Three Essays in Empirical Public Economics

by

Elizaveta Pronkina

A dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy in

Economics

Universidad Carlos III de Madrid

Advisor:

Jesús M. Carro

March 2021

Acknowledgments

This dissertation results from several years of work, and I am grateful to a large number of people for their support and help in making this possible.

First and foremost, I would like to express my sincere gratitude to my Ph.D. supervisor, Jesús M. Carro, who provided me with excellent academic guidance while writing this thesis. I highly appreciate and will always remember your suggestions in conducting research and facing challenges. Thank you a lot for your availability, time, and effort in making sure that my thesis was on track during all these years.

I also would like to express my enormous gratitude to two members of the internal committee, Matilde P. Machado and Jan Stuhler, for their invaluable comments during the internal defense and for all assistance and support during the Ph.D. period. The final result would not be the same without your presence.

I am indebted to Juan José Dolado for his insightful suggestions on my work. I immensely value your opinion and will always remember your recommendations.

Next, I would like to extend special thanks to Antonio Cabrales for his guidance and support during the job market year. I also would like to thank Jesús Fernández-Huertas Moraga for always having time to talk to me and discuss any idea. During all these years, I enormously benefited by being surrounded by the great researchers at Universidad Carlos III de Madrid: Julio Cáceres-Delpiano, Raquel Carrasco, Marco Celentani, Alan Crawford, Miguel Á. Delgado, Antonia Díaz, Martin Dumav, Andrés Erosa, Juan Carlos Escanciano, Natalia Fabra, José Luis Ferreira, William Fuchs, Luisa Fuster, Boris Ginzburg, Jesús Gonzalo, Imelda, Belén Jerez, Matthias Kredler, Warn N. Lekfuangfu, Marta C. Lopes, Francisco Marhuenda, Luigi Minale, Diego Moreno, Carmelo Núñez Sanz, Ignacio Ortuño-Ortín, Juan Pablo Rincón-Zapatero, Antonio Romero, Clara Santamaría, Johannes Schneider, Georges Siotis, Gabriel Smagghue, Carlos Velasco, Felix Wellschimied, and Emircan Yurdagul. Being a part of this community has enriched my academic profile.

I would like to acknowledge the work of Arancha Alonso Nieto, Angélica Aparicio, Laura García Llamas and Mercedes Palacios. Your excellent assistance simplified my stay at Universidad Carlos III de Madrid and in Spain during the last six years. Thank you for your numerous emails and time.

I owe thanks to my coauthor, Telmo Pérez-Izquierdo, a great researcher and a wonderful person to work with. I highly appreciate your enthusiasm and work spirit from the very first days of the joint project. Further, I would like to thank two great researchers, who are also my great friends, Sergio Feijoo-Moreira and Rubén Veiga Duarte. Your presence during my Ph.D. and my life in Madrid made this period full of joy and memories. I am also thankful to Fabrizio Leone with whom we shared the first two years of this long journey. I would like to mention Javier Rodríguez, Sebastian Panthöfer, María Bru, Miguel Ángel Cabello, Lorenzo de Masi, Jakob Henning, Weifeng Jin, Antonio Raiola, Henry Redondo, Hasin Yousaf; I was fortunate to be a part of this community.

In the last paragraph, I would like to express my greatest debt to my family. To my parents, Irina and Sergey. I am a lucky daughter to have your unconditional love and encouragement in the course of my life. My sister, Varya, thanks for brightening my world with warmth and happiness. Finally, I would like to thank Michelangelo. Each moment meant a lot to me and will always be.

Published and Submitted Content

The two chapters of this thesis have been published online as working papers. In particular,

1. A previous version of Chapter 1 has been published as:

   Pronkina, E., & Perez Izquierdo, T. J. (2020). USSR, education, work history, fertility choices, and later-life outcomes (No. 30663). Universidad Carlos III de Madrid. Departamento de Economía;

   Available at: https://e-archivo.uc3m.es/handle/10016/30663;

   Pronkina, E., & Perez Izquierdo, T. J. USSR, Education, Work History, Fertility Choices, and Later-Life Outcomes;

   Available at: Personal website;

   The material from these sources included in this thesis is not singled out with typographic means and references.

2. A previous version of Chapter 2 has been published as:

   Pronkina, E. Impact of Employment on Informal Caregiving to the Elderly Mothers in Europe;

   Available at: Personal website;

   The material from this source included in this thesis is not singled out with typographic means and references.

Abstract

In this dissertation, I study the role of policies and institutions to foster social inclusion. In particular, in two of my projects, I use the Survey of Health, Ageing, and Retirement in Europe (SHARE) dataset to analyze the impact of institutions and policies on individuals' decisions. In the first Chapter, I exploit the historical context and study how the Soviet regime changed women's choices within the Soviet sphere regarding educational attainment, labor participation, marriage and fertility. In the second chapter, I consider the role of employment status and the probability to provide informal care to elders in Europe. Finally, in the third part of my doctoral dissertation, I study the role of self-regulation to mitigate the ethnic discrimination on the largest hospitality platform, Airbnb.

The first Chapter, *"USSR, Education, Work History, Fertility Choices, and Later-Life Outcomes" (with Telmo Pérez-Izquierdo)*, investigates the difference in the impact of the Soviet regime on life decisions within the Soviet sphere. We use the retrospective SHARELIFE data to analyze the educational, labor, marriage, and fertility decisions of East Europeans from 1950 to 1990. The main identification strategy is a natural experiment in which we compare former provinces of the Russian Empire in Lithuania and Poland that were exposed to different forms of communism after WWII. For 40 years, Lithuania was a part of the USSR, whereas Poland was a part of the Eastern Bloc. We find that during communism, Lithuanian women worked two years more by age 50 relative to Polish women. This effect is half of the one found for the East-West Germany comparison. Moreover, we observe that women's educational attainment increased more than men's. We propose a potential mechanism behind this fact: an indirect channel of improved work opportunities on female education. Accordingly, this paper's findings highlight the different impacts of the Soviet regime within communist countries.

The second Chapter, *"Impact of Employment on Informal Caregiving to the Elderly Mothers in Europe"*, studies the trade-off faced by adult individuals in Europe between participating in the labor market and providing informal care to their elderly mothers. Using the SHARE data, I develop a bivariate simultaneous choice model of work and informal care. To correct

for the endogeneity of employment status in care decision, I exploit the heterogeneous impact of the Great Recession on European countries as an exclusion restriction in a non-linear setting. When individuals between 50 and below statutory retirement age participate in the labor market, the probability of providing informal care to elder mothers decreases by about nine percentage points. This finding documents the negative causal relationship between employment and the provision of informal care in Europe.

The third Chapter, *"Online Discrimination and (Self) Regulation: Evaluating the Airbnb's Nondiscrimination Policy" (with Michelangelo Rossi)*, is motivated by the following fact digital platforms have changed the ways of doing business in many markets. Still, some characteristics of the transactions occurring online remain unaltered relative to the traditional off-line settings: discrimination of minorities is one of them. Without clear legislative frameworks, in recent years platforms tried to reduce these issues with self-regulations. In this paper, we study the Airbnb's Nondiscrimination policy implemented at the end of 2016. The share of hosts who cannot reject - and potentially discriminate - guests more than doubled after two years from the policy. Accordingly, the number of guests with non-white names on the platform slightly increased. Yet, the proportion of guests with non-white names accepted by hosts who can discriminate guests did not significantly change after the policy.

Contents

1 USSR, Education, Work History, Fertility Choices, and Later-Life Outcomes (with Telmo Pérez-Izquierdo) 1

1.1 Introduction . . . 1

1.2 History of Lithuania and Poland and Soviet Ideology . . . 6

1.2.1 Brief History of Lithuania and Poland . . . 7

1.2.2 Evidence on Similarity Between Lithuania and Poland . . . 9

1.2.3 Enforcement of the Soviet Regime in Lithuania and Poland . . . 15

1.2.4 Summary of the Section . . . 17

1.3 Data and Descriptive Findings . . . 17

1.4 Methodology . . . 22

1.4.1 Benchmark Specification . . . 22

1.4.2 Heterogeneous Impact of the USSR . . . 24

1.4.3 Intensive Margin of the Impact of the USSR . . . 26

1.5 Results . . . 26

1.5.1 USSR, Education and Work History . . . 27

1.5.2 Model of Labor and Schooling Decision in the USSR . . . 29

1.5.3 Heterogeneity with Education . . . 31

1.5.4 Intensive Margin of the USSR Impact . . . 32

1.5.5 USSR, Marriage Choices, Reproductive History and Later-Life Outcomes 33

1.6 Threats for Identification . . . 36

1.7 Robustness Checks . . . 39

1.7.1 Movements During Life . . . 39

1.7.2 Interwar Borders . . . 40

1.7.3 The Soviet Union and the Eastern Bloc . . . 40

1.7.4 Placebo Analysis . . . 40

1.8 East and West Germany Comparison . . . 41

1.9 Conclusion . . . 43
1.10 Appendix: Additional Evidence on Similarity Between Lithuania and Poland . . . 45
1.10.1 Map of the Russian Empire in 1897 . . . 45
1.10.2 Relevant Districts of the Russian Empire in 1897 . . . 46
1.10.3 Demographic and Labor Information in 1897 . . . 47
1.10.4 Labor Information in 1897 . . . 49
1.10.5 Factories and Industries in 1908 . . . 50
1.10.6 Interwar Statistics: Lithuania . . . 52
1.10.7 Interwar Statistics: Poland . . . 53
1.10.8 Geographical Characteristics . . . 54
1.10.9 Soviet Ideology: Propaganda . . . 55
1.10.10 Soviet Regimes in Lithuania and Poland from 1940 to 1990 . . . 56
1.11 Appendix: SHARE Data . . . 57
1.11.1 Year of Birth . . . 57
1.11.2 Area Under the Analysis . . . 58
1.11.3 Educational Attainment . . . 59
1.11.4 Geographical Mobility . . . 60
1.11.5 Descriptive Statistics . . . 61
1.12 Appendix: Average Marginal Impact and Heterogeneity . . . 62
1.12.1 Inconsistent Estimation of the AMI in Presence of Heterogeneity . . . 62
1.12.2 Estimator of the AMI in the Presence of Heterogeneity . . . 62
1.12.3 Intensive Margin of the USSR's Impact . . . 63
1.13 Appendix: Results . . . 64
1.13.1 Years of Education . . . 64
1.13.2 Unweighted Sample . . . 65
1.13.3 Controlling for Region Fixed Effects . . . 67
1.13.4 Mechanism of Labor and Schooling Decisions in the USSR . . . 68
1.14 Appendix: Heterogeneity with Education . . . 69
1.14.1 Three Education Groups . . . 69
1.14.2 Seven Education Groups . . . 70
1.14.3 On Education Across the Place of Birth . . . 70
1.15 Appendix: Intensive Margin of the Impact of the USSR . . . 71
1.15.1 On Experience from 25 to 50 Across Birth Cohorts . . . 71
1.15.2 On Marrying Across the Year of Birth . . . 72
1.15.3 On the Number of Children Across the Year of Birth . . . 73

1.16 Economic Conditions During the Transition Period 74
1.17 Appendix: Threats for Identification 75
1.17.1 Impact of WWII 75
1.17.2 Demographic Characteristics: Sex Ratio 76
1.17.3 Demographic Characteristics: Life Expectancy at Birth 77
1.17.4 Out-Migration from Lithuania and Poland 78
1.18 Appendix: Robustness Checks 79
1.18.1 Movements During Life 79
1.18.2 Interwar Borders 80
1.18.3 The USSR and the Eastern Bloc 82
1.19 Appendix: Placebo Test 84
1.20 Appendix: East and West Germany 85
1.20.1 Fertility During Separation 85
1.20.2 Results 86

2 Impact of Employment on Informal Caregiving to the Elderly Mothers in Europe 87

2.1 Introduction 87
2.2 Data and Descriptive Findings 90
2.2.1 SHARE Data 90
2.2.2 Measure of the Great Recession 95
2.3 Empirical Strategy 96
2.3.1 Identification 97
2.3.2 Potential Omitted Variables Bias 98
2.3.3 Sample Selection in Dynamic Specification 101
2.4 Results 102
2.4.1 Main Findings 102
2.4.2 Heterogeneity Across Genders 105
2.5 Robustness Checks 106
2.6 Conclusion 106
2.7 Appendix: SHARE Data and External Statistics 108
2.7.1 Statutory Retirement Age 108
2.7.2 Frequency of Care Across Country Groups 109
2.7.3 Work Experience Before 2005 Across Genders 110
2.7.4 Measure of the Great Recession 110

2.8 Appendix: Results . . . 111
2.8.1 Change in Mother's Residential Proximity . . . 111
2.8.2 Full Specification . . . 112
2.9 Appendix: Robustness Checks . . . 113
2.9.1 Formal Care Controls . . . 113
2.9.2 Controls for Formal Care . . . 114
2.9.3 Sample of Family Respondents and All Respondents . . . 115
2.10 Appendix: Sample Selection in the Panel Analysis . . . 116

3 Online Discrimination and (Self) Regulation: Evaluating the Airbnb's Nondiscrimination Policy (with Michelangelo Rossi) 118
3.1 Introduction . . . 118
3.2 The Airbnb's Nondiscrimination Policy . . . 120
3.3 Data Description . . . 122
3.4 Identification Strategy . . . 126
3.4.1 Difference-in-Differences Design . . . 126
3.4.2 Event Study Set-Up . . . 129
3.5 Results . . . 131
3.5.1 Main Findings . . . 131
3.5.2 Impact on a Per-Night Price . . . 132
3.5.3 Other Treated Groups . . . 138
3.6 Robustness Checks . . . 140
3.6.1 Minority Groups . . . 140
3.7 Conclusion . . . 145
3.8 Appendix: Data Description . . . 146
3.8.1 Host's Ethnicity Across Cities . . . 146
3.8.2 Share of Non-White Hosts and Guests Across Cities . . . 147
3.8.3 2010 Census Tract Boundaries and Percent of the White Population Across Cities . . . 149
3.9 Appendix: Heterogeneity Across the Income of Neighborhoods . . . 150
3.10 Appendix: Extensions . . . 151
3.10.1 Weighted Analysis . . . 151
3.10.2 Weighted Sample . . . 153
3.10.3 Gender Discrimination . . . 155
3.10.4 Black Lives Matter and Trump Popularity . . . 157

List of Tables

1.1 Balancing Test . . . 12
1.2 Ethnic Structure in Two Countries during the Interwar Period . . . 14
1.3 Ethnic Structure in Two Countries after WWII . . . 14
1.4 USSR, Education and Work Experience . . . 28
1.5 USSR and Marriage History, Fertility and Later-Life Well-Being . . . 35
1.6 Communism, Education and Work Experience in East and West Germany . . . 42
A.1 Demographic and Labor Information in 1897 . . . 47
A.2 Demographic and Labor Information in 1897 (II) . . . 48
A.3 Labor Information in 1897 . . . 49
A.4 Factories' Statistics in 1894 and 1908 . . . 50
A.5 Industrial Composition in 1908 . . . 51
A.6 Geographical Characteristics of Lithuania and Poland . . . 54
A.7 Distinction Between Soviet's Influence in Lithuania and Poland from 1940 to 1990 56
A.8 Descriptive Statistics . . . 61
A.9 USSR, Education and Work Experience . . . 64
A.10 USSR, Education and Work Experience . . . 65
A.11 USSR and Marriage History, Children and Later-Life Well-Being . . . 66
A.12 USSR, Education and Work Experience Controlling for Region Fixed Effects . . . 67
A.13 USSR and Marriage History, Children and Later-Life Well-Being Controlling for Region Fixed Effects . . . 67
A.14 Impact of WWII . . . 75
A.15 Profile of Migrants from Lithuania and Poland . . . 78
A.16 USSR, Education and Work Experience and Movements During Life . . . 79
A.17 USSR, Education and Work Experience . . . 80
A.18 USSR and Marriage History, Children and Later-Life Well-Being . . . 81
A.19 USSR, Education and Work Experience in the USSR and the Eastern Bloc . . . 82

A.20 USSR and Marriage History, Children, Later-Life Well-Being in the USSR and the Eastern Bloc . . . 83
A.21 Placebo Test Similar to Table E.30 in Lippmann, Georgieff, and Senik (2020) . . 84
A.22 Communism and Marriage History, Children, Later-Life Well-Being in East and West Germany . . . 86

2.1 Care Choice Across Genders . . . 92
2.2 Care Intensity Across Genders . . . 93
2.3 Descriptive Statistics of Respondents . . . 94
2.4 Great Recession Across European Countries . . . 96
2.5 Impact of the Crisis on Controls . . . 100
2.6 Maximum Likelihood Estimation of Work and Care Choices . . . 104
2.7 Maximum Likelihood Estimation of Work and Care Choices Across Genders . . 105
B.1 Statutory Retirement Age Across Genders and Countries in 2014 . . . 108
B.2 Care Intensity Across Country Groups . . . 109
B.3 Change in the Transition of the Mother's Residential Proximity before and after the Crisis . . . 111
B.4 Maximum Likelihood Estimation of Work and Care choices . . . 112
B.5 Maximum Likelihood Estimation of Work and Care Choices with Additional Controls . . . 114
B.6 Maximum Likelihood Estimation of Work and Care Choices . . . 115
B.7 Descriptive Statistics of Respondents Across Cross-section and Dynamic Specifications . . . 116
B.8 Selection into the Panel Analysis . . . 117

3.1 Descriptive Statistics of Control and Treated Groups . . . 130
3.2 Difference-in-Differences Set-Up . . . 133
3.3 Difference-in-Differences Set-Pp: Impact on Per-Night Prices . . . 137
C.1 Difference-in-Differences Set-Up Across Neighborhoods . . . 150
C.2 Difference-in-Differences Set-Up Using the Logarithmic Weights . . . 153
C.3 Difference-in-Differences Set-Up Using Optimal Weights . . . 154
C.4 Difference-in-Differences Set-Up Controlling for Black Lives Matter and Trump Popularity . . . 158

List of Figures

1.1 Modern Poland and Lithuania Borders and Selected Provinces of the Russian Empire . . . 10

1.2 Work Experience Across Europe between 1950 and 1990 . . . 21

1.3 AMI of the USSR on Experience by Age 50 Across Education by Gender . . . 32

1.4 AMI of the USSR on Education Across Birth Cohorts by Gender . . . 33

1.5 AMI of the USSR on Experience by 50 Across Birth Cohorts by Gender . . . 34

A.1 Present-Day Lithuania and Poland and the Provinces of the Russian Empire in 1897 . . . 45

A.2 Illustration of the Identification Assumption . . . 46

A.3 Ethnic Structure of the Present-Day Republic of Lithuania in the Interwar Period 52

A.4 Ethnic Structure of Polish Provinces in 1931 by Declared Language . . . 53

A.5 USSR's Propaganda Targeted Women . . . 55

A.6 The Empirical Cumulative Distribution of the Year of Birth of the SHARE Respondents . . . 57

A.7 Regions in the Analysis in Lithuania and Poland . . . 58

A.8 Educational Attainment . . . 59

A.9 Place of Residence During the Life . . . 60

A.10 Maximum Attainable Work Experience by Age 50 Before 1990 . . . 63

A.11 USSR, Education and Work Experience . . . 68

A.12 AMI of the USSR on Experience from 25 to 50 Across Three Education Groups by Gender . . . 69

A.13 AMI of the USSR on Experience by Age 50 Across Seven Education Groups by Gender . . . 70

A.14 AMI of the USSR on Education Across the Place of Birth by Gender . . . 70

A.15 AMI of the USSR on Experience from 25 to 50 Across Birth Cohorts by Gender 71

A.16 AMI of the USSR on Marrying Across Birth Cohorts by Gender . . . 72

A.17 AMI of the USSR on the Number of Children Across Birth Cohorts by Gender . 73

A.18 GDP Per Capita in Lithuania and Poland During the Transition Period . . . 74

A.19 Sex Ratio Male/Female 76
A.20 Life Expectancy at Birth 77
A.21 Fertility Patterns in Germany 85

B.1 Years of Work Experience by 2005 110
B.2 Impact of the Great Recession on Men in Austria 110

3.1 Booking Process on Airbnb 122
3.2 Ethnicity of Users on Airbnb 124
3.3 Change in Users' Ethnicity and Instant Booking Option on Airbnb 125
3.4 Parallel Trend Assumption 128
3.5 Impact of Airbnb's Nondiscrimination Policy 134
3.6 Impact of Airbnb's Nondiscrimination Policy Across Cities 135
3.7 Parallel Trend in Per-Night Prices 136
3.8 Impact of Airbnb's Nondiscrimination Policy on Per-Night Prices 138
3.9 Impact of Nondiscrimination Policy using Different Treated Groups 139
3.10 Impact of Nondiscrimination Policy on Each Minority Group 141
3.11 Impact of Nondiscrimination Policy Across Hosts 142
3.12 Impact of Nondiscrimination Policy Across Neighborhoods 144
C.1 Host's Ethnicity Across Cities 146
C.2 Number and Share of Non-White Hosts Across Cities 147
C.3 Number and Share of Non-White Guests Across Cities 148
C.4 2010 Census Tract Boundaries and Percent of the White Population Across Cities 149
C.5 Empirical Distribution of the Number of Guests Per Month Per Listing 151
C.6 Distribution of Weights and Optimal Weights Across the Number of Guests Per Month 152
C.7 Gender of Users on Airbnb 155
C.8 Impact of Nondiscrimination Policy on Female Guests 156
C.9 "Black Lives Matter" and "Trump" Weekly Popularity in the USA in 2016 . . . 158

Chapter 1

USSR, Education, Work History, Fertility Choices, and Later-Life Outcomes (with Telmo Pérez-Izquierdo)

1.1 Introduction

> The Communist camp is neither homogeneous, monolithic, nor unchanging.
>
> *Zbigniew K. Brzezinski* in the Preface to *"Soviet Bloc: Unity and Conflict"* (1967)

In the 20th century, communist institutions appeared in several forms in Eastern Europe. According to its philosophy, women have the same equal rights as men in economic, political, and family life. In contrast, the male bread-winner family structure was prevalent in capitalist societies. In recent years, economists evaluated the impact of communist regimes on individuals' decisions (among others, Alesina and Fuchs-Schündeln, 2007), focusing on a communist versus non-communist comparison. However, communist countries were not homogeneous and experienced different forms of the Soviet regime.

In this paper, we are interested in the impact of the different forms of communism within the Soviet sphere. Indeed, communist countries were not homogeneous. Communist institutions greatly differ from the Union of Soviet Socialist Republics (USSR) to the Eastern Bloc (Brzezinski, 1967). Specifically, we study the effects of different forms of communism on women's schooling decisions, labor participation, and fertility choices during the regime to abstract from the transition period.

Both individual decisions and political regimes are endogenous to country-specific factors, making it challenging to quantify the impact of the different shades of the Soviet regime from other effects. To overcome this problem and get a causal estimate, we restrict our analysis to the former territories of the Russian Empire in Lithuania and Poland. These two countries show a similar history and patterns before the start of the Second World War. Yet, since the Molotov–Ribbentrop Pact (1939), Lithuania became one of the republics of the USSR and got exposed to the same central government as the rest of the USSR. On the other hand, after World War II (WWII), the Polish National Government continued to exist and formed the Polish People's Republic (1947 – 1989). We exploit this divergence as a natural experiment to shed light on the impact of different forms of communist regimes.

This study uses the Survey of Health, Ageing and Retirement in Europe (SHARE) (2017), retrospective SHARELIFE data (2017), and the SHARE Job Episode Panel.[1] For the first time, this survey covers the life history of respondents in all European countries. The SHARELIFE data allow tracking the education, work, marriage, fertility, and residential history over lifetimes. We focus on respondents born between 1935 and 1958 and consider their choices between 1950 and 1990 to abstract from the confounding factors during the transition period.

Our identification strategy is a natural experiment. We assume that Lithuania became part of the USSR and not of the Eastern Bloc due to exogenous factors unrelated to outcome variables relevant to this study. In particular, we compare former provinces of the Russian Empire in Lithuania and Poland that were similar before communism. The main treatment variable is meant to capture the differential implementation of the Soviet regime. On the one hand, in both countries, women were encouraged to participate in the labor market indepen-

[1]This paper uses data from SHARE Waves 1, 2, 3, 4, 5, 6 and 7 (DOIs: 10.6103/SHARE.w1.710, 10.6103/SHARE.w2.710, 10.6103/SHARE.w3.710, 10.6103/SHARE.w4.710, 10.6103/SHARE.w5.710, 10.6103/SHARE.w6.710, 10.6103/SHARE.w7.710), see Börsch-Supan et al. (2013) for methodological details.(1) The SHARE data collection has been funded by the European Commission through FP5 (QLK6-CT-2001-00360), FP6 (SHARE-I3: RII-CT-2006-062193, COMPARE: CIT5-CT-2005-028857, SHARELIFE: CIT4-CT-2006-028812), FP7 (SHARE-PREP: GA N°211909, SHARE-LEAP: GA N°227822, SHARE M4: GA N°261982) and Horizon 2020 (SHARE-DEV3: GA N°676536, SERISS: GA N°654221) and by DG Employment, Social Affairs & Inclusion. Additional funding from the German Ministry of Education and Research, the Max Planck Society for the Advancement of Science, the U.S. National Institute on Aging (U01_AG09740-13S2, P01_AG005842, P01_AG08291, P30_AG12815, R21_AG025169, Y1-AG-4553-01, IAG_BSR06-11, OGHA_04-064, HHSN271201300071C) and from various national funding sources is gratefully acknowledged(see www.share-project.org).

dently of marital status and the presence of a child, have access to public services to ease the family burden and have an abortion. On the other hand, the Soviet authorities were able to achieve only partial completion of the regime in Poland (see Brzezinski, 1967). Here we document at least two dimensions that were critically different between the two countries: religion and property rights. The weaker enforcement of communism in Poland stems from eased pressure on the Roman Catholic Church that supports less gender egalitarian norms. The other difference between the Soviet regime intensity in Lithuania and Poland is due to the degree of land collectivization. Poland was an important exception among all Eastern Bloc countries because nationwide land reforms dramatically failed (see Brzezinski, 1967; White, Batt, and Lewis, 1993).

The main threats to identification arise from potential differences between Lithuania and Poland in a pre-communist era. We address this concern in two ways. First, we restrict the analysis to Lithuanian and Polish regions that were former territories of the Russian Empire from 1895 to 1918 to isolate the potential persistence in imperial legacies. It is essential to consider only selected areas due to the documented long-lasting impact of the division of Poland among three Empires in the 19th century on the contemporary variables (e.g., Grosfeld and Zhuravskaya, 2015, for political outcomes; Bukowski, 2019, for educational choices).

