 0.000 | 0.000 |
| | (0.049) | (0.002) | (0.001) | (0.002) |
| Nb. Groups (pre-17/11) | -0.017 | -0.008 | 0.001 | 0.001 |
| | (0.015) | (0.006) | (0.001) | (0.001) |
| Local Group (pre-17/11) | -0.065 | -0.007 | -0.000 | 0.005 |
| | (0.209) | (0.063) | (0.007) | (0.011) |
| Facebook Penetration | 2.811 | 10.519*** | 5.872** | 5.738* |
| | (3.266) | (2.721) | (2.910) | (2.965) |
| Controls | ✓ | ✓ | ✓ | ✓ |
| Fixed-Effects | L.Z. | L.Z. | L.Z. | L.Z. |
| Observations | 34,451 | 34,451 | 34,451 | 34,451 |
| F-stat | 5.597 | 5.597 | 5.597 | 5.597 |
| p-value Hansen | 0.050 | 0.057 | 0.575 | 0.657 |

Notes: This table shows placebo results of our 2SLS estimates using traffic light to instrument local blockades. The four measures of online mobilization after 11/17 are: the signature rate of the Change.org petition after 11/17 (column 1), the creation of a local group post-11/17 (column 2), the number of members per inhabitant (column 3) and posts per inhabitant (column 4) in these newly created local groups. Signature rates, Number of groups, Number of members and posts per inhabitants are standardized. Standard errors are clustered at the LZ level. *: $p < 0.1$, **: $p < 0.05$, ***: $p < 0.01$.

## 4.C Analysis of Facebook Pages – Details and Robustness

### Text Pre-processing and Featurization

**Text Cleaning Steps** We process all text corpora in the same way. We remove emojis, links, accents, punctuation, social media notifications (e.g., "Gilets jaunes changed their profile picture") and stopwords from the corpus. We also lower-case the text and lemmatize words. We keep hashtags and user mentions, but drop all tokens which occur less than ten times in the Facebook corpus.[25] This leaves us with approximately 40,000 unique tokens in the corpus.

**Unit of Analysis** Most documents in our corpora are short text snippets (e.g., a phrase or a sentence). Some are longer and span over a multiple sentences (e.g., Facebook posts). To keep all documents comparable, we work with unigrams at the sentence level.

### Topic Model

**Rationale** The standard approach for topic modeling in the text as data literature is to rely on Latent Dirichlet Allocation (LDA) models. LDA models documents as a distribution over multiple topics. Though this is often a reasonable assumption, it is implausible in the case of short text snippets (such as sentences) which often refer to only one topic (**?**). For this reason, standard topic models are known to perform poorly on such short texts. As an alternative, we build a custom topic model in the spirit of **?**.

**Implementation** First, we produce word embeddings for the corpus and represent each sentence as a vector in the embedding space. We train a Word2Vec model using Gensim's implementation, with moving windows of eight tokens and ten iterations of training. We build sentence embeddings as the weighted average of the constituent word vectors, where the weights are smoothed inverse term frequencies (to assign higher weights to rare/distinctive words) (**?**). The resulting embedding space allows for a low-dimensional representation of text, in which phrases which appear in similar contexts are located close to one another. Second, we group sentence vectors together into a small set of clusters. The goal is to have different clusters for different topics in the text. We rely on the K-Means algorithm. We train the algorithm on 100,000 randomly drawn sentences and predict clusters for the rest of the corpus. We use the ten closest words to the cluster centroids to manually label topics.[26]

**Alternative Numbers of Topics** The number of topics is a hyperparameter in our topic model. For transparency and robustness, we present resulting topics when specifying 5, 10 and 20 clusters.

[25] The frequency threshold does not influence results, but allows us to remove many uncommon spelling mistakes and other idiosyncrasies related to social media data.

[26] We also considered alternative labeling options, such as term frequency - inverse cluster frequency, which yield similar results.

**Table 4.9:** Results of the Topic Model for Alternative Numbers of Clusters

Panel 1: Results of the Topic Model for 5 clusters

| **Associated words** |
|---|
| 04, nimes, arras, nime, 77, narbonne, albi, chambery, 47, orleans |
| pouvoir, etre, consequent, favoriser, necessaire, n, global, politique, specifique, constitue |
| merde, connard, salopard, pourriture, encule, putain, hont, honte, batard, ordure |
| gabin, live, sympa, app, brancher, stp, ramous, cool, stabilisateur, coupure |
| laziah, misfortune, #noussommesgiletsjaune, dellacherie, exhort, substituons, sansone, pajalo, victory, naeim |