Next, we document the similarity of these regions using historical statistics before the Soviet regime to argue our identification's plausibility. According to the Russian Imperial Census from 1897 (Grosfeld, Rodnyansky, and Zhuravskaya, 2013; Markevich and Zhuravskaya, 2018), the former territories of the Russian Empire that currently belong to Lithuania and Poland were similar in terms of religious composition, the share of women in the society and the type of jobs held by men and by women. Moreover, we study the Imperial Russian Factory Database by Gregg (2020) that systematizes the Imperial Russian censuses of manufacturers from 1894, 1900, and 1908. Overall, there is no evidence of a significant difference in industrial development, as machine power per worker, on average, was similar in Lithuania and Poland.

Our findings show that being part of the USSR increased the educational attainment for all residents. Yet, the total impact on the education of women was three times larger than on men. Moreover, cumulative work experience by age 50 increased almost by 2 years among women in the USSR, controlling for schooling, and did not change among men. We refer to working status only before the regime's fall in 1990, not including the transition period. The impact of the USSR gets larger when we consider low-educated women. Accordingly, we account for heterogeneity when estimating the impact on experience. Regarding the intensive margin of the USSR's impact on labor participation, we find that women from early birth cohorts (1935

- 1940 and 1941 - 1946) work more. This is in line with the idea that the treatment intensity is the largest for early-born individuals.

We propose a simple educational choice model to explain the USSR's stronger impact on women's education and work experience relative to men. According to the model, the USSR's policies to favor women's employment could have caused an additional positive effect on education. We argue that the enforced Soviet ideology in USSR jeopardized the possibility for women to stay outside the labor market. Thus, women had more incentives to study with the prospect of more favorable working conditions. This policy complementarity was weaker in the Eastern Bloc.

We show that the USSR gave women higher incentives to study; however, the direct impact on marriage and fertility history is also of interest as it often happens simultaneously with labor choices (Chiappori, 2015). We do not analyze any particular policy for consistency of representation, but rather focus on the USSR's overall impact. Conditional on being married at least once during a lifetime, we find that the number of marriages increased by 0.10. Using our identification strategy, we cannot establish the causal relationship between work and marriage history, but we document the correlation between labor participation and a higher probability of remarrying.

In terms of fertility outcomes, we document evidence for lower fertility. Living in the USSR resulted in a statistically significant decrease by 0.18 in the number of children. In our study, we restrict to children who were born before 1990 to abstract from the transition period after the end of the regime. Using East and West Germany comparison, we do not find any drop in the number of children among women born from 1935 to 1958.

Our paper relates to several strands of literature. First of all, we contribute to the literature about the direct impact of communism on residents. The broad comparison between a communism and a non-communism past is the first order Soviet regime's overall effect (see Fuchs-Schündeln and Schündeln, 2020). However, this comparison can be extended further to distinguish shades of communism. Up to our knowledge, this paper is the first to estimate the impact of living in the USSR compared with the Eastern Bloc. So far, most studies on the causal effects of communism, influenced by Alesina and Fuchs-Schündeln (2007), focus on East-West Germany separation and reunification. The studies on gender roles include Klüsener and Goldstein (2014) (non-marital birth), Beblo and Görges (2018) (preferences for work), Lippmann and Senik (2018) (gender gap in mathematics), Campa and Serafinelli (2019) (career success), and Lippmann et al. (2020) (the male bread-winner norm and marriage stability).[2] We add to this literature by looking at a new environment and a different regime in Europe. It can be of particular interest in light of the recent paper by Becker et al. (2020b) who stress the potential bias in East and West Germany comparison due to pre-separation differences along the newly assigned country borders after World War II.

Furthermore, apart from bringing to attention a new setting, our estimates are based on capturing the difference within communism (e.g., due to private ownership or religion in Poland during the communist period). Consequently, results are not driven by the divergence between egalitarian and traditional regimes. Our findings are not affected by how much a male-bread winner family structure discouraged women from working. This is not the case when we consider East-West Germany, and two very distinct leading regimes: communism (East Germany) versus capitalism (West Germany). In addition to our main study, we repeat our specification in the East and West Germany context, using the SHARELIFE data. In this case, the coefficient of communism on women's experience gets even larger, because the control group is different: Poland before, West Germany now. In particular, East German women worked 5 years more by age 50 than West Germans, which is twice larger than in Lithuania-Poland comparison. It confirms the divergence of the regimes during the separation. The third contribution to this literature is an analysis of individual choices under the USSR regime. Instead of focusing on the persistence of the impact of the regime, we study the factors that shaped individual choices across birth cohorts from 1935 to 1950, isolating the period of the transition from planned to market economy.

The second group of articles we contribute to is the impact of gender-egalitarian policies on educational and work choices. Given the potential complementarity among institutions, the only way to estimate policies' synergy is to look at the historical context. The case of Eastern Europe, that experienced communism, gives us the laboratory to address the research question. Previous scholars showed that egalitarian gender policies increased women's labor force participation (e.g., Fuchs-Schündeln and Schündeln, 2020). Further, in our paper, we clarify the underlying women-specific channel that increased women's participation and schooling through almost compulsory labor force participation and educational availability. This indirect impact of the USSR on women is similar to a cascade effect of simultaneous free schooling and work opportunities (e.g., Wyrwich, 2019; Duflo, 2012). In the historical context, we mention just a few papers for the reason of space: Becker and Woessmann (2008) and Valencia Caicedo (2019)

[2] By any means, this list is not exhaustive. For detailed literature on East-West Germany see Becker, Mergele, and Woessmann (2020b).

stress the role of religion to enhance female's educational attainment. Facilitating women's presence in the labor market and closing the gender gap is in the agenda of many countries (World Economic Forum, 2020). By any means, we do not claim that a country must impose the Soviet Union policies to increase women's economic role, our paper highlights the existence of complementarity between education and job opportunities that can foster policies' results.

Next, we contribute to the literature about imperial legacies in the Baltic countries. The closest article to our analysis is Polugodina and Grigoriadis (2020), in which they investigate the impact of East Prussia on the persistence of political preferences in Lithuania, Russia and Poland. Vitola and Grigoriadis (2018) document the crucial role of the German population in the development of Latvia and Estonia.

Finally, we contribute to the literature on life satisfaction by documenting the gap within two transition economies. Individuals who lived under the USSR report statistically significant lower life satisfaction and life quality in 2017 than individuals from the Eastern Bloc, and this drop gets larger for men. Our findings are derived from older cohorts, who were above 50 years old in 2017. The strand of literature about "happiness in the transition" also reports a lower life satisfaction in the countries that undergo the economic and political changes from planned to market economy after 1990 (see, for instance, Guriev and Zhuravskaya, 2009; Djankov, Nikolova, and Zilinsky, 2016; Adsera, Dalla Pozza, Guriev, Kleine-Rueschkamp, and Nikolova, 2019). Authors focus primarily on West and East comparison. In a recent paper, Guriev and Melnikov (2018) show that this gap is almost closed, but mainly due to younger generations.

The paper proceeds as follows. Section 1.2 provides historical background about Lithuania and Poland and describes the differences in the communist regimes in these two countries. We introduce the dataset and the descriptive findings in Section 1.3, and we discuss the identification strategy in Section 1.4. Section 1.5 provides the main findings, the model about the education choice, and the treatment effect's heterogeneity analysis. Robustness checks are in Sections 1.7. The further analysis of East and West Germany is in Section 1.8. Section 1.9 concludes. Additional tables and figures are in the Appendix.

1.2 History of Lithuania and Poland and Soviet Ideology

In this Section, first, we briefly describe the history of Lithuania and Poland in Sections 1.2.1-1.2.2. The similarity between two areas in pre-communist era is crucial to get the causal impact of the regime. Next, we explain the characteristics of communism and the difference in its implementation in the USSR (Lithuania) and the Eastern Bloc (Poland) (in Section 1.2.3). A reader who is aware of the past of both countries and the Soviet influence in the region can freely skip Sections 1.2.1- 1.2.3 and continue with the summary of this Section in Section 1.2.4.

1.2.1 Brief History of Lithuania and Poland

From 1569 to 1918. Nowadays, two modern countries, Poland and Lithuania, have a land border. However, starting from 1569, the Crown of the Kingdom of Poland and the Grand Duchy of Lithuania formed one state, the Polish–Lithuanian Commonwealth, that existed for over 200 years. A single elected monarch governed in both territories for over 200 years. The state was gradually falling and reducing the size during the First and Second Partition of Poland (1772 and 1793) and entirely lost its independence after the Third Partition in 1795. The Congress of Vienna (1815) brought the last changes in the country borders. Three empires divided the territory of the Polish-Lithuania Commonwealth: the Russian Empire, the Kingdom of Prussia and the Hapsburg Austria. The Russian Empire got present-day Lithuania, except for the area of Klaipeda (also known as Memel territory), which belonged to East Prussia and a large part of the central areas of current Poland. Fig. A.1 in Appendix 1.10.1 shows the modern country borders between Poland and Lithuania, and highlights the territories of the Russian Empire until 1918.

From the first years, the Tsars of the Russian Empire took an active role in getting power in the newly annexed territories and suppressed any subsequent rebellion. The new governorates got established: Vilna Governorate and the Kingdom of Poland (also known Congress Poland and, from 1867 onwards, Vistula Land). Formally, Lithuania was one of the provinces in the Russian Empire, whereas the territory of present-day Poland was an area of limited autonomy (no political autonomy) within the Russian Empire (Eberhardt, 2003, p.72). Despite this formal difference, the Russian Empire's presence was large in both countries. Repressions against the participants of uprisings took place during the 19th century. The extensive Russification policies were universal; e.g., Lithuanian and Polish languages were severely discriminated against from public use. An excellent overview of the Russian Empire's impact on education, economic, and political variables is covered in Eidintas, Bumblauskas, Kulakauskas, and Tamovaitis (2016, Chapter 3) and Grosfeld and Zhuravskaya (2015), about Lithuania and Poland, respectively.

World War I (1914 - 1918). Shortly after the rising military conflict between European empires, in 1915, German troops advanced from East Prussia and occupied Lithuania until November 1918. The Russian Empire withdrew from World War I (WWI) due to its political crisis in 1917 and claimed no interest in Lithuanian territories until the Soviet Russian government attempted to increase its power at the end of 1918. Lithuania became an independent state in 1918 but faced a series of wars with its neighbors to establish its country borders until 1920 (Crampton, 2002, p.97).

Meanwhile, Poland was also reemerging as the state during WWI and the country regained its full independence from three ruling Empires in 1918. The Second Polish Republic (SPR) was established in 1918.

The territorial disputes between Lithuania and Poland lied into the city of Vilnius that was a formal territory of Lithuania but had a large share of the Polish population historically. After a one-month war between the two countries Vilnius and the surrounding region became part of Poland for the whole interwar period (Eberhardt, 2003, p.88-89). The border of Lithuania also slightly changed comparing with the 19th century: it got the northern territories of Suvalskaia province, that before belonged to the Polish Kingdom. Further, in 1923 the district of Klaipeda was incorporated into Lithuania with autonomous status (Eberhardt, 2003, p.90-91).[3]

The Interwar Period. The history of the two countries after regaining independence followed a similar path. In the first years after WWI, Lithuania and Poland had a democratic government. However, already in 1926 after a military coup in two countries, conservative authoritarian governments took power until World War II (Crampton, 2002, p.102 for Lithunia and p.46 for Poland). An excellent outlook of Lithuania and Poland in the 20th century can be found in Eidintas et al. (2016) and Bukowski and Novokmet (2017), respectively.

World War II (1939 - 1945). World War II (WWII) brought further changes in country borders. In August 1939, Nazi Germany and the USSR signed a non-aggression agreement called the Molotov–Ribbentrop Pact, which defined the two spheres of influence between countries. According to its modified version in October 1939, Lithuania became a territory of the USSR. From 1940 the Lithuanian Soviet Socialist Republic (Lithuanian SSR) was established, but in 1941 Nazi Germany invaded the territory and stood in power until 1944. At the end of WWII, the USSR got full control over the area (Snyder, 2004, p.98, p.154).

Poland also suffered constant combat on its territory during WWII. At the beginning of WWII (1939-1941), Poland was invaded by two powers: Nazi Germany on its west and the Soviet Union on its east. Then, Nazi Germany got full control over the country until 1944. In the last year of war, 1945, Poland was under Soviet occupation (Snyder, 2004, p.154).

[3]Based on these territorial changes, in the robustness checks we consider Lithuanian regions that corresponded with interwar country boarders.

After WWII. After WWII. At the end of WWII, Lithuania increased its territory by gaining back the region of Vilnius and the district of Klaipeda. However, its political freedom got suppressed and the country forcefully joined the USSR and became one of the republics, the Lithuanian SSR. Accordingly, Lithuanian provinces were under the centralized government in Moscow and Lithuanian citizens where exposed to the Soviet regime until the dissolution of the Soviet Union.

Poland was able to regain independence after WWII, even though the country was shifted to the west.[4] The communist power was established in 1945. From 1947, the Polish national government formed the Polish People's Republic and got heavily influenced by the USSR under the Eastern Bloc agreement until 1989 (European Commission, 2014).

After 1989/1990. Lithuania and Poland restored their independence in 1990 and 1989, respectively. After that, the two countries entered a transition period from planned to market economy. Later, on May 1, 2004, Lithuania and Poland simultaneously joined the European Union.

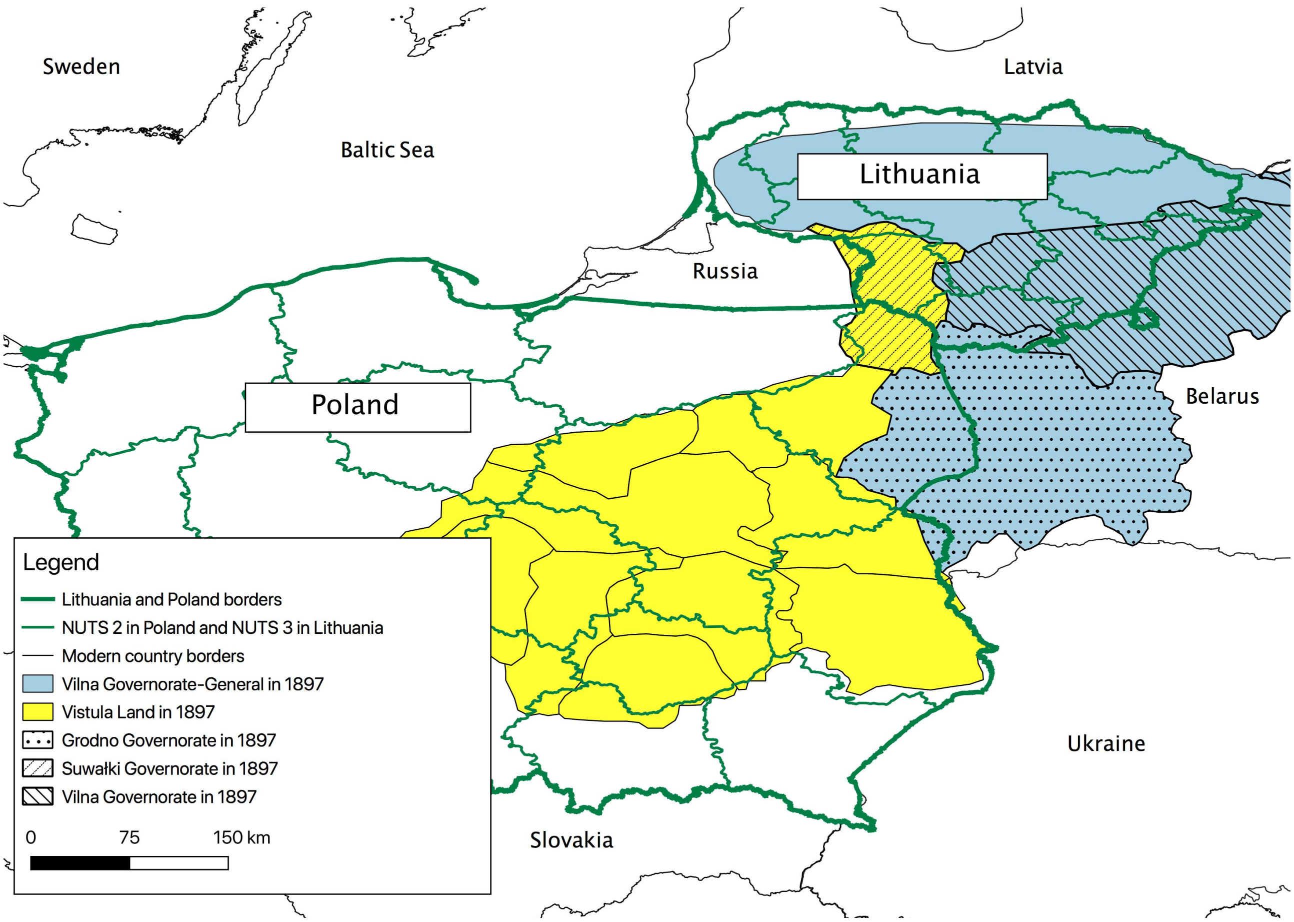
Our study argues that the institutional difference between the two countries after WWII creates a natural experiment because it was unexpected that the USSR's power would differ across Lithuania and Poland. Next, we show statistics about the similarity between the two countries in the pre-communist period.

1.2.2 Evidence on Similarity Between Lithuania and Poland

Matching present-day and historical borders during the Russian Empire. The territory of the Russian Empire was divided into governorates, those in provinces, and, finally, each province included several districts. After the Congress of Vienna, there were only minor changes in the province and district borders. In 1897, Vistula Land, the successor of the Congress of Poland, included Varshavskaia, Kalishskaia, Keletskaia, Liublinskaia, Lomzhinskaia, Petrokovskaia, Plotskaia, Radomskaia, Sedletskaia, and Suvalskaia provinces. Vistula Land covers territories of both modern Poland and Lithuania. Fig. 1.1 shows its territorial division. Vistula Land had a direct border on the East with Vilna Govenatore-General that included Grodnenskaia, Kovenskaia, and Vilenskaia provinces. Vilna Govenatore-General lies mainly in modern Lithuania, the North-West part of modern Belarus and modern Poland.

[4] Becker, Grosfeld, Grosjean, Voigtländer, and Zhuravskaya (2020a) discuss in details this exogenous shift in country boarders. Following their terminology, we can describe the territory of Poland under our analysis, as the region that lies in the intersection of Central Poland and former Russian Partition of Poland.

Fig. 1.1. Modern Poland and Lithuania Borders and Selected Provinces of the Russian Empire



Source: GIS map of country boarders in 2016 comes from Eurostat, GISCO. GIS map of the Russian Empire by province comes from Sablin, Kuchinskiy, Korobeinikov, Mikhaylov, Kudinov, Kitaeva, Aleksandrov, Zimina, and Zhidkov (2015).

The territories of almost all provinces of Vistula Land are included within the modern country borders of Poland. Equivalently, Vilna Governatore-General mainly lies in modern Lithuania. However, a significant part of Grodnenskaia (Grodno Governatore) and Vilnenskaia (Vilna Governatore) provinces is allocated in modern Belarus. Moreover, Suvalskaia (Suwalki Governatore) province is divided into two parts between modern Lithuania and Poland. Accordingly, to match the historical and present-day borders, we look at the district level whenever data allow. Fig. A.2, in Appendix 1.10.2 numbers, the relevant districts in Grodnenskaia, Vilnenskaia, and Suvalskaia provinces used in the descriptive analysis.

Russian Imperial Census in 1897. First, we exploit the Russian Imperial Census in 1897.[5]

[5]In Appendix 1.10.3, Table A.1 and Table A.2 report original raw statistics about demographic and labor characteristics in 1897 of present-day territories in Lithuania and Poland. The average value in Lithuania and Poland corresponds with an unweighted value. In Appendix 1.10.4, Table A.3 shows the industry of work among

The former territories of the Russian Empire that currently belong to Poland were, on average, higher populated than Lithuanian territories (see Column 3 in Table A.1 in Appendix 1.10.3). Accordingly, when we make the balancing test between modern-day Lithuania and Poland, we weight the sample by population in each district or province. Table 1.1 shows that the share of women in a country was about 50 in both places (though we find statistically significant differences). The religious composition that influences women's role along other variables was strikingly similar: about 75 percent of believers in the Roman Catholic church. This finding is important in light of a recent paper by Becker et al. (2020b), that stresses the potential bias in East and West Germany comparison due to the unequal share of Roman Catholics and Lutherans along the newly assigned country border after World War II.

Furthermore, the share of employed residents is almost 50 percent in both countries (see Table 1.1). This rate increases to about 75 percent for men in Lithuania and Poland, but the data for women slightly differ and are equal to 25 percent and 21 percent in Lithuania and Poland, respectively. Still, this difference is statistically insignificant once we perform a difference in means test at the district level.

Next, we report the illiteracy rate among men and women in two places, and we find a marginally significant difference between Lithuania and Poland. The illiteracy rate was slightly lower among Lithuanian women (62 percent) than among Polish ones (69 percent). This prior deviation in education can lead to upper-biased estimates. However, the magnitude of the difference is less than 25 percent of the actual literacy rate, and it is unlikely to violate our identification strategy dramatically.

Finally, women and men's prevalent job occupations look very similar among Lithuania and Poland. Agriculture is the dominant industry with the highest share of employed people in both countries. The only significant difference stems from the higher share of women working in services in Poland than in Lithuania.

the employed people in 1897.

Table 1.1: Balancing Test

	Lithuania (1)	Poland (2)	(1) - (2)	P-value
Panel I: Demographic and Labor Information in 1897				
Geographical Level: Provinces and Districts of the Russian Empire				
Percent of women	50.89	49.80	1.09	0.02
Percent of Catholics	75.88	73.34	2.54	0.60
Percent of Jews	13.54	14.54	-1.00	0.63
Percent of Orthodox	4.59	7.72	-3.13	0.47
Employed population/Age 11-60	48.49	48.21	0.29	0.92
Employed women/Women 11-60	24.87	20.77	4.11	0.14
Employed men/Men 11-60	73.62	75.42	-1.80	0.54
Geographical units	8	15		
Geographical Level: Provinces of the Russian Empire				
Percent of age 11-60	65.39	64.58	0.82	0.32
Percent of urban	10.59	22.96	-12.37	0.16
Percent of illiterate	62.35	69.80	-7.45	0.09
Percent of women illiterate	65.03	73.80	-8.77	0.08
Percent of Women Working in				
Capital owners	2.15	2.49	-0.33	0.52
Sellers	2.26	2.28	-0.02	0.98
Agriculture	46.32	35.04	11.28	0.27
Manufacture	9.14	10.69	-1.56	0.79
Services	19.40	32.11	-12.71	0.01
Undefined	18.63	15.23	3.40	0.09
Other	2.10	2.16	-0.07	0.89
Percent of Men Working in				
Capital owners	1.09	1.30	-0.21	0.42
Sellers	2.69	3.77	-1.08	0.11
Agriculture	60.82	47.48	13.34	0.15
Manufacture	9.47	15.80	-6.33	0.20
Services	13.82	18.64	-4.82	0.24
Undefined	6.61	6.80	-0.19	0.89
Other	5.50	6.22	-0.72	0.70
Geographical units	3	11		
Panel II: Factory Database in 1908				
Geographical Level: Provinces and Districts of the Russian Empire				
Number of factories	445.91	444.59	1.32	0.99
Density of factories	14.29	38.95	-24.65	0.10
Number of workers	33.17	62.69	-29.52	0.15
Revenue	69,928.37	137,204.15	-67,275.78	0.32
Power per worker	1.63	1.24	0.39	0.24
Industrial Composition in 1908				
Animal	10.01	9.81	0.20	0.98
Foods	27.50	20.07	7.43	0.31
Metals or machines	17.14	17.88	-0.74	0.91
Mineral products	10.61	11.29	-0.68	0.74
Paper	19.16	7.77	11.40	0.03
Wood	16.34	13.50	2.83	0.61
Wool	1.40	12.78	-11.38	0.39
Geographical units	8	15		

Sources: Demographic and Labor Information come from the Russian Imperial Census 1897. Data about factories come from the Imperial Russian Factory Database developed by Gregg (2020).
All statistics are weighted by the population size in each geographical unit, see Column 1 in Table A.1 in Appendix 1.10.3.

Imperial Russian Factory Database. Second, we study the Imperial Russian Factory Database, a new data source created by Gregg (2020) that systematize the Imperial Russian censuses of manufacturers of 1894, 1900, and 1908.[6]

Table 1.1 shows that, on average, Polish provinces had a higher density of factories in 1908 and a higher number of workers per factory. However, none of the differences is statistically significant. Further, the industrial composition of the two countries was very similar in 1908. About 80 percent of all factories were concentrated in five industries: foods, wood, animal, mineral products, and machine.

Interwar period. First, after WWI, the leading political regime in both countries was similar and followed the same path from a newly-born democratic government to an autocratic national regime (Eidintas et al., 2016; Bukowski and Novokmet, 2017).

Second, given the changes in country borders in Lithuania and Poland during the interwar period, we cannot provide historical statistics that match the present-day territory under analysis. However, evidence from the Lithuanian census in 1923, not including the region of Vilnius, show that country was mainly populated by Lithuanians and had one big minority group - Jews (see Table 1.2). Aggregate data about Poland come from the Polish census in 1931 that includes the vast territory of present-day Belarus and Ukraine.[7] Table 1.2 shows that overall, Polish formed the majority in the country, but Ukrainians and Jews formed the largest minority groups. However, when we look at the province level in Fig. A.4, Appendix 1.10.7, most minorities lived outside the former territory of the Russian Partition. Accordingly, we argue that the area under our analysis in Lithuania and Poland, in the interwar period, preserved its ethnic population from the former Russian Empire period as minorities concentrated in the other country territories.

Impact of WWII and Migration. WWII was a traumatic episode in the history of the two countries. Given its complexity, we discuss the impact of several channels of WWII and the potential bias in our estimations in detail in Section 1.6. Overall, the main thread for our identification arises only if one of the country was significantly more affected by WWII than the other.

After WWII. The main treatment variable in our paper is the exogenous difference between the political regimes in Lithuania and Poland after WWII. The Soviet government took the

[6] In Appendix 1.10.5, Table A.4 and Table A.5 report original raw statistics from the Imperial Russian Factory Database of present-day territories in Lithuania and Poland.

[7] Fig. A.3 in Appendix 1.10.6 pictures the ethnic composition of Lithuania in 1931 and 1939. Fig. A.4 in Appendix 1.10.7 pictures the ethnic composition of Polish provinces in 1931. The former Russian Partition roughly corresponds with Warsaw, City of Warsaw, Lodz and Kielce.

Table 1.2: Ethnic Structure in Two Countries during the Interwar Period

Lithuania in 1923			Poland in 1931		
Ethnic Group	N	%	Ethnic Group*	N	%
Lithuanians	1701900	83.9	Polish	21993400	68.9
Jews	153700	7.6	Ukrainian	4441600	13.9
Poles	65600	3.2	Jewish	2732600	8.6
Russians	50500	2.5	Belorussian	1698100	5.3
Germans	29200	1.4	German	741000	2.3
Latvians	9000	0.4	Russian	138700	0.4
Others	19100	1.0	Others	170400	0.6
Total	2029000	100	Total	31915800	100

Sources: Data about Lithuania based on Census from 1923 is from Table 2.14 in Eberhardt (2003). Data about Poland is from Table 3.24 in Eberhardt (2003).
* Corresponds with a declared language.

main role in Lithuanian SSR. During the 1950s, ethnic Russians began to migrate into Baltic Republics, but differently from Latvia and Estonia, Lithuania preserved the national character of their state (Eberhardt, 2003, p.124). The share of Russians was about 8 percent in 1959 (see Table 1.3), and they were dispersed across the country. Accordingly, Lithuanians formed the largest ethnic group in 1959. Likewise, after a change in country borders and the deportations of ethnic minorities, Poland also became ethnically homogeneous. In 1950, around 98 percent of the population in Poland were Poles (see Table 1.3).

Table 1.3: Ethnic Structure in Two Countries after WWII

Lithuania in 1959			Poland in 1950		
Ethnic Group	N	%	Ethnic Group*	N	%
Lithuanians	2150800	79.3	Polish	24448000	97.8
Russians	231000	8.5	Ukrainians	170000	0.7
Poles	230100	8.5	Germans	160000	0.6
Belorussians	30300	1.1	Belorussians	150000	0.6
Jews	24700	0.9	Jews	50000	0.2
Ukrainians	17700	0.7			
Others	26800	1.0	Others	30000	0.1
Total	2711400	100	Total	25008000	100

Sources: Data about Lithuania is from Table 2.27 in Eberhardt (2003). Data about Poland is from Table 3.42 in Eberhardt (2003).

Landscape. Next, we show no difference in the geographic and climatic characteristics in Lithuania and Poland. In Appendix 1.10.8, Table A.6 confirms the similarity in the landscape and temperature in the two countries.

Schooling Reforms. We conclude this Subsection with the summary of schooling policies in the areas under analysis. Backhaus (2019) and Bukowski (2019) overview the educational policies in Poland. For the whole 19th century, there was no mandatory education in the Russian partition of Poland, and overall the literacy rate was very low (in line with Table A.1). In 1919 the compulsory elementary schooling was introduced, and seven-year education was universal in the country. WWII had devastating consequences for the country, but the relatively fast recovery followed them, and already in 1949, the mandatory seven-year education was back in the country. The final change in compulsory education was in 1961 when eight-year schooling became mandatory. Bieliauskienė (2014, p.61, 62) gives an extensive overview of the history of the educational system in Lithuania from the 14th century. The 7-year education became compulsory in 1949. In 1959, eight-year education became mandatory.

1.2.3 Enforcement of the Soviet Regime in Lithuania and Poland

In this Section, we describe how the economic, political, and social distinctions of the Soviet regime impacted the lives of people and show evidence for the different level of enforcement of the regime in the Soviet Union and Eastern Bloc.

The Soviet Union. After the October Revolution in 1917, the Bolsheviks government started the process of taking power from the Emperor of Russia, Nicholas II Romanov, and at the end of the civil war in 1922, they formed the Union of Soviet Socialist Republics (USSR). Vladimir Lenin was the first leader of the new-born country. Importantly, there was a one-party system with a centralized planned economy that had control over all sectors.

Soviet Regime. According to Communist Party's ideology, women have the same equal rights as men in economic, political, and family life (Atkinson, Dallin, and Lapidus, 1977, p.115). The party leaders launched country-wide campaigns to bring women to the labor force from the first years and encourage their educational attainment (see Appendix 1.10.9, Fig. A.5). Home production was considered a secondary activity, and each individual's responsibility was to work in social production. To bring further women in the labor market, not working was considered as being dependent and unpatriotic (Atkinson et al., 1977, p.170). Female participation was promoted independently of the marital status and presence of children, the authorities always facilitated entry to the labor market. From the beginning, the Soviet regime promoted public services like child-care or the production of consumer durables to ease home production. However, very soon, they faced significant under the provision. Chapter 5 in Lapidus (1978) gives a detailed overview of the evolution of female participation in the USSR and the general trends in society. Previous literature on East and West Germany's separation extensively documented policies in East Germany during the Soviet influence. An interested reader can check Campa and Serafinelli (2019) and Lippmann et al. (2020).

Soviet Regime in Lithuania and Poland. The Soviet regime was exported to Baltic and Eastern European countries. However, there was an essential difference among them: Lithuania was the part of the USSR; however, Poland was an independent country that got influenced by the Soviet Union. This political freedom of the last could lead to the lagged and weaker implementation of the Soviet regime.

Here we want to illustrate two aspects that differed between the two states and illustrate only partial enforcement of the regime in Poland. The first factor stems from the Roman Catholic Church. The USSR was an atheistic state. During the first years, the church's power was dramatically suppressed; most churches were either destroyed or transformed into non-religious buildings. Believers and priests faced forced migration. Accordingly, only after the USSR's dissolution, the Pope of the Roman Catholic Church could come to Lithuania. On the contrary, religion was widespread in Poland, even during the communist period. Despite being officially an atheistic country, in 1956, Wladislaw Golumka (a de-facto leader of Poland) eased pressure on the church. In 1979, John Paul II, the Pope of the Roman Catholic Church (Polish origin), received a warm welcome during his visit to Poland. The majority of members of the communist party were believers (Brzezinski, 1967, p.36). Moreover, in 1980, the Independent Self-governing Trade Union "Solidarity" was formed and organized mass protests and strikes. However, for the first time, the Soviet Government did not intervene with military forces to suppress a conflict in an Eastern Bloc country (White et al., 1993, p.6). These mass events were never present in Lithuania during the regime.

The other difference between the Soviet regime intensity in Lithuania and Poland stems from the degree of land collectivization. The Soviet Government conducted an extensive land reform to eliminate private property rights. Brzezinski (1967, p.36) notes that nationwide collectivization companies were successful everywhere in the USSR and the Eastern Bloc but Poland. After several rebellions in 1956, Polish farmers were allowed to own land. Unprecedentedly, in the 1980s, three-quarters of all farmland in Poland was private property and a quarter of the country's workforce were private farmers (see Brzezinski, 1967; White et al., 1993). Meanwhile, mass land collectivization in Lithuania was very effective, and already by 1950, 90% of all land became state-owned and remained so until the USSR's dissolution (Girnius, 1988). Appendix 1.10.10, Table A.7 summarizes the mentioned differences above.

1.2.4 Summary of the Section

1. Lithuania and Poland formed the Polish–Lithuanian Commonwealth from 1569 to 1795.
2. The state lost full independence, and the Russian Empire governed in Lithuania and the large part of the central areas of Poland until 1918.
3. Lithuania and Poland regained independence in 1919 after the Treaty of Versailles.
4. In 1939, the USSR took the territories of Lithuania as a part of their own state. After 1947, Poland preserved its national government but got heavily influenced by the USSR as a part of the Eastern Bloc.
5. The Soviet regime brought structural changes to society, one of which is egalitarian gender roles.
6. The Soviet regime was only partly enforced in Poland compared to Lithuania due to the presence of the Roman Catholic church and private ownership in the former.
7. Lithuania and Poland regained back their full independence in 1990 and 1989, respectively.

1.3 Data and Descriptive Findings

In this Section we discuss the Data used in the paper.

SHARE and SHARELIFE Data. This paper exploits the Survey of Health, Ageing and Retirement in Europe (SHARE), the SHARELIFE, and the SHARE Job Episodes Panel[8]. The eligible participant in the SHARE survey is above 50 years old. The main survey provides socio-demographic, health, and economic information about individuals. We consider only wave 7 (2017) because it allows studying Eastern European countries for the first time.

The SHARELIFE survey aims to represent individuals' life history, and it is part of the SHARE project. It was conducted twice: in wave 3 (2007) and wave 7 (2017). We use information about all respondents who participated in the main SHARE survey in 2017 and merge the available information from the retrospective studies in 2007 and 2017. Finally, we exploit the Job Episodes Panel based on the SHARELIFE survey to follow the individual working history.

[8] See Brugiavini, Orso, Genie, Naci, and Pasini (2019) for details about construction.

The quality of re-called information can be of concern, so previous studies based on wave 3 of SHARELIFE (Brunello, Fabbri, and Fort, 2013; Kesternich, Siflinger, Smith, and Winter, 2014; Crespo, López-Noval, and Mira, 2014; Fort, Schneeweis, and Winter-Ebmer, 2016; Havari and Mazzonna, 2015) run numerous tests and argue data trustworthiness.

We restrict our analysis to individuals who lived most of their working life under the Soviet regime, so we consider only those born between 1935 and 1958 to start the working career during the regime. Our targeted sample includes 70 percent of all SHARELIFE participants in 2017 (see Appendix 1.11.1, Fig. A.6).

Also, the main results of the paper are based on the data from all present-day regions in Lithuania and the regions of modern Poland that were part of former Russian Partition: Lublin Voivodeship, Lódz Voivodeship, Masovian Voivodeship, Podlaskie Voivodeship and Swietokrzyskie Voivodeship.[9,10] In the extension, we also consider the interwar borders, that exclude Klaipeda region (also called Memel territory) in Lithuania because a part of it was Prussian partition and remove Vilnius, Alytus and Utena regions in Lithuania because parts of them were territories of Poland in the interwar period. In Appendix 1.11.2, Fig. A.7 numbers the corresponding Lithuanian and Polish modern regions.

Treatment variable. We define the USSR variable to capture the treatment. We classify a respondent in the USSR or the Eastern Bloc if he lived at age 18 (at the beginning of the working career) in Lithuania or one of those Polish regions under the analysis, respectively. By the survey construction, we can observe full individual residential history. In our sample, 50 percent of respondents never change the region of residence during their life (see Appendix 1.11.4, Fig. A.9). Still, we show that findings are robust to movements during the life, like the region in which a person was born or lived the most of the life.

Outcome variables. First, we study the educational level of an individual; and we use seven ISCED-1997 categories provided by organizers. We also define pooled categories of education: low-, secondary- and high-education to overcome the limited size of some original categories.[11] In Appendix 1.11.3, Fig. A.8 shows empirical cumulative distributions for the two education variables across individuals from the USSR and Bloc: the left-hand-side Panel is the original ISCED-1997 level, and the right-hand-side Panel shows the aggregate variable. We see that

[9]Information about the region of residents is at NUTS 3 level for Lithuania and NUTS 2 level for Poland. Accordingly, in total, we have 10 regions in Lithuania and 5 regions in Poland.

[10]Following the terminology in Becker et al. (2020a) we can describe the territory of Poland under consideration, as the region that lies in the intersection of Central Poland and former Russian Partition of Poland.

[11]The low education group includes no education at all, ISCED-1997-1 and ISCED-1997-2. Secondary education corresponds with ISCED-1997-3. The high-education group consists of ISCED-1997 above 3.

individuals from the USSR acquired more education than those born in the Eastern Bloc.

Second, we look at the labor experience variables. Using the Job Episodes Panel, we compute cumulative years of work experience at each age. We restrict to individuals who were born from 1935 to 1958, and we count only years of experience before the fall of the regime, 1990, not to confound with the transition period after the fall of the USSR. In our analysis, first, we use cumulative years of experience by age 25, which allows us to verify the trends in the early-life career. Next, we study experience by age 50 as cumulative experience during the whole life, abstracting from early retirements. Finally, we construct a labor experience between 25 and 50 years variable to eliminate the study-work trade-off early in life.

Then, we consider marital and fertility history during life: a dummy indicator to marry at least once, the number of marriages, the age at the firth birth and the number of children. To abstract from the confounding factors during the transition period in Lithuania and Poland, we consider only events before 1990.[12]

Finally, we consider life satisfaction (a categorical variable from 0 to 10), and life quality (a categorical variable from 12 to 48) measures.

Control variables. Similar to other authors who used the SHARELIFE (Brunello et al., 2013; Kesternich et al., 2014; Crespo et al., 2014; Fort et al., 2016), we include proxies for early life socioeconomic status (SES): four categories of health at age 10, a mental health problem dummy at age 10, to be a good student at math dummy; five places of birth dummies: a big city, the suburbs of a big city, a large town, a small town or rural area; the features of the individual's dwelling at age 10: three dummies for the number of books by age 10, the number of services (e.g., hot running water supply, having a toilet inside the house and others), and the number of rooms. Further, we add year of birth fixed effects.

Appendix 1.11.5, Table A.8 shows the descriptive statistics of variables listed above for the target sample.

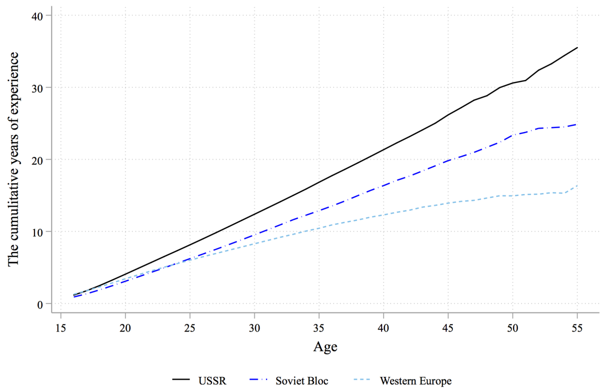
To support the idea that the Soviet regime played an essential role in women's labor decisions in the 20th century, we consider the evolution of working life among men and women in each country group according to the place of birth: the USSR (Estonia, Latvia, and Lithuania), the Eastern Bloc countries (Bulgaria, Czech Republic, Hungary, Poland, Slovakia, and Romania), and Western European countries (Western countries consists of Austria, Belgium, Cyprus, Denmark, Greece, Finland, France, Luxembourg, Malta, Italy, Portugal, Spain, Sweden, and

[12]We exclude all marriages after 1990 and children born after 1990.

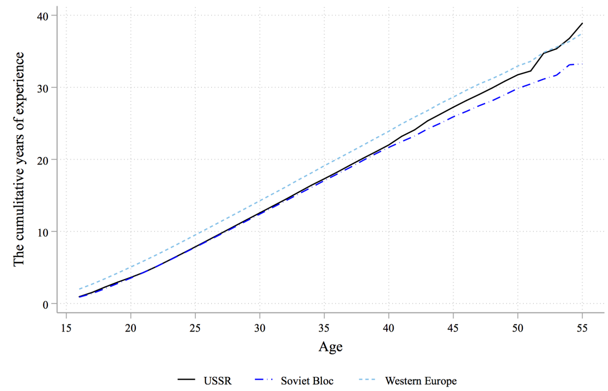
Switzerland).[13] In Fig. 1.2, we split the sample according to educational attainment in three groups: low, secondary, and high. Men's profiles look very similar across countries and education levels, whereas the profile of women notably differs. Women in the USSR accumulate more years of work experience by age 50 regardless of their schooling. This difference in the sharpest among a low-educated group. In our analysis we exploit the life decisions among respondents from Lithuania and Poland.

[13] We exclude Croatia, Israel, and Slovenia from the principal analysis because of the political regime changes. We also exclude Germany as the country was divided into two parts after World War II.

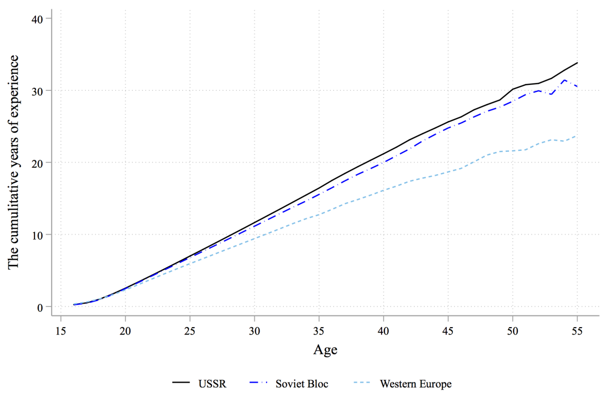
Fig. 1.2. Work Experience Across Europe between 1950 and 1990



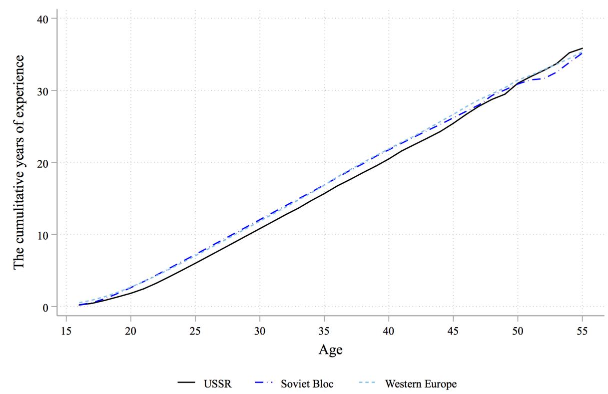
(a) Lower/Primary Education, Women



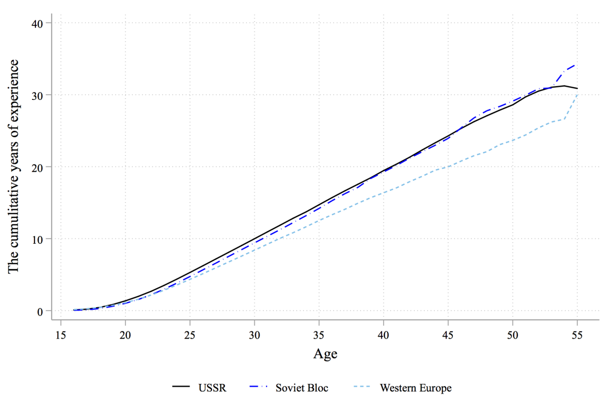
(b) Lower/Primary Education, Men



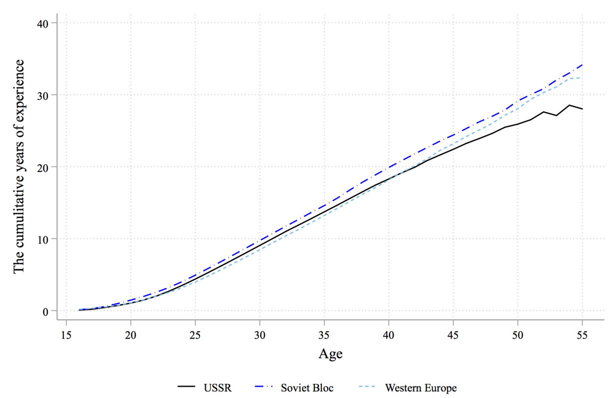
(c) Secondary Education, Women



(d) Secondary Education, Men



(e) Above Secondary Education, Women



(f) Above Secondary Education, Men

1.4 Methodology

In this Section, we describe our empirical strategy to measure the impact of the USSR on education, work experience, fertility choice, and later-life outcomes. First, we describe the benchmark specification and review the identification assumption. Second, we introduce a series of extensions that allow us to study the USSR's heterogeneous impact and the intensive margin of the effect.

1.4.1 Benchmark Specification

Our identification strategy is a natural experiment. We draw the first set of results by fitting

$$Y_i = \gamma_0 + \gamma_1 G_i + \alpha_1 Z_i + \alpha_2 G_i \cdot Z_i + \beta' X_i + \varepsilon_i \qquad [1.4.1]$$

where Y_i is an outcome variable of individual i (discussed in Section 1.3), G_i is a female dummy, Z_i indicates that the individual lived in the USSR at age 18,[14] X_i is the set of controls (see Section 1.3). ε_i is an unobserved error.

When fitting [1.4.1] and the regressions to follow, we use SHARELIFE weights provided by the SHARE to get the representative sample of individuals above 50. Finally, we allow for correlation in unobserved errors among people of the same age and from the same region, and use the cluster standard errors at the year of birth and the region of residence at age 18 level.[15]

The target sample includes only individuals who were born between 1935 and 1958. It guarantees that they start their working career during the regime. Furthermore, we can track their choices and count only life episodes before 1990, not to confound with the transition period after the USSR's dissolution.

This paper aims to identify the impact of the USSR on a variety of outcomes and investigate a gender-specific channel. As argued in Section 1.2.2, we see that individuals who lived in the USSR were more exposed to the Soviet ideology than their counterparts in the Eastern Bloc. The Soviet Union was able to enforce the leading ideology in its territories strongly. We thus want to quantify the differences between two communist regimes. Key to understanding the parameters in [1.4.1] as the impact of Soviet ideology and policies is the following assumption,

[14] We do also perform our analysis with different definitions of the treatment variable, e.g, whether an individual was born in the USSR or lived most of her life in the USSR. The main results hold (see Section 1.7).

[15] We repeat the same analysis but using the robust standard errors or clustered standard errors at the year of birth and the region of work, and the results for all outcome variables hold.

introduced in Section 1.2.1:

Assumption. *Lithuania became a part of the USSR and not the Eastern Bloc due to exogenous factors that are unrelated to the outcome variables relevant to this study. Our identification relies on the similarity between the former territories of the Russian Empire in Lithuania and Poland.*

As we argued in Section 1.2.2, we believe that the territories under consideration were similar before the adhesion of Lithuania to the USSR. Therefore, Z_i measures the impact of Soviet ideology and policies and is not confounded by pretreatment differences.

To measure the effect of the USSR, we compute the the Average Marginal Impact (AMI) in a variety of specifications. The AMI of the USSR on women is defined as the impact of the regime on an average woman. It answers the following question: how much would have changed the outcome (e.g., education or working experience) of a woman, who is around the average in terms of socioeconomic status, if she had lived in the USSR? According to [1.4.1], the AMI of the USSR on women, denoted as AMI^f, is measured by:

$$\text{AMI}^f = \mathbb{E}\left[\frac{\partial Y_i}{\partial Z_i}\Big|G_i = 1\right] = \alpha_1 + \alpha_2 \qquad [1.4.2]$$

Since, in the specification in [1.4.1], $\partial Y_i/\partial Z_i$ does only depend on G_i, the AMI can also be read as the impact of communism on women, keeping all other covariates fixed:

$$\text{AMI}^f = \mathbb{E}[Y_i|G_i = 1, Z_i = 1, X_i = x] - \mathbb{E}[Y_i|G_i = 1, Z_i = 0, X_i = x] \qquad [1.4.3]$$

The latter is a ceteris paribus effect, instead of an effect on an average woman. Once we extend the benchmark setup to account for the heterogeneous effects by education, we will lose this ceteris paribus interpretation of the AMI.

For men, the AMI of the USSR is given by $\text{AMI}^m = \alpha_1$. The parameter corresponding to the interaction term, α_2, is central to this study. This is the interaction parameter, which measures the differential impact of the Soviet Union on women when compared to men, i.e., $\alpha_2 = \text{AMI}^f - \text{AMI}^m$. It quantifies the gender-specific channel.

Our first extension to [1.4.1] is to allow for gender varying covariates. We estimate equations [1.4.4] bellow for separate samples of women and men:

$$\begin{aligned} Y_i &= \gamma^f + \alpha^f Z_i + \beta^{f\prime} X_i + \varepsilon_i \quad \text{if } G_i = 1, \\ Y_i &= \gamma^m + \alpha^m Z_i + \beta^{m\prime} X_i + \varepsilon_i \quad \text{if } G_i = 0 \end{aligned} \qquad [1.4.4]$$

Here, the superscripts $\{f, m\}$ denote female and male, respectively. In the regressions above, the effect of controls on the outcome variable, β, is allowed to vary across gender. That is, this specification is more flexible than [1.4.1]. Even though estimating an equation separately for each gender leads to a considerable sample size drop, we want to confirm that the effects estimated by [1.4.1] do not rest on the imposition of a homogeneous impact of other covariates. Moreover, in extensions, we use [1.4.4] to allow for the heterogeneous impact of Soviet ideology by education (see Section 1.4.2).

In the setup specified by [1.4.4], the USSR's impact on women is measured by $\text{AMI}^f = \alpha^f$. Equivalently, $\text{AMI}^m = \alpha^m$. In this specification, we can also study the differential impact on women by estimating $\text{AMI}^f - \text{AMI}^m = \alpha^f - \alpha^m$. As mentioned above, in the absence of heterogeneous impact of covariates, we expect these coefficients to be close to the estimate α_2 in [1.4.1].

1.4.2 Heterogeneous Impact of the USSR

Now, we extend the specification to account for the impact of the Soviet Union by education groups. Education is itself an outcome of the USSR regime and policies. However, it is also an important predictor of work experience: in early life, there is a trade-off between working and studying, whereas educated people face different working opportunities (see our model in Section 1.5.2). Fig. 1.2 in Section 1.3 show that the impact of the Soviet Union may be heterogeneous in education.

This extension has two goals. First, as we show in Appendix 1.12, the AMI may be inconsistently estimated if heterogeneity is not accounted for. Second, it allows a deeper analysis of the impact of the USSR. Indeed, as shown in Fig. 1.2, there are noticeable differences across education groups. We therefore allow the USSR coefficient in [1.4.4] to vary with the education level of the individual:[16]

$$Y_i = \gamma^f + \alpha^f(E_i) Z_i + \beta^{f\prime} X_i + \varepsilon_i \qquad [1.4.5]$$

[16] We use the regression for women for exposition. Note, however, that we also consider this extension for male respondents.

where E_i denotes educational attainment of individual i. We consider two different measures of the education level: the seven ISCED-1997 categories and a pooled low-, secondary- and high-education variable (see Section 1.3). For simplicity of exposition, we derive this section considering the pooled education variable, which can take 3 values: 0 (low), 1 (secondary), and 2 (High). So, without loss of generality:

$$\alpha^f(E_i) = \alpha_0^f + \alpha_1^f E_{1i} + \alpha_2^f E_{2i} \tag{1.4.6}$$

being E_{ji} a dummy variable indicating that individual i is in the j-th education group, for $j \in \{0, 1, 2\}$. This leads to the following regression, which includes the interaction terms between the USSR variable (treatment) and education dummies:

$$Y_i = \gamma_0^f + \gamma_1^f E_{1i} + \gamma_2^f E_{2i} + \alpha_0^f Z_i + \alpha_1^f Z_i E_{1i} + \alpha_2^f Z_i E_{2i} + \beta^{f\prime} X_i + \varepsilon_i \tag{1.4.7}$$

Estimation of [1.4.7] leads to a profile of impact of the Soviet Union across education. Now, we are able to estimate the impact of the regime on women, conditional on a fixed education level. We construct the (conditional) AMI of the USSR on women following the equation bellow:

$$\text{AMI}^f(e) = \begin{cases} \alpha_0^f & \text{if } e = 0, \\ \alpha_0^f + \alpha_1^f & \text{if } e = 1, \\ \alpha_0^f + \alpha_2^f & \text{if } e = 2 \end{cases} \tag{1.4.8}$$

We use the above equation, and its counterpart for men, to plot the profile of the the impact of the USSR across educational attainment. For a justification of the above formula, see Apendix 1.12.