Panel 2: Results of the Topic Model for 10 clusters

| **Associated words** |
|---|
| etre, n, peuple, meme, politique, faiblesse, nefaste, veritable, gouvernement, destructeur |
| annuel, beneficiaire, compenser, bonus, salaire, taxation, production, exoneration, delocalisation, embauche |
| cr, flic, flics, policier, gazer, projectile, charger, manifestant, matraque, gendarme |
| zappe, zapper, tpmp, humoriste, fakenew, interviewe, conversation, cnew, interviewer, bfmtv |
| orlane, magdalena, grilo, correia, gourdon, leal, caudrelier, malaury, macedo, khaye |
| connard, merde, encule, bouffon, conard, pd, salope, enculer, fdp, batard |
| adhesion, charte, valider, definir, modalite, eventuel, prealable, specifique, necessaire, proposer |
| 04, nimes, arras, albi, nime, royan, 77, narbonne, chambery, 47 |
| courage, courag, bravo, felicitation, formidable, bisou, bisous, genial, soutien, continuation |
| sansone, dutie, facilitate, soldiers, auv, weier, unterstutzen, #jiletsjaune, ausbeutung, seem |

Panel 3: Results of the Topic Model for 20 clusters

| **Associated words** |
|---|
| beneficiaire, compenser, salaire, bonus, annuel, exoneration, plafonner, taxation, embauche, reduction |
| omo, #noussommesgiletsjaune, laziah, houpette, nooooon, jeoffrey, chab, limitatif, exhort, cageot |
| aller, faire, voir, la, etre, oui, vraiment, merde, savoir, meme |
| englos, royan, sisteron, pontivy, arras, seclin, hendaye, douai, roanne, albi |
| twitter, diffuse, info, publier, fb, diffuser, relater, page, interview, information |
| adhesion, structuration, proposer, proposition, definir, charte, structurer, concertation, revendication, necessaire |
| maud, johanna, gomes, anai, melanie, gregory, rudy, armand, melissa, mathias |
| bisous, courage, felicitation, courag, bisou, bravo, formidable, soutien, genial, coucou |
| asservissement, domination, peuple, deposseder, destructeur, gouvernance, oppression, politique, veritable, appauvrissement |
| recours, illegal, sanction, infraction, poursuite, condamnation, delit, penal, abusif, commettre |
| 41, 52, 58, 47, 38, 61, 69, 37, 46, 82 |
| canette, chaussette, bouteille, cendrier, plastique, peintur, toilette, saucisson, scotch, brosse |
| cr, flic, flics, frapper, tabasser, matraquer, policier, gazer, matraque, tabasse |
| mafieux, imposteur, larbin, escroc, acolyte, magouilleur, maffieux, corrompu, dictateur, sbire |
| kassav, akiyo, diritti, sempr, dittaturer, etait, popolo, quando, anch, infami |
| stupide, pathetique, affliger, pitoyable, malsain, stupidite, abject, irrespectueux, insultant, grossier |
| 15h, 17h30, 16h30, 10h, 14h00, 11h, gare, 8h30, 18h, 18h30 |
| laziah, #noussommesgiletsjaune, gourdon, misfortune, orlane, grilo, victory, duquesnoy, dellacherie, macedo |
| #jiletsjaune, created, soldiers, #assembleenationale, #coletesamarelo, #parisprotest, dutie, unterstutzen, #france3, sansone |
| connard, encule, batard, salope, fdp, merde, conard, enculer, pd, salopard |

Notes: This table shows the clusters defined by our the topic model when requesting alternative numbers of topics (5, 10, and 20). For each topic, we report the closest words to the cluster centroid (measured by cosine similarity).

### Sentiment Analysis

**Implementation** To measure emotional content in Facebook messages, we use a dictionary-based approach that assigns to a sentence a sentiment score ranging from -1 (very negative) to 1 (very positive). For each sentence, the sentiment score is obtained as the average of the sentiment scores of its constituent words. We rely on the VADER (Valence Aware Dictionary for Sentiment Reasoning) library for our main results. As a robustness check, we rely on French TextBlob as an alternative dictionary for word sentiment.