Computation of the unconditional AMI of the USSR for women becomes more complex in the presence of heterogeneity. Nevertheless, writing the model as in equation [1.4.5] allows us to show that:

$$\text{AMI}^f = \mathbb{E}\left[\frac{\partial Y_i}{\partial Z_i}\middle| G_i = 1\right] = \mathbb{E}[\alpha^f(E_i)|G_i = 1] \tag{1.4.9}$$

Thus, by the Law of Iterated Expectations, $\text{AMI}^f = \mathbb{E}\left[\text{AMI}^f(E_i)\right]$, where $\text{AMI}^f(e)$ is given in equation [1.4.8]. That is, once we have constructed the profile of impact of the USSR across education from the estimands in equation [1.4.7], we need to average it using the distribution of education in the female subsample. We provide the specific formula for the AMI in Appendix 1.12.

1.4.3 Intensive Margin of the Impact of the USSR

In this Section, we adapt the methodological tools described in 1.4.2 to study the intensive margin of the impact of the USSR. We exploit the difference across birth cohorts in our sample. The exposure of these cohorts to Soviet Union and its policies varies within our sample as the regime fell roughly in 1990. As we can see in Fig. A.10, the maximum attainable years of work experience before the fall of the USSR is higher for older cohorts, as the regime spanned the whole working lives of these individuals. In contrast, younger cohorts were in the middle of their career when the regime fell. However, it is worth noting that these younger individuals did perform their educational choices under the regime, being uncertain about its eventual end.

We can exploit these variation to measure the intensive margin of the impact of the Soviet Union. Indeed, we claim that this is the heterogeneous impact of the USSR by birth cohort. Therefore, we adapt the specification in [1.4.5] to account for heterogeneity:

$$Y_i = \gamma^f + \alpha^f(C_i)Z_i + \beta^{f\prime}X_i + \varepsilon_i \qquad [1.4.10]$$

where C_i denotes the birth cohort of individual i. Assuming that there are $\mathcal{C}$ cohorts:

$$\alpha^g(C_i) = \alpha_0^g + \sum_{j=1}^{\mathcal{C}-1} \alpha_j^g C_{ji} \qquad [1.4.11]$$

being C_{ji} a dummy variable indicating that individual i belongs to the j-th cohort. Following the same reasoning as in 1.4.2, we obtain the intensive margin of the effect of the USSR on woman by computing:

$$\text{AMI}^f(c) = \begin{cases} \alpha_0^f & \text{if } c = 0, \\ \alpha_0^f + \alpha_c^f & \text{if } c \neq 0 \end{cases} \qquad [1.4.12]$$

where the α's are obtained from [1.4.10].

1.5 Results

First, this Section reports the findings on education and experience. Next, we introduce the model of educational choice that guides the mechanism of the USSR impact. Then, we study the heterogeneous impact across education and explore the difference across birth cohorts. Finally, we report the findings on marriage history, fertility, and later-life well-being. All the results correspond with the life outcomes of individuals born between 1935 and 1958 during the regime, before 1990.

1.5.1 USSR, Education and Work History

First, Column 1 in Table 1.4 shows the impact of the USSR on the education level measured by seven ISCED-1997 categories. *Panel I* reports the results for the pooled sample of men and women. The female coefficient, γ_1, is negative as expected, that confirms the general trend in lower schooling among women. The USSR coefficient, γ_3, shows evidence that, on average, people who were grown in the Soviet Union get more education. It reflects the availability of education in the country. The gender-specific impact, γ_2, shows that the USSR affected more women than men; and looking at the Average Marginal Impact (AMI) for women, we see that they accumulate 80 percent a level more of education than the ones from the Eastern Bloc.[17] This finding relates to the magnitude of the USSR coefficient on the subsample of women and men, *Panel II* and *Panel III* respectively. We will discuss the origin of this differential impact on women after presenting the results of experience.

Next, we check the impact of Soviet Union on work experience by age 25, by age 50 and between 25 and 50 years old. We begin with the regime's total impact, not isolating the effect on schooling, in Columns 2 - 4. In *Panel I*, the coefficient on female is negative in all specifications, which is in line with the intuition about lower women's labor attachment, particularly during the 20th century. The USSR estimate in Column 2 is negative and significant, that reflects the lower labor attachment by age 25 that, indeed, is very likely in the presence of higher opportunities to study in the USSR. In Column 4, we abstract from the early-life trade-off between schooling and working and look at the experience from 25 to 50 years old to estimate the impact of work enforcement policies. Accordingly, the USSR coefficient is equal to 0.46 and statistically significant at 10 percent level (see Column 4), implying that, on average, individuals under the strong Soviet regime accumulate close to half a year more of work experience during those periods of life. Finally, the women-specific impact is larger than the one for men, and it is equal to 1.42 years more of experience. These findings hold when we look separately on men and women in *Panel II* and *III*. Moreover, the interaction term, γ_2, is strongly significant for all outcome variables, indicating a strong differential effect of the USSR on women. Additionally, the AMI of the USSR on women increases over the life cycle: by age 50 women in the USSR accumulate almost 2 years of work experience more than women in the Eastern Bloc.

[17] Following the specification in [1.4.4], we can also compute the differential impact of the USSR on women as the difference in the USSR coefficients in the subsample of women (*Panel II*) and men (*Panel III*).

Table 1.4: USSR, Education and Work Experience

		Cumulative work experience								
	Education	No control for education			Controls for three education levels			Heterogeneity with education		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Variables		By 25	By 50	25-50	By 25	By 50	25-50	By 25	By 50	25-50
Panel I: Both men and women										
Female	-0.289***	-0.485**	-2.085***	-1.600***	-0.596***	-2.062***	-1.466***			
	(0.0646)	(0.210)	(0.446)	(0.311)	(0.203)	(0.428)	(0.297)			
Female × USSR	0.519***	1.163***	2.120***	0.956**	1.480***	2.299***	0.819**			
	(0.106)	(0.280)	(0.554)	(0.387)	(0.265)	(0.539)	(0.380)			
USSR	0.300***	-0.741***	-0.275	0.466*	-0.440**	-0.00417	0.436*			
	(0.0852)	(0.205)	(0.356)	(0.242)	(0.204)	(0.367)	(0.251)			
Education:										
Secondary					-0.312	0.998*	1.310***			
					(0.231)	(0.575)	(0.439)			
High					-2.191***	-0.924	1.267***			
					(0.311)	(0.608)	(0.421)			
AMI of the USSR on women	0.819***	0.423*	1.845***	1.422***	1.039***	2.294***	1.255***			
P-value: AMI=0	0.000	0.052	0.000	0.000	0.000	0.000	0.000			
R^2	0.338	0.0925	0.391	0.540	0.131	0.397	0.547			
N	2190	2190	2190	2190	2190	2190	2190			
Panel II: Women										
USSR	0.852***	0.395*	1.972***	1.577***	1.047***	2.359***	1.312***	1.840***	4.482***	2.642***
	(0.0736)	(0.226)	(0.483)	(0.358)	(0.231)	(0.522)	(0.401)	(0.533)	(1.227)	(0.877)
Education:										
Secondary					-0.235	1.665*	1.900***	-0.149	1.878**	2.027***
					(0.326)	(0.862)	(0.678)	(0.342)	(0.920)	(0.727)
High					-2.204***	-0.320	1.884***	-2.104***	0.0544	2.159***
					(0.360)	(0.787)	(0.604)	(0.427)	(0.900)	(0.679)
Education × USSR:										
Secondary × USSR								-1.203**	-2.942**	-1.739*
								(0.563)	(1.303)	(0.950)
High × USSR								-0.907	-2.727**	-1.820**
								(0.598)	(1.234)	(0.871)
AMI of the USSR	0.852***	0.395*	1.972***	1.577***	1.047***	2.359***	1.312***	1.148***	2.715***	1.566***
P-value: AMI=0								0.000	0.000	0.000
R^2	0.407	0.0933	0.326	0.454	0.128	0.335	0.466	0.129	0.336	0.467
N	1277	1277	1277	1277	1277	1277	1277	1277	1277	1277
Panel III: Men										
USSR	0.294***	-0.669***	-0.136	0.533***	-0.396*	0.112	0.508**	0.521	1.648*	1.127*
	(0.0891)	(0.229)	(0.327)	(0.194)	(0.229)	(0.336)	(0.198)	(0.490)	(0.885)	(0.574)
Education:										
Secondary					-0.488	-0.252	0.236	-0.394	-0.110	0.284
					(0.352)	(0.729)	(0.490)	(0.374)	(0.780)	(0.526)
High					-2.214***	-1.861**	0.353	-2.141***	-1.662	0.479
					(0.498)	(0.895)	(0.558)	(0.565)	(1.012)	(0.631)
Education × USSR:										
Secondary × USSR								-1.320**	-1.974*	-0.654
								(0.566)	(1.021)	(0.653)
High × USSR								-0.983	-1.989*	-1.006
								(0.697)	(1.191)	(0.733)
AMI of the USSR	0.294***	-0.669***	-0.136	0.533***	-0.396*	0.112	0.508**	-0.340	0.276	0.616***
P-value: AMI=0								0.131	0.417	0.004
R^2	0.282	0.155	0.553	0.720	0.197	0.558	0.721	0.199	0.559	0.721
N	913	913	913	913	913	913	913	913	913	913

Note: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. The cluster standard errors at the year of birth and the region of residence at age 18 are in parentheses. We restrict to individuals who were born from 1935 to 1958 in the former territories of Russian Empire in Lithuania and Poland. In *Panel (I)* we report the estimated coefficient γ_2 from Equation 1.4.1, and *AMI* from Equation 1.4.2. In *Panel (II)* and *Panel (III)*, we report the estimated coefficient α^f, α^m from Equation 1.4.4, and *AMI* from Equation 1.4.8. All regressions control for constant, a four categories of health at age 10, a mental health problem dummy at age 10, to be a good student at math dummy; five place of birth dummies: a big city, the suburbs of a big city, a large town, a small town or rural area; the features of the individual's dwelling at age 10: the number of books by age 10, the number of services, the number of rooms, and the year fixed effects.

Despite education being an outcome of the regime, it is also an important predictor of labor participation. Next, we include it as a control variable. Columns 5 - 7 show the results. In this case, we include education as an independent variable in the work model, so the associated coefficient with education captures partially the impact of the USSR itself. Indeed, looking at the subsample of women and men (*Panel II and III*), the impact of the USSR on the employability (Column 7) gets smaller comparing with Column 4. Moreover, there is a different pattern across genders. Women with secondary and high education accumulate more years of experience from 25 to 50 compared with low educated; coefficients for education are statistically significant at 1 percent level. The findings remain the same if we control for years of education (see Table A.9 in Appendix 1.13.1). It means that either women find easier a job once they acquired education; or women who decided to take more education, want to participate in the labor market. In our analysis, we cannot disentangle these two channels. However, there is no association between men's education and work experience between 25 to 50. It is in line with an idea that work is not a choice for men, and they need to work regardless of schooling.

Since the Soviet Union changed the acquired schooling, we account for the heterogeneity of the impact of the USSR on experience across educational attainment. Columns 8 - 10 show the results. In this case, the AMI increases even further for men and women comparing with the estimate in which we ignored the potential heterogeneity (Column 2 - 7). On average, women in the USSR accumulated almost 3 extra years of experience by age 50 compared with women in the Eastern Bloc. The impact for men is significant only when we abstract from early-life trade-off.

Next, we repeat the same analysis as above, but following the identification strategy similar to Fuchs-Schündeln and Masella (2016). Instead of controlling for the USSR dummy, we now include each region identifier to isolate any regional differences. Table A.12 in Appendix 1.13.3 reports the results for the full sample. As we see, the magnitudes of all coefficients are similar to *Panel I* in Table 1.4. It confirms no systematic differences across regions, along with acquired education and work history. It is not a formal test, but since including region fixed effects do not change our conclusion about the coefficients; in the rest of this article, we only control for the USSR variable.

1.5.2 Model of Labor and Schooling Decision in the USSR

Why did the Soviet Union have a stronger impact on women's education and work experience? The first channel is related to the work enforcement: zero unemployment policies, work propaganda, and other macroeconomic factors specific to the USSR. The second channel is through education availability, in particular, making it free and universal. The results in Table 1.4 suggest that these first two channels are less strong for men. Lastly, the third gender-specific channel regards the indirect impact of the USSR on women's experience through higher incentives to study in the presence of future rights and obligation to work.

To see the third women-specific mechanism, we can build a model of educational choice. For simplicity, let us assume that, at the moment when a woman makes a binary decision about schooling, *educ*, she also forms beliefs about her future possibility to be employed, π. Likely, women in the USSR had higher beliefs about the probability to find a job because of the stronger power of the state to enforce the announced policies: $\pi_{USSR} > \pi_{Bloc}$. Equivalently women in the Eastern Bloc could form higher beliefs about the possibility not to participate, i.e., to remain housewives, compared with women in the USSR. Moreover, assume that the chances to find a job differ between these two states due to macroeconomic factors. The zero unemployment policy and the planned mechanism of getting jobs were better enforced in the USSR that in the Eastern Bloc. Let us denote the future employment rate as ϕ, and $\phi_{USSR} > \phi_{Bloc}$. For simplicity, we can assume that this probability to find a job does not change with the education level. Then, her expected income becomes $\pi\phi \cdot w(educ) + \pi(1-\phi) \cdot h + (1-\pi) \cdot h$, where $w(educ)$ is the labor income, $w(1)$ or $w(0)$, h is the utility from staying home (spouse wage, unemployment benefits, other source of utility). Consider that the costs of education, $c(educ)$, are increasing in the level of education, $c(1) > c(0)$. Since the USSR government tried to achieve free and universe education, it is reasonable to assume that this education cost is relatively smaller in the USSR than in the Eastern Bloc: $c_{USSR}(1) - c_{USSR}(0) < c_{Bloc}(1) - c_{Bloc}(0)$.

Then, a woman in the USSR decides to study if

$$\pi_{USSR}\phi_{USSR} \cdot (w(1) - w(0)) \geq c_{USSR}(1) - c_{USSR}(0).$$

On the other hand, a woman in the Bloc chooses schooling if

$$\pi_{Bloc}\phi_{Bloc} \cdot (w(1) - w(0)) \geq c_{Bloc}(1) - c_{Bloc}(0).$$

Here, we implicitly assume that the relative wage range is the same across the USSR and the Eastern Bloc, we do it because both countries use the wage scale. We get that the wage premium of education is the same in both countries, but what varies the beliefs about being employed is the probability to find a job and the cost of education. According to this model, if a woman forms high beliefs about her job opportunities, π, then she studies more. So, this simple model explains why schooling became more attractive for women in the USSR, once the employment was less a choice but an obligation later in life. For men, this indirect channel is not present because for them π is equal to 1 independently of education; so, the only two channels that matter for men are the direct channel of work enforcement, ϕ, and the direct educational channel measured by relative costs, $c(1) - c(0)$. Appendix 1.13.4, Fig. A.11 illustrates the proposed mechanism.

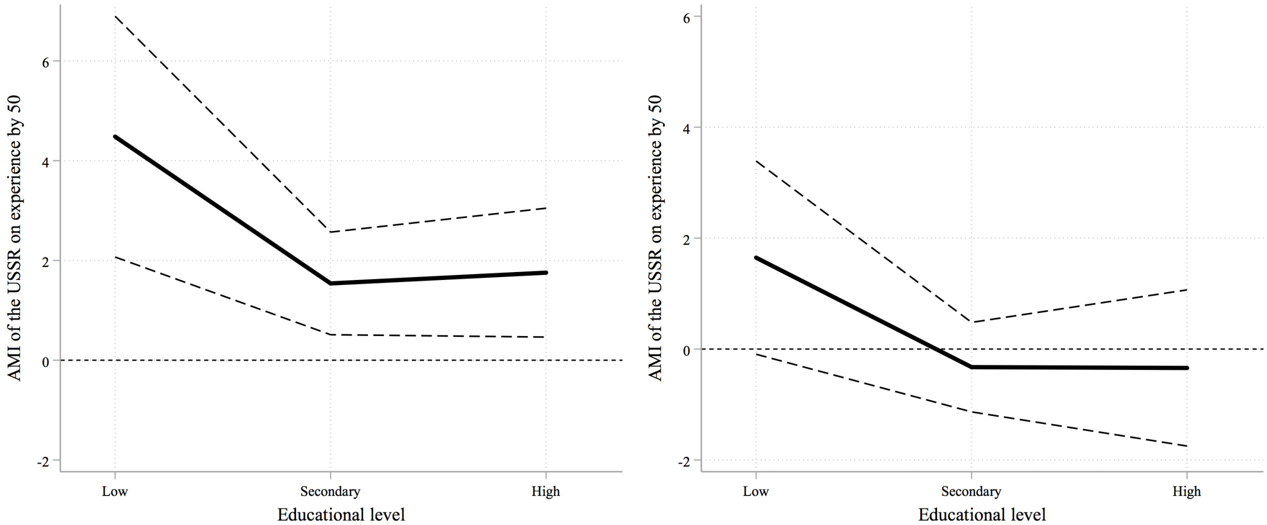
We are not the first who claim this positive impact of communism on the education and women's participation. However, up to our knowledge, we contribute to this literature by studying individual choices during the Soviet Union and showing directly three channels of the Soviet regime on education and experience. The closest article for that regard is Campa and Serafinelli (2019) in which they show that gender role attitudes and the importance of career success significantly differ among East and West German women. Regarding the impact of communism on the education in East Germany, Fuchs-Schündeln and Masella (2016) point out that the Soviet governments promoted free education to all citizens, in part because schools were the perfect place to implement propaganda.

1.5.3 Heterogeneity with Education

In this Section, we show the heterogeneity of the impact of the USSR on experience across education groups.

First, we report how the USSR's impact on cumulative work experience by age 50 depends on education for men and women. Fig. 1.3 shows the AMI of the USSR on experience by age 50 and the 95 percent confidence interval. Fig. 1.3a confirms that the USSR has a positive and significant impact on women's participation, and this effect is larger for women with lower and secondary education. On average, women with low-education accumulate more than 4 years of experience at the end of their life comparing to those who were born in the Eastern Bloc. This group accounts for about one-fourth of all women in the USSR (see Fig. A.8). The impact for low-educated men is also the largest (see Fig. 1.3b), which means that the USSR regime brought the least educated men to the labor market, i.e., by reducing their unemployment span. In Appendix 1.14, we report the results for work experience from age 25 to 50, and we see that the USSR impacted the most the low and secondary education groups (see Fig. A.12). Additionally, we repeat the same estimate using seven education levels; Fig. A.13 confirms the results, even though looking at more refined groups makes estimates noisier.

Fig. 1.3. AMI of the USSR on Experience by Age 50 Across Education by Gender



(a) Women **(b)** Men

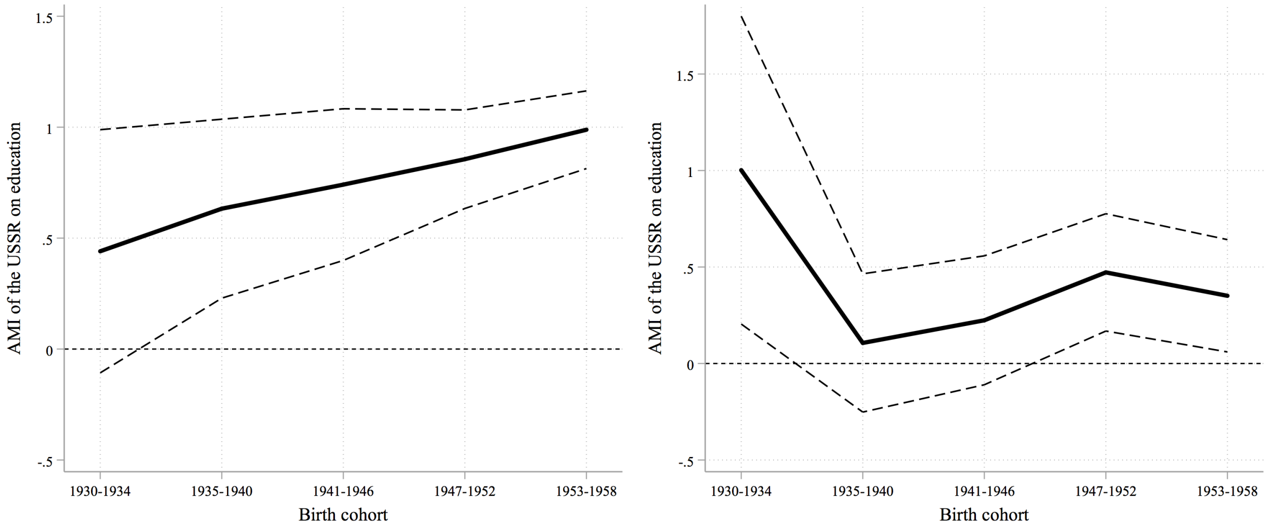
1.5.4 Intensive Margin of the USSR Impact

So far, we studied the intensive margin of the treatment, whereas in this Section, we look at the intensive margin. How do the results vary across birth cohorts? In the main analysis, we focus on the respondents born after 1935, so that they made decisions during communism. Only in this Section, we add a cohort from 1930 to 1934 to verify potential cofounders in the pre-communist era.

In terms of acquired education, the impact increases over the 20th century. Women who were born in the USSR from 1953 to 1958 accumulate, on average, one level more of education (see Fig. 1.4a); the impact is smaller for men than for women, but for the former it also gets larger among recent cohorts (see Fig. 1.4b).

Employability follows the opposite pattern across the birth cohorts. Our hypothesis about lagged enforcement of the Eastern Bloc regime compared with the USSR should lead to a stronger impact of the Soviet Union among individuals who were born early. Fig. 1.5a confirms it as a cohort from 1935 to 1940 work, on average, almost four additional years by age 50. For men, there is no significant impact on cumulative years of experience at 50, as shown in Fig. 1.5b. In Appendix, Fig. A.15a shows similar findings for experience from age 25 to 50 in the case of women. On the other hand, we see a slightly significant impact of the USSR on low educated men's employability in Fig. A.15b. Male work enforcement is better described in the case of experience between ages 25 and 50, since this removes the early-life trade-off between

Fig. 1.4. AMI of the USSR on Education Across Birth Cohorts by Gender



(a) Women **(b)** Men

work and study.

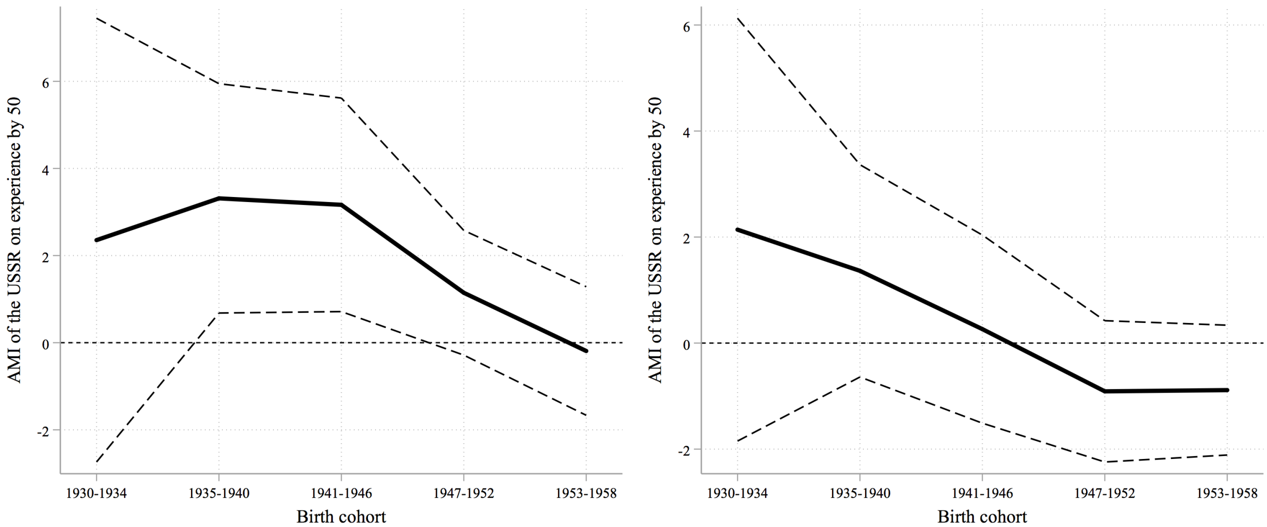
1.5.5 USSR, Marriage Choices, Reproductive History and Later-Life Outcomes

We already saw that the Soviet Union brought more women to the labor market but an increase in participation could have consequences in the family structure. In this Section, we study if the Soviet regime has impact on marriage, fertility and later-life satisfaction.

First, we discuss the marriage outcomes in Columns 1 and 2 in Table 1.5. *Panel I* shows that the USSR coefficient is always statistically significant and increases the probability of marrying over life and the number of legal marriages, Column 1 and 2, respectively. On average, people in the USSR are more likely to marry by 0.05 during life, mainly driven by men (see Column 1 in *Panel III*). This evidence is line with the Bachelor tax (also known as the tax on childlessness) that was in place in the USSR until its dissolution, but in Poland only from 1946 to 1973. For that regard, the most recent cohorts should be more affected. In Appendix 1.15.2, Fig. A.16 exactly confirms that the AMI of the USSR for men increases for the more recent cohorts, 1953-1958. However, the more extended implementation of the Bachelor tax in Lithuania than in Poland is only one possible explanation.

Despite that the chances to be married during the life did not change for women in Lithua-

Fig. 1.5. AMI of the USSR on Experience by 50 Across Birth Cohorts by Gender



(a) Women **(b)** Men

nia, the number of marriages during life increased by 0.07 compared with Poland (Column 2).[18] This difference can be the result of women's empowerment, better education, changes in the divorce law, or the social norms about remarrying. All these factors ease the termination of an unhappy relationship. Our analysis is agnostic about the causality between any policy in the USSR and marriage choice. However, we document the correlation between simultaneous changes in labor participation and the higher probability of quitting the marriage and remarrying.

Next, we consider the fertility outcomes, starting with the age at first birth (see Column 3). The USSR coefficient is not significant, but the female-specific impact is positive and significant, meaning that women in Lithuania were older, on average, at the moment of the first birth than women in Poland. The AMI of USSR on women is almost 1 year. When we control for education (as a level or years of education) and allow for heterogeneity of the impact, the USSR coefficient becomes smaller, 0.75, but still significant at 1 percent level.[19] It means that the institutional norms imposed in the USSR, made women to postpone the moment of delivery. However, we do not see any impact on men. This gender-difference in the response can be due to unequal distribution of child bearing and caring time among parents. Regarding the positive and significant impact of the USSR on women is partially explained

[18]The chances to be married is measured as a binary variable (Column 1), whereas the number of marriages takes only positive integer values (Column 2).
[19]The results are available upon request.

Table 1.5: USSR and Marriage History, Fertility and Later-Life Well-Being

	Marriage and fertility history					
	(1)	(2)	(3)	(4)	(5)	(6)
Variables	Ever-married	Number of marriages	Age delivery	Number of children	Life satisfaction	Life quality
Panel I: Both men and women						
Female	0.0235	0.00360	-3.065***	0.0529	-0.220	-1.320***
	(0.0173)	(0.0121)	(0.248)	(0.0747)	(0.137)	(0.447)
Female × USSR	-0.0307	0.0231	1.419***	-0.0869	0.160	1.341**
	(0.0212)	(0.0255)	(0.349)	(0.108)	(0.206)	(0.611)
USSR	0.0537***	0.0565***	-0.558**	-0.161*	-0.605***	-3.165***
	(0.0189)	(0.0209)	(0.274)	(0.0857)	(0.168)	(0.476)
AMI of the USSR on women	0.023	0.080***	0.861***	-0.248***	-0.445***	-1.824***
P-value: AMI=0	0.202	0.000	0.001	0.002	0.004	0.000
R^2	0.0519	0.0699	0.201	0.137	0.0884	0.132
N	2163	2087	1990	1990	2172	2135
Panel II: Women						
USSR	0.0193	0.0731***	1.014***	-0.274***	-0.342**	-1.589***
	(0.0200)	(0.0180)	(0.255)	(0.0868)	(0.159)	(0.530)
R^2	0.0860	0.0733	0.138	0.164	0.125	0.167
N	1268	1227	1180	1180	1265	1245
Panel III: Men						
USSR	0.0547***	0.0568***	-0.624**	-0.128	-0.780***	-3.469***
	(0.0190)	(0.0217)	(0.282)	(0.0884)	(0.180)	(0.510)
R^2	0.0976	0.208	0.134	0.189	0.130	0.146
N	895	860	810	810	907	890

Note: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. The cluster standard errors at the year of birth and the region of residence at age 18 are in parentheses. We restrict to individuals who were born from 1935 to 1958 in the former territories of Russian Empire in Lithuania and Poland. We consider only children born before 1990. In *Panel (I)* we report the estimated coefficient γ_2 from Equation 1.4.1, and AMI from Equation 1.4.2. In *Panel (II)* and *Panel (III)*, we report the estimated coefficient α^f, α^m from Equation 1.4.4. All regressions control for constant, a four categories of health at age 10, a mental health problem dummy at age 10, to be a good student at math dummy; five place of birth dummies: a big city, the suburbs of a big city, a large town, a small town or rural area; the features of the individual's dwelling at age 10: the number of books by age 10, the number of services, the number of rooms, and the year fixed effects.

We also document a statistically significant decrease in the number of children by 0.14. This findings consider only children who were born before 1990 to isolate the impact of the transition period. There is no evidence about the intensive margin of the impact of the Soviet Union on the marriage and fertility history.[20] What mainly matters is the extensive margin of living in the USSR.

Further, we check the impact on later-life well-being. There is evidence for overall disappointment about living in Lithuania than Poland. Columns 5 and 6 show that in the pooled analysis, the USSR coefficient is statistically significant at 1 percent level and is equal to -0.70

[20]The results are available upon request.

and -3.23 for life satisfaction and life quality, respectively. Both magnitudes get even larger when we restrict to men (see *Panel III*). So far, in our analysis, we always abstract from the transition period in Lithuania and Poland after the fall of the USSR, as we were able to restrict to events that happened before 1990. For what concerns the later-life well-being, the data do not allow us to do that. Accordingly, we cannot disentangle the impact of living in the USSR (and not the Eastern Bloc) with the difference in the transition period between Lithuania and Poland. In terms of GDP per capita (PPP, constant 2017 international $) as a proxy for economic conditions in two countries after 1990, we see that in 1995 GDP per capita was lower in Lithuania, but a country had higher economic growth in the next years and from 2003 it leveled and exceeded the GDP per capita in Poland (see Appendix 1.16, Fig. A.18). It implies, that the economic growth was higher in Lithuania and it is unlikely to explain the lower life satisfaction of individuals. However, Guriev and Zhuravskaya (2009) identify the other potential channels for unhappiness in East European countries, like human capital depreciation, higher inequality, deterioration of public goods; any of it could be a confounder in the USSR coefficient. Accordingly, with our study we highlight the important difference within the transition economies but we leave open its origin.

To conclude this Subsection, we want to make sure that area differences do not drive our findings. We exclude the overall USSR impact and control for region identifiers. Table A.13 in Appendix 1.13.3 confirms findings in *Panel I* in Table 1.5.

1.6 Threats for Identification

The potential threat to our identification assumption can be if the two countries had different exposure to WWII. Here we discuss each concern separately and how we deal with it.

During WWII. When the combats began in 1939, the almost entire population got mobilized if men went to the front, women got mainly involved in war-related activities in hometowns. Accordingly, due to the absence of men population, a large positive female labor demand shock could create a mechanical impact on the cumulative years of experience. To isolate this problem, we consider only individuals who were born after 1935, and the oldest in our sample was only 4 years old when the war began. Accordingly, the reported years of work experience are always after the war. Moreover, Schweitzer (1980) shows that in the US context, this upward shift in female employment lasted only during WWII, and shortly after the war, the female employment went back to the pre-war period.

WWII could also destruct the process of human capital formation and lead to massive breaks in education. We expect that it is less of a concern in our case as most of the respondents were born in the postwar period. For what regards the impact on more recent cohorts, up to our knowledge, the closest article is Ichino and Winter-Ebmer (2004). They show that there is a negative impact of WWII on the educational attainment of individuals born in 1930 - 1939 in more affected European countries, but individuals born in 1940 - 1949 do not experience any loss in schooling. Accordingly, we believe that the direct impact of WWII on education does not affect our findings.

Men Shortage After WWII. If Lithuania is more affected by WWII in terms of men's loss, then higher work experience among Lithuanian women can be due to it. To verify this hypothesis, given the lack of comparable post-war statistics and official data about war destruction for Lithuania and Poland, we closely follow the idea of Becker et al. (2020a, Online Appendix, Section VI.B). We exploit the Life in Transition Survey III conducted by the European Bank for Reconstruction and Development (EBRD) in 2016 and consider Question 9.24a: *Were you, your parents or any of your grandparents physically injured or were your parents or any of your grandparents killed during the Second World War?*.[21] We restrict to respondents who were born from 1935 to 1958 as we do in the main analysis using the SHARE data. In Appendix 1.17, Table A.14 shows that, on average, about 17 percent of respondents who currently reside in Lithuania report to have close relatives who were direct victims of WWII, and this percent triples among respondents from Poland.[22] Accordingly, we find supportive evidence that men's losses were large in Poland and not in Lithuania, so women in Poland could face less competition on the labor market than women in Lithuania. It implies that in our analysis, we find a lower bound of the USSR coefficient.

Differential Capital Destruction due to WWII. Up to our knowledge, there is no comparable data between Lithuania and Poland about industrial stock losses due to WWII. However, even if those statistics would be available, it is hard to claim the direction of the overall impact. On the one hand, if Poland experienced higher industry destruction than Lithuania, it could lead to an immediate drop in the labor attachment due to the lack of jobs and the decrease in the utility of working. Accordingly, it could explain the lower education and work experience among Poles. On the other hand, the higher industry destruction leads to job creations in eco-

[21] Data: `https://www.ebrd.com/what-we-do/economic-research-and-data/data/lits.html`

[22] This impact is statistically significant when we isolate birth fixed effects. Similar to Becker et al. (2020a) in Online Appendix, Section VI.B, we run a regression of having a relative directly affected by WWII on the dummy being in Poland and the set of birth fixed effects and female dummy, then the Poland coefficient is equal to .435 and statistically significant at 1 percent level. We include only the regions that were former territories of the Russian Empire and cluster standard errors at the primary sampling unit.

nomic recovery sectors in the short run. Accordingly, our USSR coefficient will be downward biased if Poland is more affected by WWII.

The intensive margin analysis documents the increase in female's work experience by age 50 or between ages 25 and 50 is driven by cohorts from 1935 to 1946. Accordingly, the youngest individual was 25 years old in 1960, which is more than 10 years after WWII. By that year, we expect that the immediate negative impact that vanished out and remained recovery channel should lead to a downward biased USSR coefficient. Yet, we want to admit that our argument is informal, and we cannot rule out the other scenario.

Demographic Composition After WWII. The other explanation for the difference between the two regimes that we find can be confounded with the two countries' differential demographic imbalances. Brainerd (2017) documents evidence about gender imbalances in Russia after WWII. Our identification strategy fails not if there is an imbalance in two countries due to WWII but only if there is a differential composition across two countries. Accordingly, we plot the sex ratio (men over women) for cohorts under the analysis using data from the Soviet Census 1959 and the Polish Statistical Yearbook 1955. Fig. A.19a and A.19b show results for Lithuania and Poland, respectively. We document that there is a shortage of men among early cohorts within countries, but cohorts under the study experience a balanced distribution. Across countries, there is a notable imbalance among cohorts from 1935 to 1940 in Lithuania than Poland. Accordingly, we repeat our analysis restricting to cohorts from 1941 to 1958, and all our findings hold.

In terms of life expectancy, two countries follow very similar patterns (see Fig. A.20a and A.20b for women and men, respectively). Women born from 1960 to 1990 in Lithuania have longer life expectancy than in Poland. Similarly, men born before 1975 in Lithuania expect to live longer than the ones in Poland. Accordingly, there is an unlikely shortage of labor forces in Lithuania comparing with Poland due to differential life expectancy.

Differential Out-Migration During Communism. If the out-migration was larger in Lithuania than in Poland, then the higher number of years of work experience among Lithuanian women could be due to lower labor market competition. We believe that it was not the case, given that the USSR almost abolished immigration and emigration abroad Light (2012). Regarding internal immigration, Lithuania was able to preserve its own identity and keep the dominant ethnic majority Eberhardt (2003). Likewise, the out-emigration from Poland was tiny after 1951 (Becker et al., 2020a, footnote 14).

Differential Out-Migration After 1990. After the fall of the Iron Curtain it was a mass wave of migrants from Lithuania and Poland. If the profile of individuals who left Lithuania and Poland were different then it could bias our results. To check that it was not the case, using the SHARE dataset we find migrants from Lithuania and Poland who currently reside in the other European country. Then, we compare their profiles along with control variables used in our analysis. We do not find any statisctically significant difference among them (see Table A.15 in Appendix 1.17.4).

1.7 Robustness Checks

In this Section, we show that our findings are robust to several tests. First, in Section 1.7.1, we show that our specification is robust to movements during life. Then, in Section 1.7.2 we consider the interwar borders. In Section 1.7.3, we amplify the sample to rule out that the results are driven by only Lithuania versus the part of Poland comparison. Section 1.7.4 reports the placebo analysis.

1.7.1 Movements During Life

So far, we assign the treatment based on the region in which a respondent lived at age 18. However, it might be that a person moved during life. Since we observe the full residential history, we can identify the region in which a respondent lived in any year.[23] Only half of the individuals in our target sample change the region of residence during life (see Fig. A.9 in Appendix 1.11.4). Still, we also identify the region of birth and the region in which lived the most of life. These two other definitions mainly change the value for respondents in the Eastern Bloc (i.e., $Z_i = 0$), as we consider only a part of Polish regions, and during the Eastern Bloc period, immigration within Poland was not restricted. Meanwhile, the value for respondents in Lithuania almost did not change because of controlled overseas migration during the regime.[24]

In Appendix 1.18.1, Table A.16 reports the results for education and work experience using two new USSR variables (*Panel II* and *Panel III*). Qualitatively, all the coefficients remain unchanged; quantitatively, the magnitudes are almost identical and within the one standard deviation interval of the original findings.

[23]By the survey's construction; we can observe only individuals who resided in one of the EU countries at the moment of the survey.

[24]Almost 50 percent of our sample never changed the region of residence. So, individuals in our target sample do not move much during their lives in line with our intuition.

1.7.2 Interwar Borders

Our analysis is based on present-day geographical regions in Lithuania and Poland, but during the interwar period, the country borders were different. In this Section, we exclude the Klaipeda region because of the territorial dispute with Prussia and Vilnius, Alytus and Utena regions that belonged to Poland before WWII. In Appendix 1.18.2, Table A.17 shows that despite the loss in the number of observations, all our findings hold.

1.7.3 The Soviet Union and the Eastern Bloc

Along all the article we restrict to the territories of the Russian Empire in Lithuania and Poland to make a cleaner comparison and abstract from potential differences in the pre-communist era. In this Section, we extend our sample to study whether the Soviet ideology had a specific impact on the former territories of the Russian Empire, or whether it is an overall change in all territories exposed to the regime. If the Soviet ideology was more effectively implemented in the Soviet Union, we should see similar findings enlarging the sample. Accordingly, we consider Lithuania versus Poland setting and all the Eastern Bloc versus all the Baltic countries (former USSR). In Appendix 1.18.3, *Panel II* and *Panel III* in Table A.19 confirm the main findings. Including other regions and countries increases the sample size, but the magnitude of the USSR's total impact on women's education is always roughly 0.8 level, and the impact on women's experience by 50 gets only larger and reaches 2.4, when not accounting for schooling. Table A.20 shows that the findings of the marriage history, the number of children, and life satisfaction are also robust to enlarging the sample.

1.7.4 Placebo Analysis

Finally, we run a permutation test similar to Lippmann and Senik (2018) and Lippmann et al. (2020) in which we divide regions randomly into two groups. Group 1 includes 10 regions, and it resembles the hypothetical treated group, and Group 2 consists of 5 remained regions, the control group.[25] Next, we check how the estimate of the heterogeneous impact of the USSR on women and the AMI of the USSR on women changes due to the different composition of regions in Group 1 and Group 2. Due to the sample structure, we always have at least 5 USSR regions in the hypothetical treated group, so we expect that the USSR coefficient remains significant, even though in fewer cases. Next, we define a dummy variable that equals one if

[25] This test is similar to Table E30 in Lippmann et al. (2020).

the "Female × USSR" coefficient and "AMI of USSR on women" are statistically significant at 10, 5, or 1 percent level. We regress this dummy on the number of USSR regions in Group 1 as an independent variable using Ordinary Least Squares. Table A.21 in Appendix 1.19 shows results considering work experience between age 25 and 50: the best fit happens when we assign regions correctly. As more USSR regions included in the hypothetical treated group, as better the predictions of the model.

1.8 East and West Germany Comparison

Most of the literature on the impact of the communist regime exploit Germany's forced division after World War II. For 41 years, a country was divided into two parts East Germany (called the German Democratic Republic) was the part of the Eastern Bloc, and West Germany (called the Federal Republic of Germany) promoted the traditional male-breadwinner society. These two systems were strikingly different for what regards the women's questions and sex-role policies. Becker et al. (2020b) provide a recent review of the related literature and excellent discussion on the identification assumption and possible biases in the results. One of their worries lies in the pre-determined difference between East and West Germany before the separation, along with a set of variables. In particular, we consider labor force participation as an outcome; that is why we should be careful about the magnitude of our findings in this Section.

Up to our knowledge, no study has been done on the gender-specific impact of communism in East Germany during the regime for 1935-1958 birth cohorts. As before, we start with the simultaneous impact on the educational choice and cumulative years of experience before the regime's fall. To close the gap in the literature and validate further our findings, we strictly follow Alesina and Fuchs-Schündeln (2007) to divide German regions into two groups, defining an *East Germany* variable.[26] Table 1.6 reports the results. East German women get half a level more of education comparing with West German women (see Column 1 in *Panel I*), the same pattern holds for men, but the magnitude shrinks to 19 percent a level of education. These findings are in line with our conclusion for the target sample.

Moreover, Eastern Germans accumulate 0.4 years more of experience between 25 to 50 years. However, the impact is considerably larger when restricting to women in East Germany,

[26] *East Germany* is equal to one if a respondent was born in Brandenburg, Mecklenburg-Western Pomerania, Saarland, Saxonia, Saxonia-Anhalt, and Thuringia. It is equal to zero if Baden-Wuerttemberg, Bavaria, Bremen, Hamburg, Hesse, Lower Saxony, North Rhine-Westphalia, Rhineland-Palatinate, and Schleswig-Holstein. We leave out Berlin as we do not know if a respondent lived in the Eastern or Western part of the city.

Table 1.6: Communism, Education and Work Experience in East and West Germany

	Education	Cumulative work experience								
		No control for education			Controls for three education levels			Heterogeneity with education		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Variables		By 25	By 50	25-50	By 25	By 50	25-50	By 25	By 50	25-50
Panel I: Both men and women										
Female	-0.464***	-0.0282	-6.272***	-6.244***	-0.466***	-6.447***	-5.981***			
	(0.0528)	(0.164)	(0.487)	(0.443)	(0.151)	(0.478)	(0.431)			
Female × East Germany	0.312***	0.556*	5.481***	4.925***	0.780***	5.440***	4.661***			
	(0.108)	(0.290)	(0.632)	(0.532)	(0.286)	(0.643)	(0.526)			
East Germany	0.182**	-0.748***	-0.320	0.428*	-0.503**	-0.125	0.378			
	(0.0734)	(0.215)	(0.337)	(0.255)	(0.203)	(0.341)	(0.258)			
Education:										
Secondary					-0.232	1.712**	1.944***			
					(0.239)	(0.717)	(0.618)			
High					-2.907***	-0.867	2.040***			
					(0.275)	(0.726)	(0.626)			
AMI of East Germany on women	0.494***	-0.192***	5.161***	5.352***	0.276***	5.315***	5.039***			
P-value: AMI=0	0.000	0.373	0.000	0.000	0.204	0.000	0.000			
R^2	0.247	0.165	0.444	0.515	0.294	0.463	0.520			
N	2240	2241	2241	2241	2240	2240	2240			
Panel II: Women										
East Germany	0.477***	-0.174	5.093***	5.268***	0.246	5.026***	4.780***	-0.198	3.497	3.695
	(0.0783)	(0.212)	(0.503)	(0.454)	(0.216)	(0.508)	(0.446)	(1.204)	(2.437)	(2.351)
Education:										
Secondary					-0.0865	1.705*	1.792**	-0.0433	1.550	1.594*
					(0.276)	(0.892)	(0.780)	(0.281)	(0.946)	(0.823)
High					-2.693***	0.496	3.189***	-2.968***	0.428	3.396***
					(0.349)	(0.966)	(0.845)	(0.379)	(1.045)	(0.926)
Education × USSR:										
Secondary × East Germany								0.130	1.728	1.598
								(1.202)	(2.480)	(2.407)
High × East Germany								1.034	1.441	0.407
								(1.257)	(2.495)	(2.381)
AMI of East Germany	0.477***	-0.174	5.093***	5.268***	0.246	5.026***	4.780***	0.173	4.885***	4.712***
P-value: AMI=0								0.516	0.000	0.000
R^2	0.280	0.149	0.271	0.293	0.264	0.277	0.306	0.268	0.277	0.307
N	1158	1158	1158	1158	1158	1158	1158	1158	1158	1158
Panel III: Men										
East Germany	0.199***	-0.819***	-0.298	0.520***	-0.528***	0.0934	0.622***	0.371	0.154	-0.217
	(0.0743)	(0.224)	(0.288)	(0.133)	(0.199)	(0.257)	(0.133)	(0.734)	(1.082)	(0.693)
Education:										
Secondary					-0.994**	-0.546	0.448	-0.684	-0.296	0.388
					(0.419)	(0.690)	(0.435)	(0.486)	(0.819)	(0.527)
High					-3.733***	-4.407***	-0.674	-3.646***	-4.591***	-0.945*
					(0.436)	(0.719)	(0.458)	(0.503)	(0.852)	(0.552)
Education × USSR:										
Secondary × East Germany								-1.380*	-0.913	0.467
								(0.792)	(1.122)	(0.701)
High × East Germany								-0.483	0.825	1.308*
								(0.791)	(1.164)	(0.727)
AMI of East Germany	0.199***	-0.819***	-0.298	0.520***	-0.528***	0.0934	0.622***	-0.548***	0.051	0.599***
P-value: AMI=0								0.005	0.836	0.000
R^2	0.196	0.212	0.709	0.888	0.354	0.761	0.895	0.357	0.763	0.895
N	1082	1083	1083	1083	1082	1082	1082	1082	1082	1082

Note: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. The cluster standard errors at the year of birth and the region of residence at age 18 are in parentheses. We restrict to individuals who were born from 1935 to 1958 in Germany excluding Berlin. East Germany is equal to one if a respondent was born in Brandenburg, Mecklenburg-Western Pomerania, Saarland, Saxonia, Saxonia-Anhalt, and Thuringia. It is equal to zero if Baden-Wuerttemberg, Bavaria, Bremen, Hamburg, Hesse, Lower Saxony, North Rhine-Westphalia, Rhineland-Palatinate, and Schleswig-Holstein. In *Panel (I)* we report the estimated coefficient γ_2 from Equation 1.4.1, and AMI from Equation 1.4.2. In *Panel (II)* and *Panel (III)*, we report the estimated coefficient α^f, α^m from Equation 1.4.4, and AMI from Equation 1.4.8. All regressions control for constant, a four categories of health at age 10, a mental health problem dummy at age 10, to be a good student at math dummy; five place of birth dummies: a big city, the suburbs of a big city, a large town, a small town or rural area; the features of the individual's dwelling at age 10: the number of books by age 10, the number of services, the number of rooms, and the year fixed effects.

and the AMI reaches about 5.2 years of work experience by age 50 before the fall of the regime (see Columns 3 and 4 in *Panel I*). Becker et al. (2020b) alarm about the possible upper bias in the women's labor participation due to preexistent trends before the forced separation, but five additional years of experience are unlikely only due to the pre-trend.

The East Germany coefficient is twice larger than the USSR estimate representing the divergence between the leading ideas in West and East Germany, compared with the treatment intensity in the USSR (Lithuania) and the Eastern Bloc (Poland). The leading regime in West Germany indeed favored a one-bread winner family structure, and women were encouraged to participate exclusively in domestic production. All our findings of women's experience become twice larger, looking at East Germany (see *Panel II*). Naturally, the impact on men almost does not change from the previous setting (see *Panel III*).

In Appendix 1.20, Table A.22 reports the results about marriage and fertility history before 1990. The communist regime did not impact total fertility among individuals from 1935 - 1958 birth cohorts.[27]

1.9 Conclusion

In this paper, we find a significant impact of the Soviet Union on schooling, labor decisions, demographic choices, and later-life outcomes during the regime. Our analysis uses the recently available retrospective SHARELIFE data (2017) and the SHARE Job Episode Panel. The target sample includes individuals born from 1935 to 1950 and their individual choices from 1950 to 1990.

Our identification exploits that Lithuania became a part of the USSR and not the Eastern Bloc due to exogenous factors unrelated to the outcome variables relevant to this study. The treatment is being in the USSR (Lithuania) and not in the Eastern Bloc (Poland). The identification riles on the similarity between the former territories of the Russian Empire in Lithuania and Poland. One of the distinguishing features of the Soviet ideology is egalitarian gender policies and full employment target. Being in the USSR made Lithuanians increase individual educational attainment and cumulative work experience.

[27] This finding corresponds with birth events before 1990 among individuals from 1935-1958 birth cohorts. Accordingly, it does not contradict with Goldstein and Kreyenfeld (2011), in which they look at total fertility rate after 1980 (based on Human Fertility Database, 2011) and in the lack of data for more recent years rely on cohort projection. Boelmann, Raute, and Schönberg (2020) also exploit different years and birth cohorts. Indeed, if we follow the strategy in Goldstein and Kreyenfeld (2011) but with newly available data we find that total fertility was similar among birth cohorts under our consideration (see Fig. A.21, in Appendix 1.20.

Moreover, we document the underlying gender-specific channel that increased women's participation in the USSR through the higher educational incentives, as can be referred to as a cascade impact of the schooling availability and work opportunities on women's labor participation. Next, we find a higher number of marriages during life and decrease in the fertility. Finally, there is evidence about lower life satisfaction about living in the USSR than in the Eastern Bloc.

Apart from studying the Soviet Union's impact on individual choices directly, we also want to exploit the unique environment created due to the regime. Nowadays, there are still some countries struggling to bring more women to the labor market. In this paper, we want to highlight one of the Soviet Union's results: the combination of educational and job opportunities is necessary to attract more women to the labor market. By any means, we do not claim that it is beneficial for a country to copy the same economic-political system from the USSR or that the Soviet Union policies are necessary and unique to achieve this goal. However, the sole availability of education and vague job opportunities likely generate a smaller boost in women's participation. We argue that it is essential that future policymakers consider increasing the efficiency of future policies to promote working choices among women.