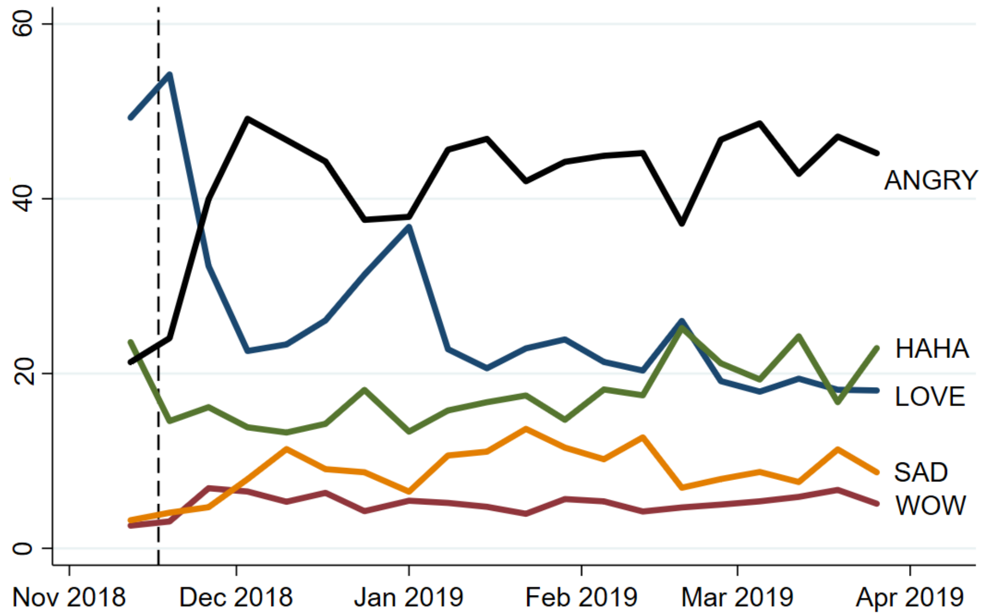
**Examples of Negative and Positive Sentences** We present the five most negative and positive sentences according to our sentiment analyzer.

**Most positive sentences**:
*cela servirait à l aida suivante. On compte sur vous pour cette fabuleuse aventure.*
*cela servirait à l aida suivante. On compte sur vous pour cette fabuleuse aventure.*
*est magnifique !! ...*
*Vous êtes superbes !!!*
*Merci a Thierry pour cette superbe mouette sur le QG*

**Most negative sentences**:
*l horrible*
*et j ai l impression d avoir très mal aux fesses.*
*Il nous prend vraiment pour des cons.*
*«Je suis embarrassé de partager un plateau avec lui. Je le tiens pour un individu très dangereux*
*qu'une bombe peut être très dangereuse. Il y avait encore des gens dans la rue à cette heure. »*

**Alternative Measure of Sentiment** The dictionary-based method has several drawbacks in our context. First, irony (a well-known feature of the French psyche) can lead to poor predictions. The following messages may be classified as positive by the method described above despite being negative: "Making America Great Again gave us everything but good"; "Congratulations to the government, #1 in keeping peaceful demonstrators out of the streets". Second, training sets in French are not as widely available as in English, and they are often extracted from very different contexts (for example, movie reviews). To overcome these problems, we take advantage of the fact that users can react to Facebook posts, using the following reactions: *love, haha, wow, angry, sad*. For each post in our corpus, we compute the weekly share of each of these reactions, displayed in Figure 4.16.

**Figure 4.16:** Sentiment evolution on Facebook pages




Notes: This figure shows the evolution of the weekly shares of different reactions to Facebook posts (in %). The dashed line corresponds to 11/17.

### Political Partisanship Model

**Penalized Multinomial Logistic Regression** Our principal classification method is penalized multinomial logistic regression (**?**). We consider the five largest French political parties: le Rassemblement National (RN), les Républicains (LR), la République en Marche (LREM), le Parti Socialiste (PS) and la France Insoumise (LFI). We parametrize the probability that a text snippet $\mathbf{x}$ is from party $k$ as:

$$\mathrm{P}\big(\text{party} = k \,|\, \mathbf{x}\big) = \frac{\exp(\mathbf{w_k} \cdot \mathbf{x} + b_k)}{\sum_j \exp(\mathbf{w_j} \cdot \mathbf{x} + b_j)}$$

in which $\mathbf{w_k}$ are specific coefficients to be estimated for party $k$. Given the large size of the vocabulary, we further penalize the multinomial logistic regression with the L1-norm (Lasso) to force some coefficients to zero. As some unigrams are not informative of political partisanship, the penalization mitigates over-fitting of the training set by shrinking coefficients.

Note that other supervised models could be used to perform the prediction (e.g., support vector machines, neural networks with BERT, etc.). We work with a logistic regression, as the workings of the model remain interpretable and can be easily validated by the researcher. It is notably straight-forward to assess whether the most predictive unigrams for each party make sense in the French context – see Tables 4.11 and 4.12.