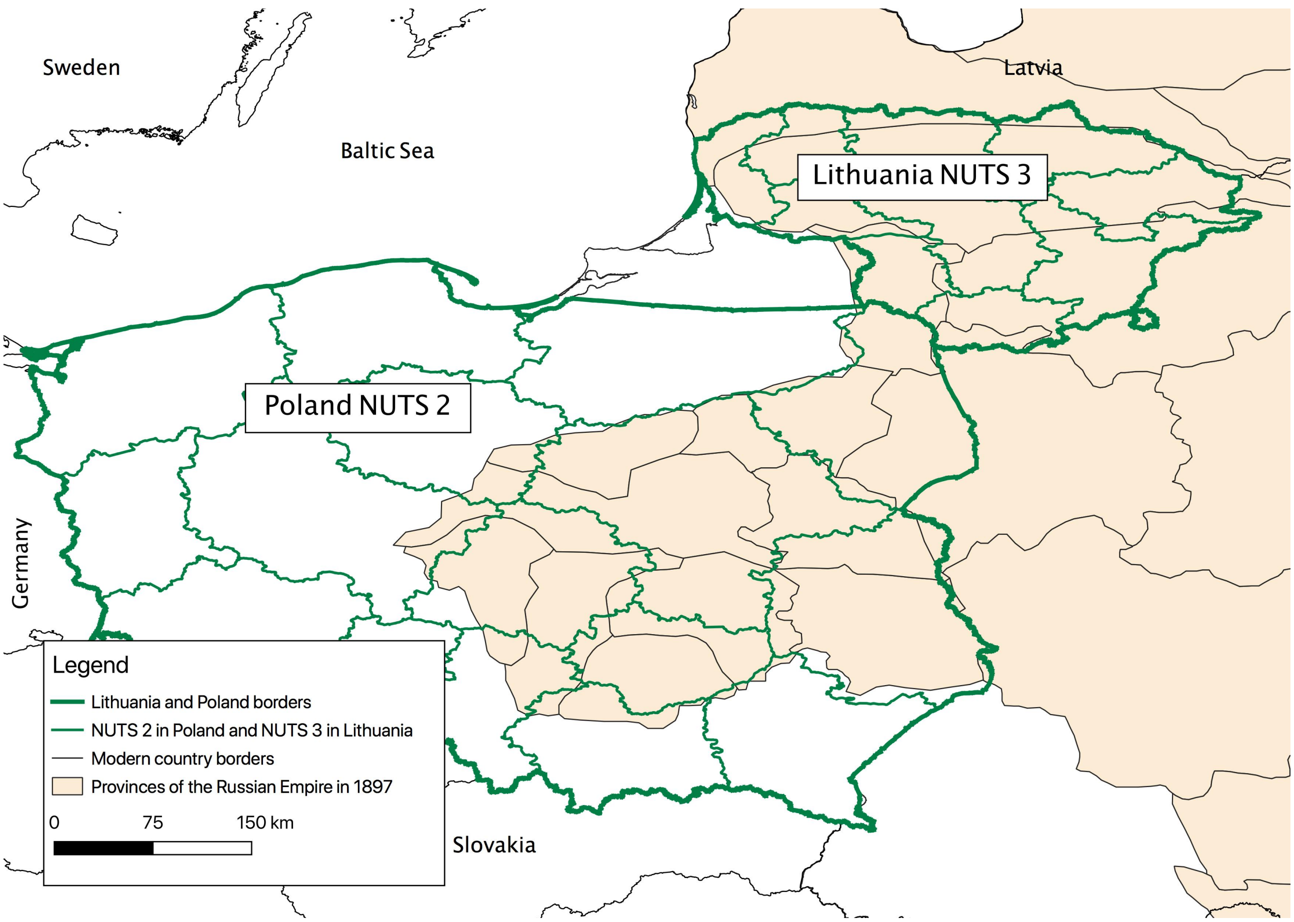
This paper's other important implication is the critical distinction between the Soviet exposure between Lithuania and Poland for more than 40 years. Up to our knowledge, this notable difference in the political-economic regime is little pronounced in the literature, and often, researchers pool all Eastern European countries together for what regards the Soviet inheritance. Within post-Soviet states, Baltic countries were part of the Soviet Union, along with other countries that formed the Eastern Bloc. We argue that contemporaneous policies should be tailored to the historical context and not ignore communism's difference in these countries.

Appendix A

1.10 Appendix: Additional Evidence on Similarity Between Lithuania and Poland

1.10.1 Map of the Russian Empire in 1897

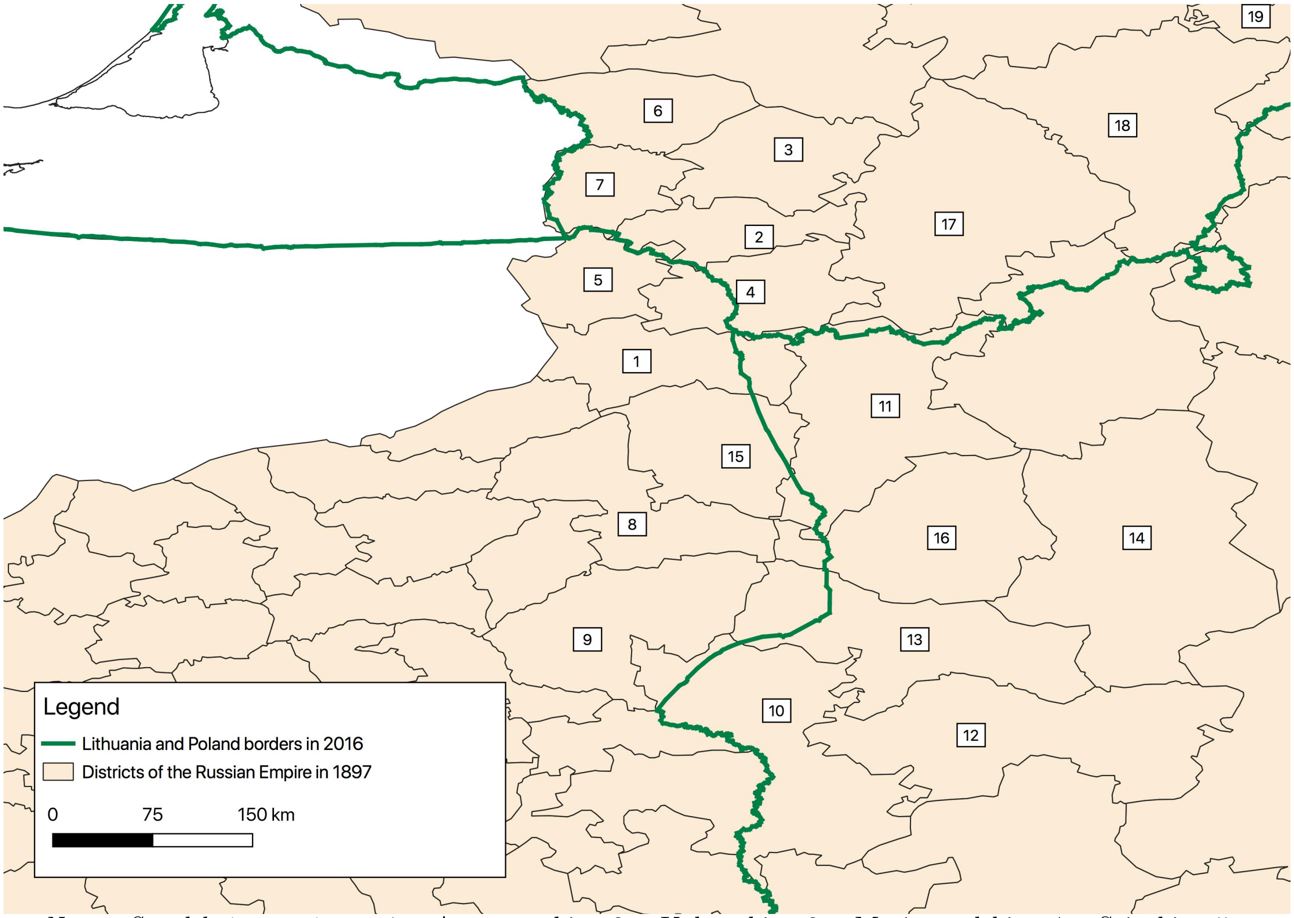
Fig. A.1. Present-Day Lithuania and Poland and the Provinces of the Russian Empire in 1897



Source: GIS map of country borders in 2016 comes from Eurostat, GISCO. GIS map of the Russian Empire by province comes from Sablin et al. (2015).

1.10.2 Relevant Districts of the Russian Empire in 1897

Fig. A.2. Illustration of the Identification Assumption



Note: *Suvalskaia province:* 1 - Avgustovskiy; 2 - Kalvarskiy; 3 - Mariampolskiy; 4 - Seinskiy; 5 - Suvalkskiy; 6 - Vladislavovskiy; 7 - Volkovyshskiy; *Grodnenskaia province:* 8 - Belostokskiy; 9 - Belskiy; 10 - Brestskiy; 11 - Grondenskiy; 12 - Kobrinskiy; 13 - Pruzhanskiy; 14 - Slonimskiy; 15 - Sokolskiy; 16 - Volkovyskiy; *Vilenskaia province:* 17 - Trokiskiy; 18 - Vilnenskiy; 19 - Zventsyanyskiy. Suvalskaia province was a part of Vistula Land in 1897, Grodnenskaia and Vilenskaia provinces were a part of Vilna Governorate-General in 1897.
Source: GIS map of country borders in 2016 comes from Eurostat, GISCO. GIS map of the Russian Empire by districts comes from Kessler and Markevich (2019).

1.10.3 Demographic and Labor Information in 1897

Table A.1: Demographic and Labor Information in 1897

Province / District (Id)[#]	Population	Square km^2	Density	Percent of[1]				Employed population Age 11−60		
				Women	Catholics	Jews	Orthodox	All	Women	Men
Panel I: Lithuania in 2016										
Vilna Governorate-General:										
Kovenskaia	1,544,564.0	40,191.2	38.4	51.3	76.4	13.8	3.0	49.6	26.5	74.8
Vilenskaia:	738,945.0	17,275.4	42.1	50.3	75.8	12.8	8.2	41.9	18.1	66.4
Sventsyanskiy (19)	172,231.0	5,228.0	32.9	51.2	78.1	7.1	9.9	34.9	12.7	59.1
Trokskiy (17)	203,401.0	5,862.3	34.7	50.2	83.8	9.5	4.5	37.8	13.2	63.1
Vilenskiy (18)	363,313.0	6,185.1	58.7	49.5	65.5	21.8	10.1	53.0	28.4	77.1
Vistula Land:										
Suvalskaia:	328,865.0	6,551.1	51.0	50.9	78.8	8.9	3.1	52.2	27.3	78.5
Kalvarskiy (2)	70,425.0	1,329.1	53.0	50.4	82.4	9.3	3.7	51.8	26.6	77.2
Mariampolskiy (3)	114,262.0	2,178.1	52.5	50.0	79.6	10.3	4.5	51.9	26.0	77.7
Vladislavovskiy (6)	67,295.0	1,774.1	37.9	51.5	80.8	7.4	1.3	51.2	28.0	76.8
Volkovyshskiy (7)	76,883.0	1,269.7	60.6	51.6	72.3	8.6	2.8	54.1	28.5	82.1
*Lithuania**	2,612,374.0	64,017.8	43.8	50.8	77.0	11.8	4.7	47.9	23.9	73.2
Panel II: Poland in 2016										
Vistula Land:										
Varshavskaia	1,931,867.0	11,336.6	170.4	49.4	71.5	18.2	5.4	56.2	28.2	83.0
Kalishskaia	840,597.0	11,336.6	74.1	50.7	82.9	8.5	1.1	48.1	21.8	75.7
Keletskaia	761,995.0	10,093.0	75.5	51.0	87.5	10.9	1.2	43.2	18.1	69.9
Liublinskaia	1,160,662.0	16,831.3	69.0	49.2	62.6	13.5	21.4	44.0	16.8	70.2
Lomzhinskaia	579,592.0	10,545.2	55.0	48.2	77.1	15.8	5.5	47.1	16.4	75.0
Petrokovskaia	1,403,901.0	12,249.4	114.6	50.3	72.9	15.9	1.6	51.5	23.7	79.8
Plotskaia	553,633.0	9,430.8	58.7	50.2	80.7	9.3	3.1	50.6	23.1	78.5
Radomskaia	814,947.0	12,352.5	66.0	50.1	83.6	13.8	1.5	44.0	16.6	71.7
Sedletskaia	772,146.0	14,317.7	53.9	49.7	66.9	15.7	15.6	41.8	13.7	69.7
Suvalskaia:	254,048.0	5,767.6	46.1	50.5	73.7	11.6	9.2	43.7	19.6	68.5
Avgustovskiy (1)	79,214.0	2,024.6	39.1	48.9	67.0	11.6	19.0	42.0	16.8	65.9
Seinskiy (4)	92,910.0	1,472.7	63.1	49.6	72.6	10.4	1.4	40.6	19.8	65.8
Suvalkskiy (5)	81,924.0	2,270.3	36.1	52.8	81.4	12.8	7.3	48.3	22.1	73.9
Vilna Governorate-General:										
Grodnenskaia:	481,601.0	9,073.2	53.2	49.2	51.0	18.6	28.8	41.7	14.8	67.8
Belostokskiy (8)	206,615.0	2,904.1	71.1	47.2	47.2	28.8	20.1	52.3	22.8	77.7
Belskiy (9)	164,441.0	3,562.2	46.2	50.5	36.5	14.9	48.3	36.9	11.4	63.8
Sokolskiy (15)	110,545.0	2,606.8	42.4	49.9	69.3	12.2	18.1	35.9	10.0	61.9
*Poland**	9,554,989.0	123,333.9	76.0	49.8	73.7	13.8	8.6	46.5	19.3	73.6

Sources: The original source of all information listed in the table is the Russian Imperial Census 1897. Most of data come from RISTAT: Electronic Repository of Russian Historical Statistics `https://ristat.org` See Kessler and Markevich (2019) for details. Data on population, square and density are taken directly from `http://www.demoscope.ru/weekly/ssp/census.php?cy=0` Data on confession at the district level are taken directly from Volume 6 (p. 9, 40, and 43) `http://istmat.info/files/uploads/15771/perepis_1897_vypusk_6.pdf`. Data on employment at the district level come from "Russian Empire Occupations in the Late 19th-Early 20th Centuries. First All-Russia 1897 Census" `http://hcod.asu.ru/en/`

[#] District Id in parentheses corresponds with Fig. A.2.

[1] Percent relates to total population in the province (guberniya) or district (uezd).

* Information about Lithuania and Poland corresponds with averages across provinces in Lithuania and Poland, respectively.

Table A.2: Demographic and Labor Information in 1897 (II)

Province	Percent of[1]			Percent of
	Age 11-60	Urban	Illiterate	Women Illiterate
Panel I: Lithuania in 2016				
Vilna Governorate-General:				
Kovenskaia	65.6	9.3	58.1	58.9
Vilenskaia	65.2	12.4	71.2	77.0
Vistula Land:				
Suvalskaia	64.8	12.6	62.6	66.8
*Lithuania**	65.2	11.4	64.0	67.6
Panel II: Poland in 2016				
Vistula Land:				
Varshavskaia	66.3	43.8	60.9	64.6
Kalishskaia	62.6	13.8	72.1	73.6
Keletskaia	62.9	9.2	77.3	80.4
Liublinskaia	65.2	13.9	76.2	81.8
Lomzhinskaia	65.0	12.9	70.4	76.3
Petrokovskaia	64.4	36.4	69.1	72.9
Plotskaia	63.3	15.9	66.5	67.8
Radomskaia	63.2	12.3	77.7	80.3
Sedletskaia	64.9	15.2	69.1	73.8
Suvalskaia	64.8	12.6	62.6	66.8
Vilna Governorate-General:				
Grodnenskaia	65.4	15.9	70.8	80.4
*Poland**	64.3	18.9	71.0	75.2

Sources: The original source of all information listed in the table is the Russian Imperial Census 1897. Most of data come from RISTAT: Electronic Repository of Russian Historical Statistics `https://ristat.org` See Kessler and Markevich (2019) for details.

[1] Percent relates to total population in the province (guberniya) or district (uezd) but information about women illiterate. The percent of "Women Illiterate" is computed with respect to only women.

* Information about Lithuania and Poland corresponds with averages across provinces in Lithuania and Poland, respectively.

1.10.4 Labor Information in 1897

Table A.3: Labor Information in 1897

Province	Percent of women working in[1]								Percent of men working in[1]							
	Undefined	Capital owners	Sellers	Agric.	Manufac.	Services	Other	100 %	Undefined	Capital owners	Sellers	Agric.	Manufac.	Services	Other	100 %
***Panel I: Lithuania in 2016*:**																
***Vilna Governorate-General*:**																
Kovenskaia	20.2	2.0	2.2	49.1	9.1	15.7	1.6	100.0	7.3	1.2	2.9	60.8	9.5	13.1	5.2	100.0
Vilenskaia	17.6	2.6	3.0	34.0	11.4	28.0	3.4	100.0	5.6	0.9	2.7	60.9	10.4	13.0	6.6	100.0
***Vistula Land*:**																
Suvalskaia	13.4	1.7	0.7	60.9	4.2	17.5	1.5	100.0	5.8	1.2	1.6	60.8	7.3	18.9	4.4	100.0
Lithuania	17.1	2.1	2.0	48.0	8.2	20.4	2.2	100.0	6.2	1.1	2.4	60.8	9.1	15.0	5.4	100.0
***Panel II: Poland in 2016*:**																
***Vistula Land*:**																
Varshavskaia	17.5	3.8	3.2	15.7	14.7	42.5	2.5	100.0	9.3	1.6	5.2	27.6	20.7	27.2	8.4	100.0
Kalishskaia	16.4	2.3	1.3	46.3	6.2	26.1	1.3	100.0	7.5	1.6	2.9	58.8	14.7	11.3	3.2	100.0
Keletskaia	11.1	1.6	1.9	54.6	2.6	27.1	1.2	100.0	4.4	1.1	4.2	65.3	9.3	11.9	3.8	100.0
Liublinskaia	13.5	1.8	2.2	43.7	5.0	31.8	2.1	100.0	4.7	0.8	3.3	55.2	10.3	21.2	4.7	100.0
Lomzhinskaia	17.5	2.2	2.2	43.1	5.9	27.0	2.1	100.0	5.6	1.0	2.7	51.3	8.6	27.3	3.5	100.0
Petrokovskaia	12.8	2.2	1.7	17.2	28.9	34.3	2.9	100.0	8.1	1.6	4.3	30.8	30.6	12.9	11.9	100.0
Plotskaia	22.5	3.5	1.2	42.0	3.4	26.0	1.3	100.0	10.4	2.2	2.3	51.7	9.9	20.3	3.4	100.0
Radomskaia	14.4	1.9	2.4	47.5	3.1	29.1	1.7	100.0	5.5	1.1	3.9	59.6	12.6	12.3	5.1	100.0
Sedletskaia	14.7	2.1	1.9	43.6	3.8	31.1	2.8	100.0	4.6	1.0	3.0	60.0	10.5	15.8	5.2	100.0
Suvalskaia	13.4	1.7	0.7	60.9	4.2	17.5	1.5	100.0	5.8	1.2	1.6	60.8	7.3	18.9	4.4	100.0
***Vilna Governorate-General*:**																
Grodnenskaia	14.0	2.5	5.5	26.9	17.4	30.7	3.0	100.0	4.2	0.8	3.4	52.5	12.4	21.4	5.2	100.0
Poland	15.4	2.4	2.3	38.1	9.1	30.6	2.1	100.0	6.4	1.3	3.5	51.3	14.0	18.2	5.4	100.0

Sources: The original data source is the Russian Imperial Census 1897. Data come from RISTAT: Electronic Repository of Russian Historical Statistics `https://ristat.org`. See Kessler and Markevich (2019) for details.
[1] This percent is defined to all employed women or men respectively.

1.10.5 Factories and Industries in 1908

Table A.4: Factories' Statistics in 1894 and 1908

Province/	Number of factories		Density of factories		Number of workers		Revenue		Power per worker	
District (Id)	1894	1908	1894	1908	1894	1908	1894	1908	1894	1908
Panel I: Lithuania in 2016										
Vilna Governorate-General:										
Kovenskaia	717.0	678.0	17.8	16.9	16.1	33.2	39,484.2	75,521.6	2.4	1.6
Vilenskaia:	89.0	109.0	5.6	10.1	71.6	83.2	50,528.3	84,910.4	2.0	1.5
Sventsyanskiy (19)	12.0		2.6		9.3		18,688.5		1.2	
Trokskiy (17)	8.0	7.0	1.6	1.4	22.6	39.7	60,419.6	76,453.0	4.3	2.4
Vilenskiy (18)	69.0	102.0	12.7	18.8	39.7	43.5	72,476.7	93,367.8	0.6	0.6
Vistula Land:										
Suvalskaia:	65.0	28.0	13.0	5.3	44.4	67.5	16,691.5	13,047.9	1.2	2.4
Kalvarskiy (2)	1.0	6.0	0.9	5.1	6.0	15.8	2,880.0	10,540.0	0.2	3.9
Mariampolskiy (3)	10.0	8.0	5.2	4.2	14.1	22.5	19,438.5	16,081.4	2.7	1.1
Vladislavovskiy (6)	10.0	3.0	6.4	1.9	11.2	7.7	28,790.3	3,200.0	0.2	2.1
Volkovyshskiy (7)	44.0	11.0	39.4	9.9	13.1	21.5	15,657.1	22,370.0	1.6	2.3
*Lithuania**	871.0	815.0	12.1	10.7	132.1	183.8	35,568.0	57,826.6	1.9	1.8
Panel II: Poland in 2016										
Vistula Land:										
Varshavskaia	684.0	711.0	60.3	62.7	74.8	74.0	125,286.9	166,400.8	0.7	0.7
Kalishskaia	408.0	326.0	36.0	28.8	36.9	48.3	50,452.9	86,854.6	1.5	1.5
Keletskaia	269.0	190.0	26.7	18.8	27.4	55.6	42,901.4	83,106.1	1.3	1.3
Liublinskaia	301.0	264.0	17.9	15.7	44.7	45.5	59,045.2	73,197.5	1.1	1.7
Lomzhinskaia	228.0	204.0	21.6	19.3	10.9	19.2	21,483.0	16,012.1	1.7	1.6
Petrokovskaia	864.0	1,087.0	70.5	88.7	126.1	148.2	216,513.0	433,449.0	0.8	0.9
Plotskaia	155.0	152.0	16.4	16.1	24.1	19.5	32,560.5	27,574.7	2.7	1.4
Radomskaia	459.0	311.0	37.2	25.2	30.0	56.5	46,526.9	86,997.1	1.2	1.3
Sedletskaia		147.0		10.3		33.9		35,034.3		1.4
Suvalskaia:	62.0	25.0	12.2	4.7	17.9	46.8	11,128.8	47,445.7	2.4	3.1
Avgustovskiy (1)	13.0	6.0	7.3	3.4	5.8	20.0	17,024.8	92,516.0	2.8	2.2
Seinskiy (4)	17.0	4.0	13.1	3.1	5.1	9.2	7,130.9	28,310.0	3.3	5.6
Suvalkskiy (5)	32.0	15.0	16.0	7.5	6.9	17.5	9,230.7	21,511.1	1.0	1.5
Vilna Governorate-General:										
Grodnenskaia:	282.0	266.0	36.1	34.1	73.3	82.1	25,047.3	46,560.4	0.4	0.7
Belostokskiy (8)	221.0	212.0	86.6	83.1	38.0	37.0	38,608.8	88,884.0	0.2	0.6
Belskiy (9)	43.0	38.0	13.7	12.1	16.5	29.8	19,785.3	28,001.4	0.7	1.4
Sokolskiy (15)	18.0	16.0	7.9	7.0	18.9	15.2	16,747.7	22,795.8	0.2	0.3
*Poland**	3,712.0	3,683.0	33.5	29.5	466.0	629.5	63,094.6	100,239.3	1.4	1.4

Sources: Data come from the Imperial Russian Factory Database developed by Gregg (2020). District Id corresponds with Fig. A.2.
* Information about Lithuania and Poland correspond with averages across corresponding provinces.