**Model Validation** For model validation, we shuffle the corpus and split it into 80% training data and 20% test data. We build the classifier in the training set and evaluate its performance in the test set. Table 4.10 presents the confusion matrix of the trained classifier. The model has an accuracy score of 55,5%. A random guess would correctly infer the author's party 20% of the time. Our model thus assigns the correct party to a text snippet between two and three times more often than a guess at random would.[27]

**Table 4.10:** Confusion Matrix

| | **RN** | **LFI** | **LR** | **LREM** | **PS** |
|---|---|---|---|---|---|
| **RN** | 1231 | 173 | 227 | 192 | 210 |
| **LFI** | 284 | 934 | 308 | 200 | 241 |
| **LR** | 254 | 277 | 1045 | 241 | 155 |
| **LREM** | 318 | 184 | 300 | 1100 | 140 |
| **PS** | 265 | 197 | 165 | 131 | 1228 |

[27] For comparison, **?** predict party affiliation with an accuracy between 60 and 80% for two parties. In this case, a guess at random would get the label right 50% of the time.

**Table 4.11:** Most Predictive Hashtags and Ats for Each Party

| LFI | PS | LREM | LR | RN |
|---|---|---|---|---|
| @jlmelenchon | @socialistesan | @larem | @republicains | @jpgarraud |
| @franceinsoumise | @borisvallaud | #circo8004 | @lesrepublicains | #mlafrance |
| #montpellier | @faureolivier | @richardferrand | @francoisfillon | @gilbertcollard |
| @mathildepanot | @partisocialiste | @agnesrunacher | @laurentwauquiez | #gard |
| #franceinsoumise | @jlbricout | @mounir | @michelbarnier | @mlp |
| @mbompard | @luccarvounas | @elysee | @fabiendifilippo | @steevebriois |
| @prudhommeloic | @cpiresbeaune | @moreaujb23 | @patrickhetzel | @gwgoldnadel |
| @ugobernalicis | @slefoll | @ephilippepm | @damienabad | @thierrymariani |
| @ericcoquerel | @lamiaela | @enmarchefr | @oserlafrance | @mairieperpignan |
| @fiassemblee | @potierdominique | @emmanuelmacron | @julienaubert84 | #circo6210 |
| @mressiguier | @jjthirson | @terresdeprogres | @chjacob77 | #bethune |
| #generationmelenchon | #laval | #larem | @valerieboyer13 | @ludovicpajot |
| #melenchon2022 | #paris20 | @renaissance | @aulnaysousbois | #marine2017 |
| #ariege | @daviddhabib | @brunobonnelloff | @cestrosi | #legrandjury |
| #rdvactufi | #mayenne | @laustmartin | @jctaugourdeau | @fredericbort |
| @simonnet2 | @guillaumegarot | @sttravert | @ericwoerth | @fn |
| #blanquer | @alaindavid | @gabrielattal | @mairieboisdarcy | @gabirobfrance |
| @manonaubryfr | #devoirdevigilance | @rolandlescure | @anniegenevard | #directhdf |
| #lagauchelavraie | @mkaramanli72 | @yaelbraunpivet | @lesrepublicain | #les4v |
| @lachaudb | #hidalgo2022 | @jblemoyne | @nicolassarkozy | @gillespennelle |
| #aubervilliers | @ericpliez | @oliviagregoire | @alainjuppe | #onarrive |
| #pg | @bcazeneuve | @christophearend | @fxbellamy | #vauvert |
| #loisfi2019 | @zerochomeurld | #rouen | #loiret | @nmeizonnet |
| @aquatennens | @mandinette77 | @bgriveaux | @phdumont | @andrerougeoff |
| @leilachaibi | @ericandrieueu | @encommunasso | @juliendive | @sebastmenard |
| #programmemelenchon | @lonycan | #larepubliqueenmarche | @vpecresse | #francaisreveillezvou |
| #pantin | @avecmarietta | @gdarmanin | @xavierbertrand | #perpignan |
| #loisfi2021 | @espritcivique | @fderugy | @eciotti | @histoiredefran7 |
| #loisfi | @audreypulvar | @nathalieloiseau | @gillescarrez | @vdnbruay |
| #seinesaintdenis | @sartheagauche | @jeunesmacron | @brunobeschizza | #entoutefranchis |
| @melenchon | @lyeslouffok | #renaissance | @villedenice | #pjlterrorisme |
| @alexiscorbiere | @juanico | #formation | @constancelegrip | #mifexpo |
| #enqueteeaufi | @brnomorel | @brunepoirson | #circo6207 | @marion |
| #debatbfmtv | @rglucks1 | @gilleslegendre | #avecjacob | @rnational |
| #esr | @dassouline | #directcirco | #circo4903 | @bleugardlozere |
| @younousomarjee | #sarthe | @stanguerini | @avecvalerie | @sebchenu |
| @carolinefiat54 | @giselebiemouret | #granddebatnational | #lr | #hongrie |
| @deputeratenon | @gabriellesiry | @auroreberge | @yvespdb | @valeurs |
| @gauche | @ithissaintjean | @seblecornu | @guyteissier | #bfmpolitiqu |
| #mun34000 | @fhollande | @barbarapompili | @marinebrenier | @jerome |
| @almokeur | @mxsauvage | @ademontchalin | @auvergnerhalpes | #npdcp |
| @insoumisjeune | @lamontagne | #tousanticovid | @renaudmuselier | @brunobilde |
| @lepg | @hsaulignac | @olivierdussopt | @dmeslot | @helenelaportern |
| @clem | @obiencourt | @enmarche | @dubymuller | #80kmh |
| #stoplawfare | @micheledelaunay | @fionalazaar | @jcgaudin | @brunogollnisch |
| @nveau | @printempspdl | @emmwargon | @phgosselin | #journeemenl |
| #amfis2021 | #gaspillagealimentaire | #masante2022 | #ump | #legrandrdv |
| #melenchon | @lafabriquepdl | #tours | @bernstephane | #plfss2019 |
| #onlacherien | #mayenn | @verdierjouclas | @aurelienpradie | #policier |
| #polatable | @debatretraites | #pplcyberhaine | #ardeche | @franckallisio |

**Table 4.12:** Most Predictive Words for Each Party

| LFI | PS | LREM | LR | RN |
|---|---|---|---|---|
| insoumis | mayenne | denormandie | ardenne | mlp |
| bifurcation | rabault | marcheur | peltier | aliot |
| citoyenne | mans | anglade | dc | marine |
| insoumission | riom | brexit | barnier | bruay |
| youtube | 94 | tempete | ardeche | bardella |
| repression | lamia | tresorerie | nicois | buissiere |
| partagez | socialiste | obstruction | villefranche | compatriote |
| casse | menetrol | 127 | ain | minier |
| rsa | variant | reussir | montargis | islamiste |
| larive | petain | hydrogene | sarrebourg | rachline |
| factieux | garot | perso | kuster | bethune |
| autoritaire | determinee | contenu | marc | clandestin |
| proces | gaspillage | raphan | ardechois | communautariste |
| planification | sarthois | parly | aubenas | ecrite |
| participez | plsr | bouleverser | savignat | communique |
| greviste | tranchee | hurler | gaulliste | islamisme |
| zemmour | alfortville | progressiste | casseur | communiste |
| conserver | rapporter | maurice | fillon | beuvry |
| luc | enfance | pragmatique | elargissement | immigration |
| obono | atlantiques | 2030 | cavaillon | pas |
| salutaire | mourenx | simplification | forissier | confronter |
| ambulancier | qqs | performant | architecte | lievin |
| veolia | go | habitation | gaullisme | elue |
| ressiguier | isf | cler | cope | laxisme |
| neonazi | appui | adoption | deficit | racaille |
| 6e | compensation | innovation | bonheur | monteil |
| back | encommun | recommandation | technocratique | agresser |
| afcult | montee | equilibre | cre | off |
| populaire | vigilance | detail | lorion | henin |
| blog | qd | beneficier | cantal | patriote |
| assistant | revenu | ambitieux | belfort | tva |
| olympe | laval | idf | bioethique | fn |
| duplex | hidalgo | borne | mep | migratoire |
| dignement | tabou | rapprocher | republicains | verlaine |
| lfi | denonce | nou | carrez | guadeloupeen |
| regie | gauche | rapporteure | surprendre | communiquer |
| pamiers | toul | thierry | ump | gauchiste |
| batonnier | nb | approche | medecine | soumission |
| ugo | fleur | infos | annemasse | ue |
| ariege | isabelle | hulot | bretagne | algerie |
| autain | sexe | envie | loiret | fra |
| insoumise | inegalite | absurdie | depense | vardon |
| ministere | assurance | efficacement | principal | bataclan |
| eau | trinque | agglomeration | echec | perpignan |
| riche | allonnes | al | imbecile | ensauvagement |
| meprendre | decalage | bergerac | estimer | ideologie |
| signataire | christopher | sgdb | verbe | rn |
| misere | 63 | rationnel | prelevement | calai |
| deceder | cri | afghan | logique | heros |
| surveillance | tsolidaire | budge | gasoil | ravier |