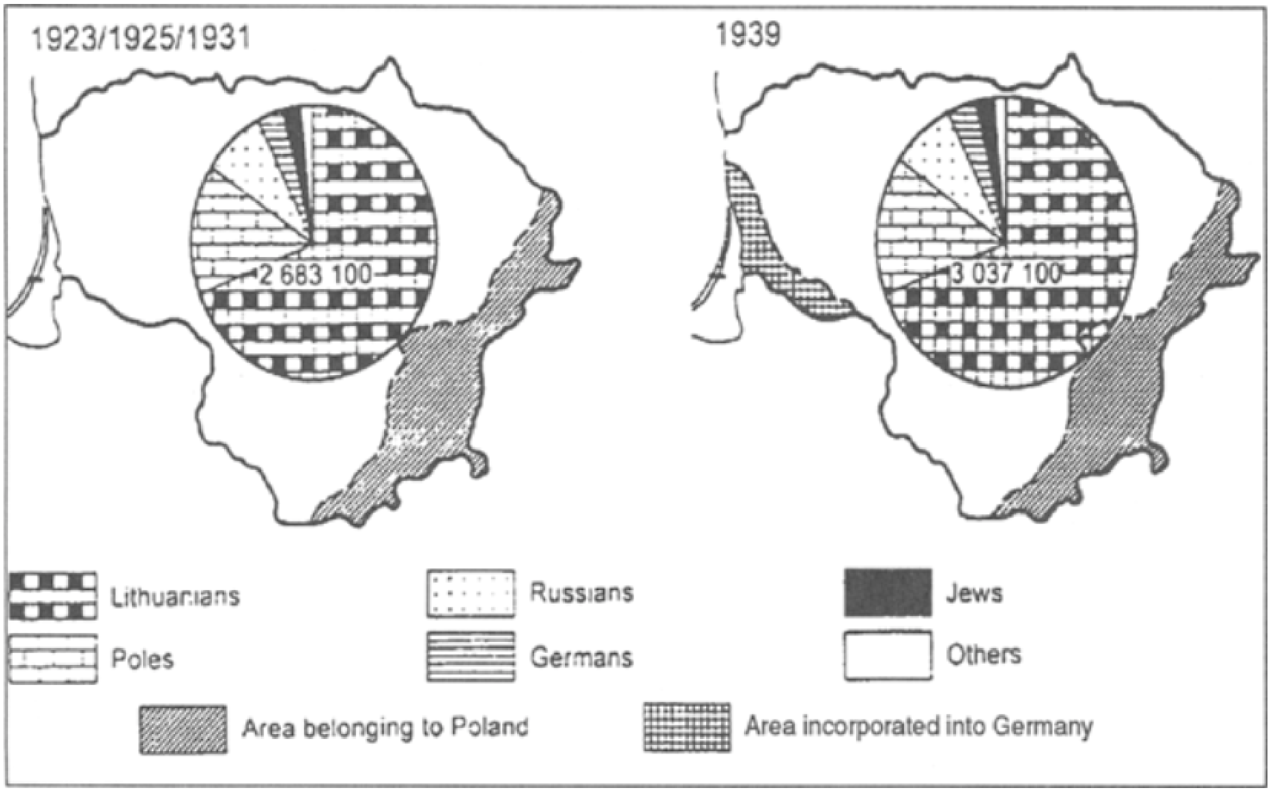
Table A.5: Industrial Composition in 1908

Province / District (Id)	Animal	Chemicals	Cotton	Flax /hemp/jute	Foods	Metals/ Machines	Mineral Products	Mixed Materials	Paper	Silk	Wood	Wool	Missing information	100 %
Panel I: Lithuania in 2016														
Vilna Governorate-General:														
Kovenskaia	7.5	4.0		0.9	31.6	18.7	7.7	3.1	13.9		10.5	1.0	1.2	100.0
Vilenskaia:	13.7		2.0		10.8	13.0	14.0	2.0	35.6		20.7	2.9	0.0	100.0
Sventsyanskiy (19)													100.0	100.0
Trokskiy (17)						14.3	14.3		42.9		28.6		0.0	100.0
Vilenskiy (18)	13.7		2.0		10.8	11.8	13.7	2.0	28.4		12.7	2.9	2.0	100.0
Vistula Land:														
Suvalskaia:	26.5				27.5	18.2	19.9				39.8		0.0	100.0
Kalvarskiy (2)	16.7				33.3		16.7				33.3		0.0	100.0
Mariampolskiy (3)					25.0		25.0				50.0		0.0	100.0
Vladislavovskiy (6)					33.3						66.7		0.0	100.0
Volkovyshskiy (7)	36.4				18.2	18.2	18.2				9.1		0.0	100.0
*Lithuania**	15.9	4.0	2.0	0.9	23.3	16.6	13.9	2.5	24.8		23.6	2.0	0.0	100.0
Panel II: Poland in 2016														
Vistula Land:														
Varshavskaia	9.3	5.1	1.4	0.4	11.3	28.6	11.0	10.0	13.6	0.4	8.6	0.4	0.0	100.0
Kalishskaia	4.3	1.5	3.7		30.7	9.2	7.4	9.5	4.0		10.4	17.5	1.8	100.0
Keletskaia	0.5	1.6		0.5	37.4	18.4	14.2		7.9		18.9		0.5	100.0
Liublinskaia	3.8	1.5			31.8	34.8	9.1		2.3		12.1	1.5	3.0	100.0
Lomzhinskaia	9.8	0.5	2.0		23.5	7.4	18.1	4.4	16.7		13.2	3.9	0.5	100.0
Petrokovskaia	3.0	2.6	17.2	1.3	4.7	7.9	10.9	5.2	4.3	1.7	6.3	34.8	0.0	100.0
Plotskaia	5.3				38.8	10.5	10.5		4.6		27.0		3.3	100.0
Radomskaia	30.2	5.1			10.9	13.8	12.9	1.6	6.8		13.8	3.2	1.6	100.0
Sedletskaia	18.4	4.8			17.7	10.2	12.9		10.2		22.4		3.4	100.0
Suvalskaia:	35.0				40.0	31.7	6.7				20.0		0.0	100.0
Avgustovskiy (1)	16.7				50.0						33.3		0.0	100.0
Seinskiy (4)					50.0	50.0							0.0	100.0
Suvalkskiy (5)	53.3				20.0	13.3	6.7				6.7		0.0	100.0
Vilna Governorate-General:														
Grodnenskaia:	25.5		0.5	0.5	6.2	5.2	9.3	1.4	0.5	1.4	22.0	39.0	0.0	100.0
Belostokskiy (8)	5.2		0.5	0.5	1.9	5.2	2.8	1.4	0.5	1.4	1.9	78.8	0.0	100.0
Belskiy (9)	2.6				10.5		15.8				57.9	13.2	0.0	100.0
Sokolskiy (15)	68.8										6.2	25.0	0.0	100.0
*Poland**	13.2	2.8	4.9	0.7	23.0	16.2	11.2	5.4	7.1	1.2	15.9	14.3	0.0	100.0

Sources: Data come from the Imperial Russian Factory Database developed by Gregg (2020). District Id corresponds with Fig. A.2.
* Information about Lithuania and Poland correspond with averages across corresponding provinces.

1.10.6 Interwar Statistics: Lithuania

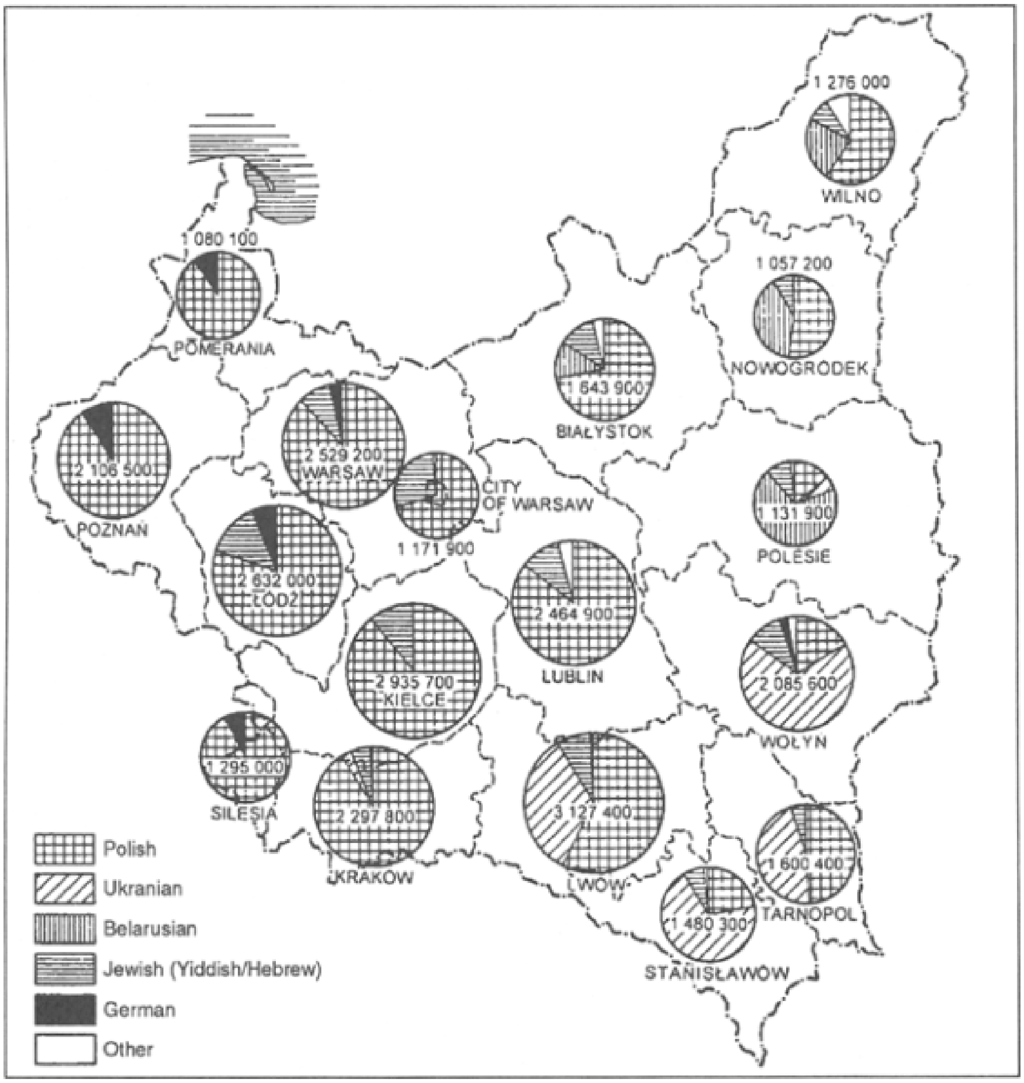
Fig. A.3. Ethnic Structure of the Present-Day Republic of Lithuania in the Interwar Period



Source: Fig. 2.2 in Eberhardt (2003) based on censuses from 1923, 1925, and 1931.

1.10.7 Interwar Statistics: Poland

Fig. A.4. Ethnic Structure of Polish Provinces in 1931 by Declared Language



Source: Fig. 3.6 in Eberhardt (2003) based on census 1931.

1.10.8 Geographical Characteristics

Table A.6: Geographical Characteristics of Lithuania and Poland

Province	Elevation	Temperature		Cloudiness		Precipitation		Actual evaporation		Potential evaporation	
		January	July	January	July	January	July	January	July	January	July
***Panel I: Lithuania in 2016*:**											
***Vilna Governorate-General*:**											
Kovenskaia	83.68	-5.13	17.19	19.14	37.54	44.71	85.79	2.50	95.07	2.50	95.07
Vilenskaia	154.88	-5.82	17.64	17.35	39.35	41.76	89.18	3.00	98.47	3.00	98.47
***Vistula Land*:**											
Suvalskaia	85.40	-4.88	17.52	17.80	38.40	43.40	93.20	3.00	97.40	3.00	97.40
Lithuania	107.99	-5.28	17.45	18.10	38.43	43.29	89.39	2.83	96.98	2.83	96.98
***Panel II: Poland in 2016*:**											
***Vistula Land*:**											
Varshavskaia	98.33	-3.40	19.03	18.83	42.83	22.67	83.17	5.00	75.33	5.00	107.33
Kalishskaia	106.75	-2.50	18.52	18.75	42.25	22.25	95.00	5.25	82.75	5.25	105.25
Keletskaia	263.00	-3.00	18.60	19.00	43.00	31.00	102.00	6.00	104.00	6.00	107.00
Liublinskaia	173.33	-4.07	18.67	19.33	44.33	31.67	93.33	5.33	95.00	5.33	108.00
Lomzhinskaia											
Petrokovskaia	143.25	-2.85	18.55	18.25	41.50	21.75	97.25	5.25	84.00	5.25	105.00
Plotskaia	143.00	-3.30	18.00	19.00	42.00	29.00	97.00	4.00	89.00	4.00	103.00
Radomskaia	150.00	-3.30	19.10	19.00	44.00	24.00	97.00	6.00	88.00	6.00	109.00
Sedletskaia	125.00	-4.55	18.70	16.00	44.50	34.00	77.50	4.50	81.00	4.50	108.00
Suvalskaia	85.40	-4.88	17.52	17.80	38.40	43.40	93.20	3.00	97.40	3.00	97.40
***Vilna Governorate-General*:**											
Grodnenskaia	136.67	-4.97	18.38	16.67	42.75	39.25	79.58	4.25	98.92	4.25	105.08
Poland	148.81	-3.55	18.62	18.31	43.02	28.40	91.31	5.06	88.67	5.06	106.41

Sources: Data come from Grosfeld and Zhuravskaya (2015). In Appendix Table A5 and A6 they describe each variable and show the descriptive statistics. The original data source that Grosfeld and Zhuravskaya (2015) use is Global GIS dataset.

1.10.9 Soviet Ideology: Propaganda

Fig. A.5. USSR's Propaganda Targeted Women

(a) To Study

(b) To Work

Note: Left one: - Woman, learn to read and write! - Oh, Mother! If you were literate, you could help me! A poster by Elizaveta Kruglikova advocating female literacy. 1923. Right one: Viktor Ivanov, Glory to the Soviet Working Women!, 1964

1.10.10 Soviet Regimes in Lithuania and Poland from 1940 to 1990

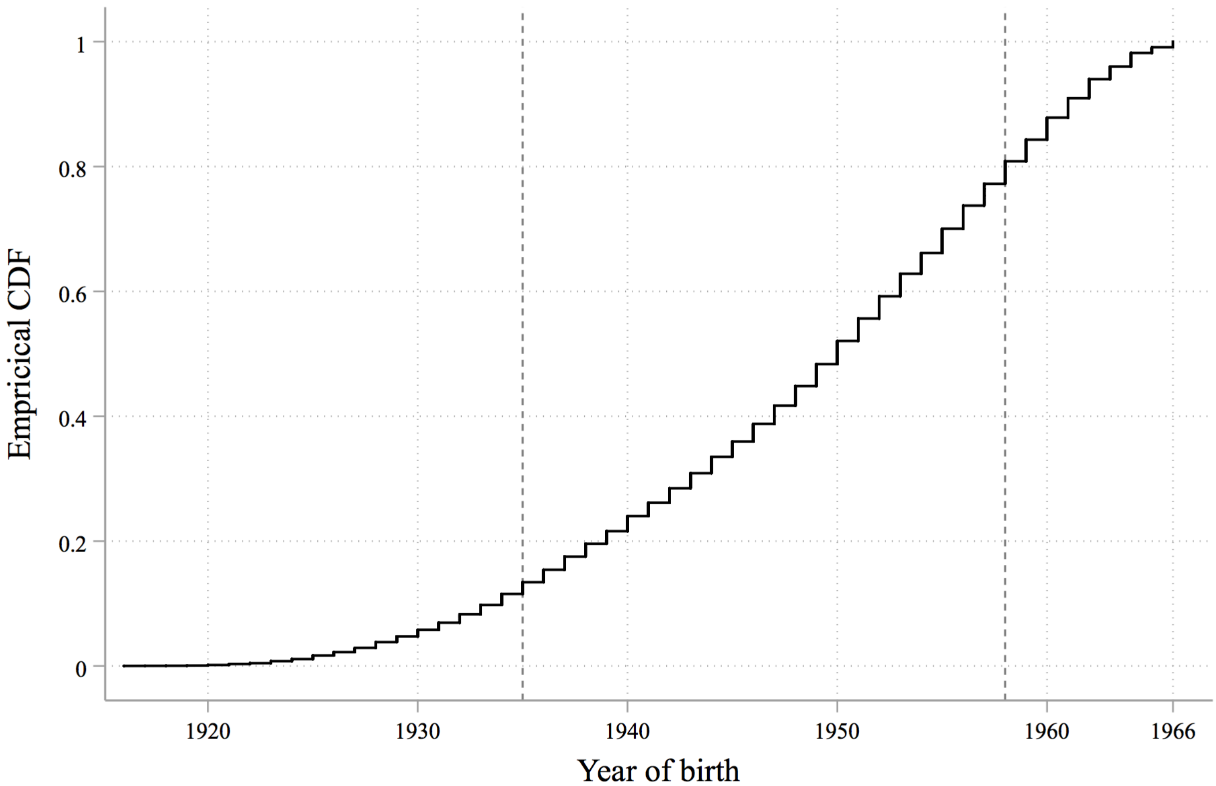
Table A.7: Distinction Between Soviet's Influence in Lithuania and Poland from 1940 to 1990

Aspect	Lithuania Former USSR	Poland Former Eastern Bloc	Sources
Schooling reforms	From 1949: mandatory 7-year education From 1959: mandatory 8-year education	From 1919: mandatory 7-year education From 1949: mandatory 7-year education From 1961: mandatory 8-year education	Lithuania: Bieliauskienė (2014, p.61, 62) Poland: Backhaus (2019), Bukowski (2019)
Property rights	From 1940, 1945: The Declaration on the Nationalization of Land was in place. Farmers became landholders and could not be landowners. 1950: 90% of all farms became state-owned. By 1952, the mass collectivization was completed.	From 1949: Nation-wide collectivization began. 1956: rebellions took place. Farmers were allowed to own lands. 1980: 3/4 of all farmland was private property; 1/4 of workforce were private farmers.	Lithuania: Girnius (1988), Poland: Brzezinski (1967, p.36)
Roman Catholic Church	The state was atheistic, and the Soviet authorities suppressed dramatically the church.	1956: Wladyslaw Gomulka (the First Secretary of the Polish United Workers' Party) eased pressure on the Roman Catholic Church and turned to a more 'Polish' form of communism. Karol Wojtyla of Krakow became Pope of the Roman Catholic Church, taking the name John Paul II. 1979: John Paul II received warm welcome during his visit to Poland. Poles got hope to oppose their government. 1980: the Solidarity labor movement was founded.	Poland: Brzezinski (1967, p.36)

1.11 Appendix: SHARE Data

1.11.1 Year of Birth

Fig. A.6. The Empirical Cumulative Distribution of the Year of Birth of the SHARE Respondents



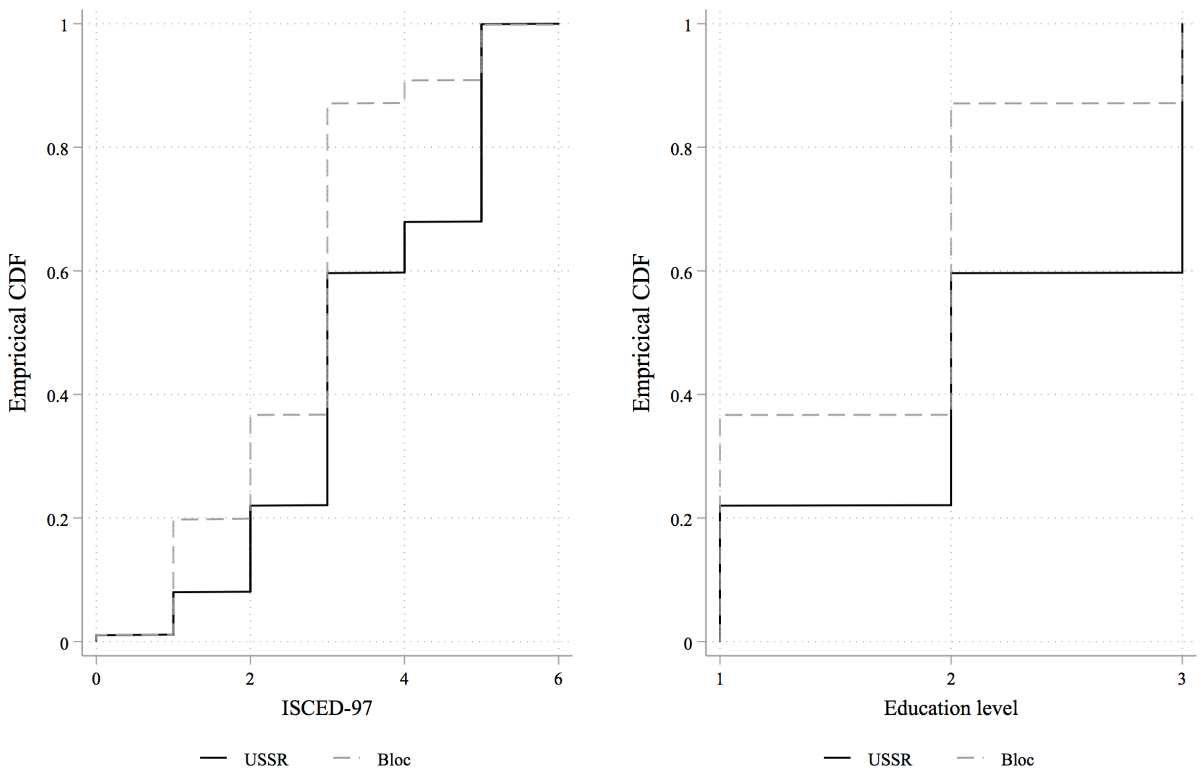
1.11.2 Area Under the Analysis

Fig. A.7. Regions in the Analysis in Lithuania and Poland

Note: *Lithuania*: 4801 - Alytus County, 4802 - Kaunas County, 4803 - Klaipeda County, 4804 - Marijampole County, 4805 - Panevezys County, 4806 - Siauliai County, 4807 - Taurage County, 4808 - Telsiai County, 4809 - Utena County, 4810 - Vilnius County; *Poland*: 2903 - Lublin Voivodeship; 2905 - Lódz Voivodeship; 2907 - Masovian Voivodeship; 2910 - Podlaskie Voivodeship; 2913 - Swietokrzyskie Voivodeship.
Source: GIS map of country borders in 2016 comes from Eurostat, GISCO. The regiond numbering corresponds with the SHARE.

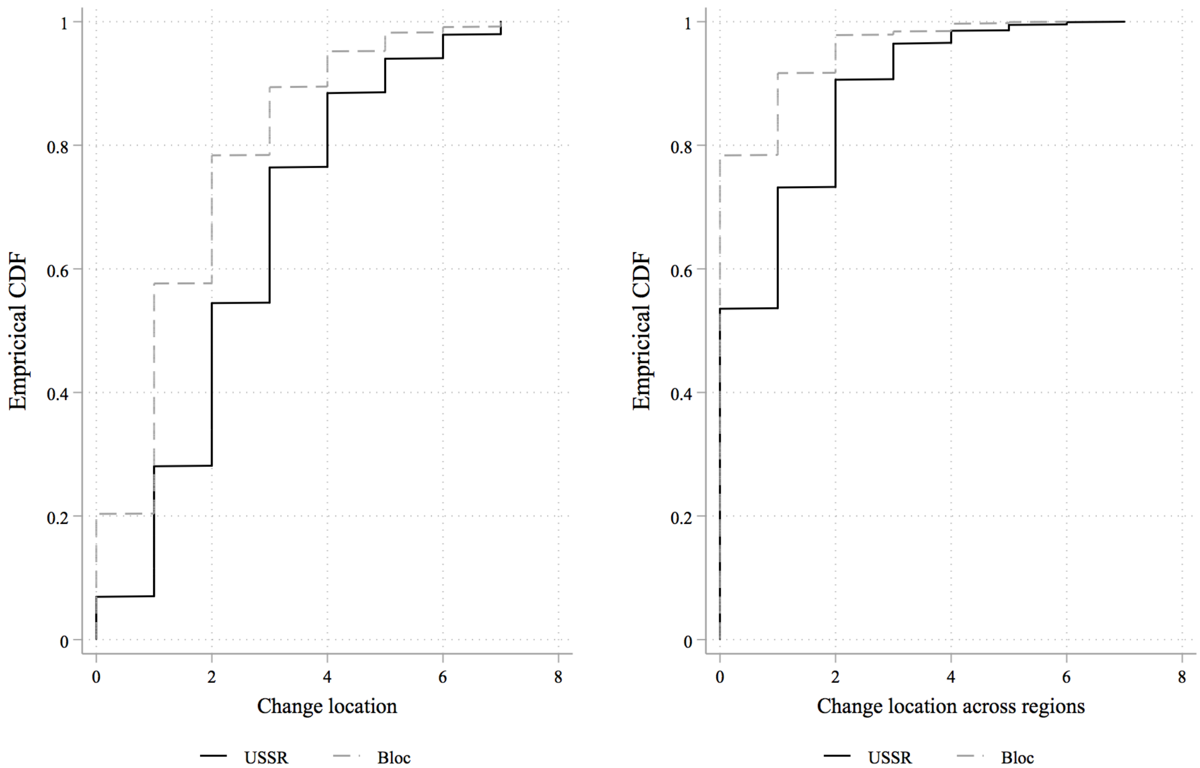
1.11.3 Educational Attainment

Fig. A.8. Educational Attainment



1.11.4 Geographical Mobility

Fig. A.9. Place of Residence During the Life



1.11.5 Descriptive Statistics

Table A.8: Descriptive Statistics

	Statistics				
	Mean	se	Min	Max	N
Female	0.57	0.50	0	1	2190
Treatment variable:					
USSR at age 18	0.10	0.30	0	1	2190
Outcome variables:					
(I) Education:					
Low	0.35	0.48	0	1	2190
Secondary	0.49	0.50	0	1	2190
High	0.16	0.36	0	1	2190
Years in education	10.40	3.16	2	27	2156
(II) Cumulative years of work experience before 1990:					
By age 25	5.55	3.23	0	18	2190
By age 50	20.29	8.21	0	42	2190
Between age 25 and 50	14.75	6.59	0	25	2190
(III) Marriage history and the number of children before 1990:					
Ever-married	0.95	0.22	0	1	2161
Number of marriages	1.04	0.19	1	4	2086
Age at birth of first child	24.37	3.85	12	47	1988
Number of children	2.52	1.19	1	9	1988
(IV) Life satisfaction:					
Life satisfaction	6.90	2.14	0	10	2171
Life quality	35.24	6.79	14	48	2134
Control variables:					
Health at age 10:					
Poor	0.02	0.13	0	1	2190
Fair	0.06	0.24	0	1	2190
Good	0.28	0.45	0	1	2190
Very good	0.64	0.48	0	1	2190
Mental problem during childhood	0.01	0.09	0	1	2190
A bad student at math at age 10	0.71	0.45	0	1	2190
The place of birth:					
A big city	0.07	0.25	0	1	2190
The suburbs of a big city	0.01	0.09	0	1	2190
A large town	0.11	0.31	0	1	2190
A small town	0.09	0.28	0	1	2190
A rural area	0.73	0.44	0	1	2190
Number of books at age 10:					
¡11 books	0.53	0.50	0	1	2190
11-25 books	0.25	0.43	0	1	2190
¿26 books	0.22	0.42	0	1	2190
Number of services of the individual's dwelling at age 10:					
No	0.82	0.39	0	1	2190
1 service	0.05	0.22	0	1	2190
¿1 services	0.14	0.34	0	1	2190
Number of bedrooms at age 10:					
¡2 bedrooms	0.39	0.49	0	1	2190
2 bedrooms	0.37	0.48	0	1	2190
3 bedrooms	0.18	0.38	0	1	2190
¿3 bedrooms	0.07	0.25	0	1	2190
Property was dispossessed	0.04	0.19	0	1	2190

Note: We further control for the year of birth.

1.12 Appendix: Average Marginal Impact and Heterogeneity

1.12.1 Inconsistent Estimation of the AMI in Presence of Heterogeneity

Here, we show that an estimator $\hat{\alpha}^f$ from [1.4.4] may not be consistent estimate of the AMI for women when the effect is heterogeneous in education. We consider that the data follows the model in [1.4.5]. First, define the deviation of the USSR's impact from its (conditional) mean:

$$v_i = \alpha^f(E_i) - \mathbb{E}[\alpha^f(E_i)|G_i = 1] = \alpha_1^f\left(E_{1i} - \mathbb{E}[E_{1i}|G_i = 1]\right) + \alpha_2^f\left(E_{2i} - \mathbb{E}[E_{2i}|G_i = 1]\right) \quad [1.12.1]$$

With this definition, we can rewrite [1.4.5] as

$$Y_i = \gamma_0^f + \mathbb{E}[\alpha^f(E_i)|G_i = 1]Z_i + \beta^{f\prime} + \underbrace{Z_i v_i + \varepsilon_i}_{=\nu_i} \quad [1.12.2]$$

Therefore, an estimator based on [1.4.4] is consistent for $\text{AMI}^f = \mathbb{E}[\alpha^f(E_i)|G_i = 1]$ if and only if

$$\mathbb{E}[Z_i v_i] = \alpha_1^f\,\mathbb{E}[Z_i\left(E_{1i} - \mathbb{E}[E_{1i}|G_i = 1]\right)] + \alpha_2^f\,\mathbb{E}[Z_i\left(E_{2i} - \mathbb{E}[E_{2i}|G_i = 1]\right)] = 0 \quad [1.12.3]$$

We believe that this is not the case, since being born in the USSR highly correlates with educational attainment, and we expect the impact of the Soviet Union to be heterogeneous in education (i.e. $\alpha_1^f \neq 0$ and $\alpha_2^f \neq 0$), as it is shown in Fig. 1.2.

1.12.2 Estimator of the AMI in the Presence of Heterogeneity

We assume that the data satisfies equation [1.4.5]. Then,

$$\frac{\partial Y_i}{\partial Z_i} = \alpha^f(E_i) = \alpha_0^f + \alpha_1^f E_{1i} + \alpha_2^f E_{2i} \quad [1.12.4]$$

First, we compute the AMI of USSR on women, conditional on a fixed level of education

$e \in \{0, 1, 2\}$. This is given by

$$\text{AMI}^f(e) = \mathbb{E}\left[\frac{\partial Y_i}{\partial Z_i}\Big| G_i = 1, E_i = e\right] = \mathbb{E}\left[\alpha_0^f + \alpha_1^f E_{1i} + \alpha_2^f E_{2i}\Big| G_i = 1, E_i = e\right] \qquad [1.12.5]$$

A case by case evaluation of the above equation results in the expression in [1.4.8].

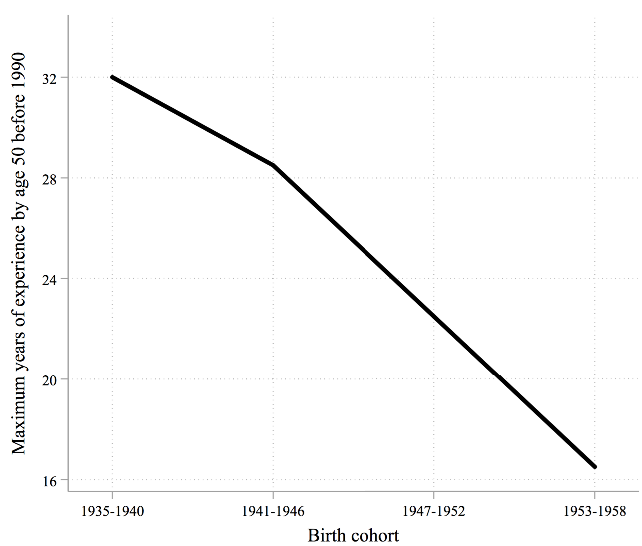
To find the (unconditional) AMI of the USSR on women, recall that $\text{AMI}^f = \mathbb{E}[\alpha^f(E_i)|G_i = 1]$. Thus, since $\text{P}(E_{ji} = 1) = \text{P}(E_i = j)$, we have that

$$\begin{aligned} \text{AMI}^f &= \mathbb{E}\left[\alpha_0^f + \alpha_1^f E_{1i} + \alpha_2^f E_{2i}|G_i = 1\right] \\ &= \alpha_0^f + \alpha_1^f\,\text{P}(E_{1i} = 1|G_i = 1) + \alpha_2^f\,\text{P}(E_{2i} = 1|G_i = 1) \\ &= \alpha_0^f + \alpha_1^f\,\text{P}(E_i = 1|G_i = 1) + \alpha_2^f\,\text{P}(E_i = 2|G_i = 1) \end{aligned} \qquad [1.12.6]$$

Thus, in presence of heterogeneity, we construct the AMI of the USSR following equation [1.12.6] (an its counterpart for men). The parameters $(\alpha_0^f, \alpha_1^f, \alpha_2^f)$ come from estimating regression [1.4.7]. Probabilities are replaced by in-sample proportions. We compute standard errors using the δ-method.

1.12.3 Intensive Margin of the USSR's Impact

Fig. A.10. Maximum Attainable Work Experience by Age 50 Before 1990



1.13 Appendix: Results

1.13.1 Years of Education

Table A.9: USSR, Education and Work Experience

		Cumulative work experience								
	Education	No control for education			Controls for three education levels			Heterogeneity with education		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Variables		By 25	By 50	25-50	By 25	By 50	25-50	By 25	By 50	25-50
Panel I: Both men and women										
Female	-0.150	-0.522**	-2.139***	-1.617***	-0.564***	-2.154***	-1.590***			
	(0.149)	(0.221)	(0.472)	(0.327)	(0.214)	(0.466)	(0.323)			
Female × USSR	0.520*	1.188***	2.164***	0.976**	1.333***	2.216***	0.883**			
	(0.287)	(0.286)	(0.570)	(0.398)	(0.269)	(0.554)	(0.390)			
USSR	0.259	-0.721***	-0.236	0.484*	-0.648***	-0.210	0.438*			
	(0.226)	(0.208)	(0.362)	(0.247)	(0.196)	(0.363)	(0.253)			
Education:										
Years in education					-0.280***	-0.101	0.179***			
					(0.0365)	(0.0881)	(0.0643)			
AMI of the USSR on women	0.779	0.467	1.927	1.460	0.685	2.006	1.320			
P-value: AMI=0	0.000	0.035	0.000	0.000	0.001	0.000	0.000			
R^2	0.407	0.0957	0.391	0.536	0.140	0.392	0.540			
N	2158	2158	2158	2158	2158	2158	2158			
Panel II: Women										
USSR	0.709***	0.439*	2.037***	1.599***	0.626***	2.064***	1.438***	1.645*	5.737***	4.092***
	(0.204)	(0.232)	(0.487)	(0.360)	(0.215)	(0.476)	(0.362)	(0.891)	(1.956)	(1.399)
Education:										
Years in education					-0.264***	-0.0376	0.226***	-0.246***	0.0261	0.272***
					(0.0405)	(0.104)	(0.0839)	(0.0475)	(0.127)	(0.102)
Education × USSR:										
USSR × Years in education								-0.0908	-0.327**	-0.236**
								(0.0704)	(0.154)	(0.113)
AMI of the USSR								0.704	2.345	1.642
P-value: AMI=0								0.004	0.000	0.000
R^2	0.435	0.0986	0.328	0.449	0.135	0.328	0.455	0.136	0.330	0.457
N	1259	1259	1259	1259	1259	1259	1259	1259	1259	1259
Panel III: Men										
USSR	0.320	-0.655***	-0.0920	0.563***	-0.556***	-0.0208	0.536***	-0.468	2.335	2.803**
	(0.230)	(0.231)	(0.331)	(0.198)	(0.214)	(0.320)	(0.196)	(0.899)	(1.700)	(1.103)
Education:										
Years in education					-0.309***	-0.223*	0.0861	-0.307***	-0.189	0.119
					(0.0590)	(0.128)	(0.0844)	(0.0673)	(0.146)	(0.0964)
Education × USSR:										
USSR × Years in education								-0.00799	-0.213	-0.205**
								(0.0790)	(0.152)	(0.0984)
AMI of the USSR								-0.551	0.112	0.663
P-value: AMI=0								0.013	0.733	0.001
R^2	0.405	0.158	0.555	0.719	0.213	0.560	0.721	0.213	0.561	0.722
N	899	899	899	899	899	899	899	899	899	899

Note: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. The cluster standard errors at the year of birth and the region of residence at age 18 are in parentheses. We restrict to individuals who were born from 1935 to 1958 in the former territories of Russian Empire in Lithuania and Poland. In *Panel (I)* we report the estimated coefficient γ_2 from Equation 1.4.1, and AMI from Equation 1.4.2. In *Panel (II)* and *Panel (III)*, we report the estimated coefficient α^f, α^m from Equation 1.4.4, and AMI from Equation 1.4.8. All regressions control for constant, a four categories of health at age 10, a mental health problem dummy at age 10, to be a good student at math dummy; five place of birth dummies: a big city, the suburbs of a big city, a large town, a small town or rural area; the features of the individual's dwelling at age 10: the number of books by age 10, the number of services, the number of rooms, and the year fixed effects.

1.13.2 Unweighted Sample

Table A.10: USSR, Education and Work Experience

		Cumulative work experience								
	Education	No control for education			Controls for three education levels			Heterogeneity with education		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Variables		By 25	By 50	25-50	By 25	By 50	25-50	By 25	By 50	25-50
Panel I: Both men and women										
Female	-0.257***	-0.507***	-2.070***	-1.563***	-0.639***	-2.103***	-1.464***			
	(0.0615)	(0.191)	(0.409)	(0.293)	(0.184)	(0.404)	(0.287)			
Female × USSR	0.519***	1.101***	1.991***	0.890***	1.468***	2.233***	0.765**			
	(0.0988)	(0.258)	(0.493)	(0.343)	(0.245)	(0.490)	(0.339)			
USSR	0.324***	-0.493***	0.0887	0.581***	-0.180	0.325	0.505**			
	(0.0798)	(0.187)	(0.316)	(0.213)	(0.177)	(0.318)	(0.220)			
Education:										
Secondary					-0.719***	0.418	1.137***			
					(0.180)	(0.374)	(0.271)			
High					-2.553***	-1.365***	1.188***			
					(0.202)	(0.390)	(0.275)			
AMI of the USSR on women	0.843	0.608	2.080	1.472	1.288	2.558	1.270			
P-value: AMI=0	0.000	0.001	0.000	0.000	0.000	0.000	0.000			
R^2	0.328	0.0884	0.489	0.639	0.158	0.496	0.643			
N	2190	2190	2190	2190	2190	2190	2190			
Panel II: Women										
USSR	0.861***	0.573***	2.034***	1.461***	1.261***	2.413***	1.152***	2.127***	4.154***	2.027***
	(0.0651)	(0.192)	(0.391)	(0.274)	(0.195)	(0.408)	(0.289)	(0.442)	(1.035)	(0.735)
Education:										
Secondary					-0.655**	1.001*	1.655***	-0.131	1.923**	2.054***
					(0.258)	(0.569)	(0.412)	(0.315)	(0.799)	(0.598)
High					-2.534***	-0.835	1.698***	-2.115***	0.541	2.656***
					(0.274)	(0.544)	(0.396)	(0.376)	(0.788)	(0.584)
Education × USSR:										
Secondary × USSR								-1.364***	-2.416**	-1.052
								(0.490)	(1.122)	(0.801)
High × USSR								-0.976**	-2.626**	-1.650**
								(0.491)	(1.066)	(0.781)
AMI of the USSR								1.283	2.390	1.107
P-value: AMI=0								0.000	0.000	0.000
R^2	0.380	0.0758	0.438	0.575	0.144	0.445	0.584	0.151	0.449	0.585
N	1277	1277	1277	1277	1277	1277	1277	1277	1277	1277
Panel III: Men										
USSR	0.324***	-0.479**	0.263	0.742***	-0.160	0.566*	0.727***	0.697	2.151***	1.454***
	(0.0814)	(0.206)	(0.315)	(0.192)	(0.194)	(0.311)	(0.196)	(0.437)	(0.796)	(0.535)
Education:										
Secondary					-0.802***	-0.483	0.319	-0.314	0.297	0.610
					(0.273)	(0.514)	(0.342)	(0.349)	(0.714)	(0.490)
High					-2.616***	-2.305***	0.311	-2.420***	-1.472*	0.949*
					(0.318)	(0.531)	(0.328)	(0.484)	(0.865)	(0.554)
Education × USSR:										
Secondary × USSR								-1.368***	-2.158**	-0.790
								(0.512)	(0.909)	(0.598)
High × USSR								-0.705	-2.033*	-1.328*
								(0.621)	(1.061)	(0.677)
AMI of the USSR								-0.141	0.605	0.746
P-value: AMI=0								0.472	0.054	0.000
R^2	0.275	0.149	0.599	0.762	0.221	0.609	0.763	0.229	0.613	0.764
N	913	913	913	913	913	913	913	913	913	913

Note: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. The cluster standard errors at the year of birth and the region of residence at age 18 are in parentheses. We restrict to individuals who were born from 1935 to 1958 in the former territories of Russian Empire in Lithuania and Poland. In *Panel (I)* we report the estimated coefficient γ_2 from Equation 1.4.1, and AMI from Equation 1.4.2. In *Panel (II)* and *Panel (III)*, we report the estimated coefficient α^f, α^m from Equation 1.4.4, and AMI from Equation 1.4.8. All regressions control for constant, a four categories of health at age 10, a mental health problem dummy at age 10, to be a good student at math dummy; five place of birth dummies: a big city, the suburbs of a big city, a large town, a small town or rural area; the features of the individual's dwelling at age 10: the number of books by age 10, the number of services, the number of rooms, and the year fixed effects.

Table A.11: USSR and Marriage History, Children and Later-Life Well-Being

Variables	Marriage history and the number of children					
	(1) Ever-married	(2) Number of marriages	(3) Age delivery	(4) Number of children	(5) Life satisfaction	(6) Life quality
Panel I: Both men and women						
Female	0.0215*	0.00966	-2.901***	0.0270	-0.180*	-1.200***
	(0.0112)	(0.00956)	(0.241)	(0.0687)	(0.106)	(0.370)
Female × USSR	-0.0319**	0.0103	1.254***	-0.102	0.105	1.162**
	(0.0161)	(0.0226)	(0.339)	(0.0948)	(0.183)	(0.542)
USSR	0.0256**	0.0659***	-0.483*	-0.189**	-0.807***	-3.643***
	(0.0129)	(0.0196)	(0.252)	(0.0751)	(0.140)	(0.425)
AMI of the USSR on women	-0.006	0.076	0.771	-0.291	-0.702	-2.482
P-value: AMI=0	0.554	0.000	0.002	0.000	0.000	0.000
R^2	0.0204	0.0589	0.143	0.118	0.0839	0.161
N	2163	2087	1990	1990	2172	2135
Panel II: Women						
USSR	-0.00737	0.0797***	0.860***	-0.323***	-0.694***	-2.397***
	(0.0109)	(0.0164)	(0.265)	(0.0693)	(0.141)	(0.447)
R^2	0.0396	0.0656	0.0809	0.118	0.0838	0.156
N	1268	1227	1180	1180	1265	1245
Panel III: Men						
USSR	0.0263**	0.0564***	-0.576**	-0.154*	-0.842***	-3.737***
	(0.0126)	(0.0199)	(0.258)	(0.0784)	(0.150)	(0.448)
R^2	0.0446	0.105	0.111	0.154	0.110	0.182
N	895	860	810	810	907	890

Note: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. The cluster standard errors at the year of birth and the region of residence at age 18 are in parentheses. We restrict to individuals who were born from 1935 to 1958 in the former territories of Russian Empire in Lithuania and Poland. We consider only children born before 1990. In *Panel (I)* we report the estimated coefficient γ_2 from Equation 1.4.1, and AMI from Equation 1.4.2. In *Panel (II)* and *Panel (III)*, we report the estimated coefficient α^f, α^m from Equation 1.4.4. All regressions control for constant, a four categories of health at age 10, a mental health problem dummy at age 10, to be a good student at math dummy; five place of birth dummies: a big city, the suburbs of a big city, a large town, a small town or rural area; the features of the individual's dwelling at age 10: the number of books by age 10, the number of services, the number of rooms, and the year fixed effects.

1.13.3 Controlling for Region Fixed Effects

Table A.12: USSR, Education and Work Experience Controlling for Region Fixed Effects

		Cumulative work experience					
	Education	No control for education			Controls for three education levels		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Variables		By 25	By 50	25-50	By 25	By 50	25-50
Female	-0.292***	-0.501**	-2.066***	-1.565***	-0.612***	-2.034***	-1.421***
	(0.0648)	(0.213)	(0.454)	(0.316)	(0.205)	(0.435)	(0.300)
Female × USSR	0.507***	1.248***	2.167***	0.919**	1.557***	2.339***	0.782**
	(0.106)	(0.288)	(0.571)	(0.394)	(0.274)	(0.555)	(0.386)
Secondary					-0.280	1.096*	1.376***
					(0.243)	(0.612)	(0.463)
High					-2.263***	-1.010	1.253***
					(0.321)	(0.650)	(0.448)
R^2	0.364	0.0976	0.393	0.541	0.138	0.400	0.548
N	2155	2155	2155	2155	2155	2155	2155

Note: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. The cluster standard errors at the year of birth and the region of residence at age 18 are in parentheses. We restrict to individuals who were born from 1935 to 1958 in the former territories of Russian Empire in Lithuania and Poland. We report the estimated coefficient γ_2 from Equation 1.4.1. All regressions control for constant, a four categories of health at age 10, a mental health problem dummy at age 10, to be a good student at math dummy; five place of birth dummies: a big city, the suburbs of a big city, a large town, a small town or rural area; the features of the individual's dwelling at age 10: the number of books by age 10, the number of services, the number of rooms, and the year and region fixed effects.

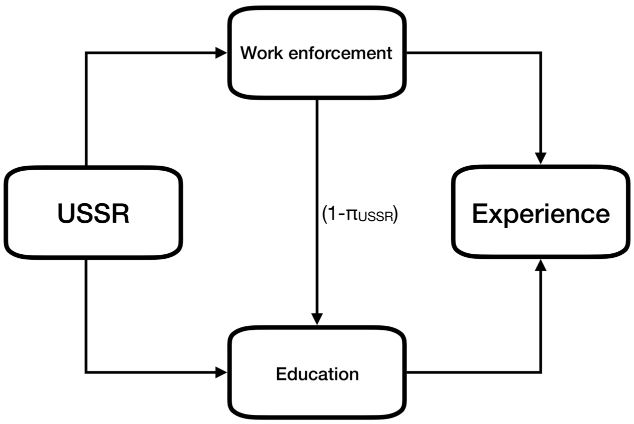
Table A.13: USSR and Marriage History, Children and Later-Life Well-Being Controlling for Region Fixed Effects

	Marriage history and the number of children					
	(1)	(2)	(3)	(4)	(5)	(6)
Variables	Ever-married	Number of marriages	Age delivery	Number of children	Life satisfaction	Life quality
Panel I: Both men and women						
Female	0.0235	0.00853	-3.126***	0.0461	-0.222	-1.304***
	(0.0176)	(0.0114)	(0.253)	(0.0743)	(0.139)	(0.444)
Female × USSR	-0.0307	0.0244	1.357***	-0.0758	0.129	1.357**
	(0.0218)	(0.0241)	(0.350)	(0.108)	(0.210)	(0.612)
R^2	0.0676	0.0884	0.210	0.149	0.0970	0.162
N	2129	2053	1958	1958	2137	2101

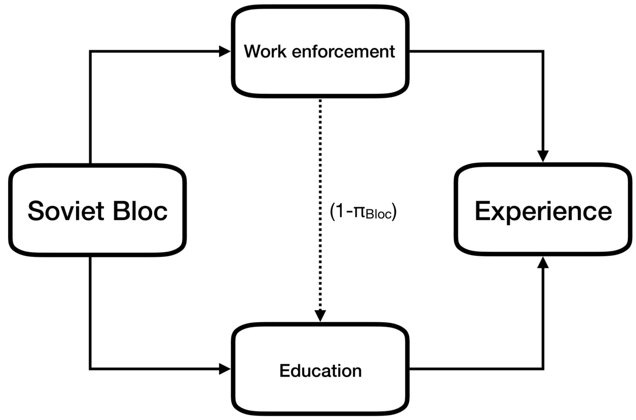
Note: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. The cluster standard errors at the year of birth and the region of residence at age 18 are in parentheses. We restrict to individuals who were born from 1935 to 1958 in the former territories of Russian Empire in Lithuania and Poland. We consider only children born before 1990. We report the estimated coefficient γ_2 from Equation 1.4.1. All regressions control for constant, a four categories of health at age 10, a mental health problem dummy at age 10, to be a good student at math dummy; five place of birth dummies: a big city, the suburbs of a big city, a large town, a small town or rural area; the features of the individual's dwelling at age 10: the number of books by age 10, the number of services, the number of rooms, and the year and region fixed effects.

1.13.4 Mechanism of Labor and Schooling Decisions in the USSR

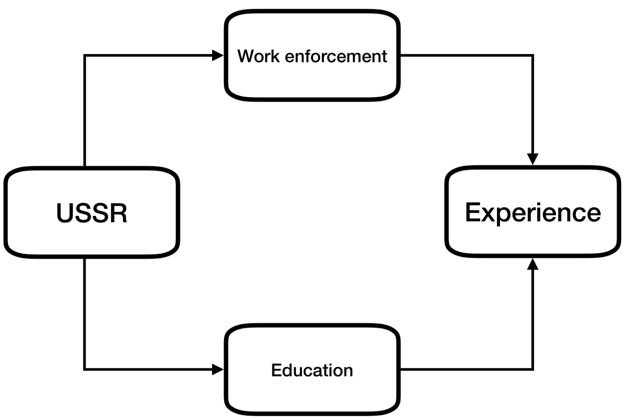
Fig. A.11. USSR, Education and Work Experience



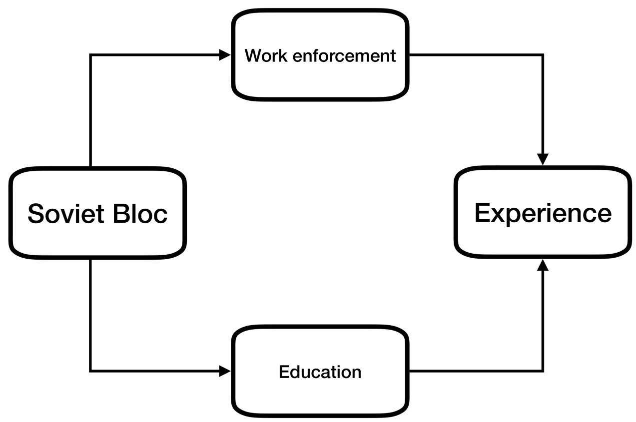
(a) Women in the USSR



(b) Women in the Eastern Bloc



(c) Men in the USSR

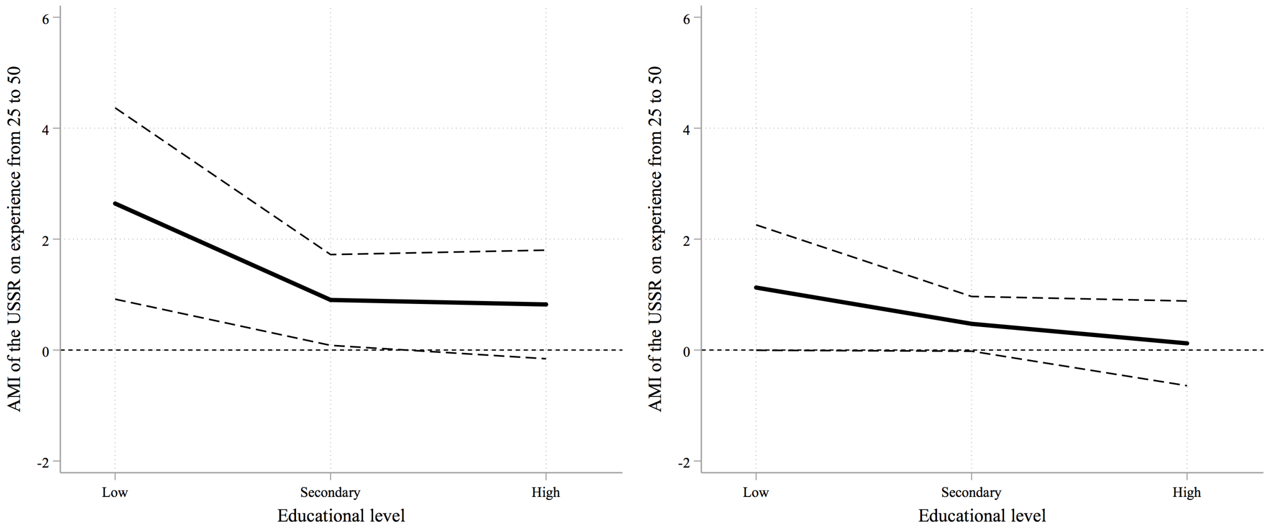


(d) Men in the Eastern Bloc

1.14 Appendix: Heterogeneity with Education

1.14.1 Three Education Groups

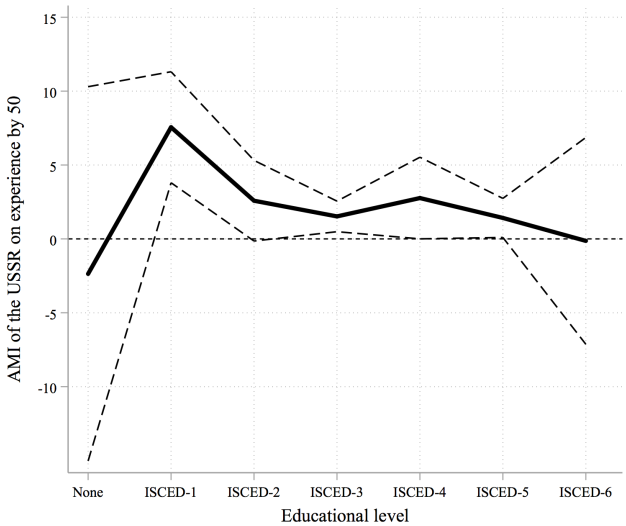
Fig. A.12. AMI of the USSR on Experience from 25 to 50 Across Three Education Groups by Gender



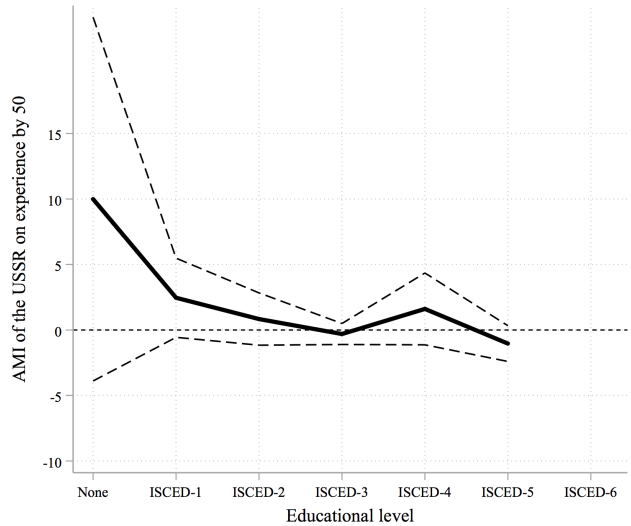
(a) Women **(b)** Men

1.14.2 Seven Education Groups

Fig. A.13. AMI of the USSR on Experience by Age 50 Across Seven Education Groups by Gender



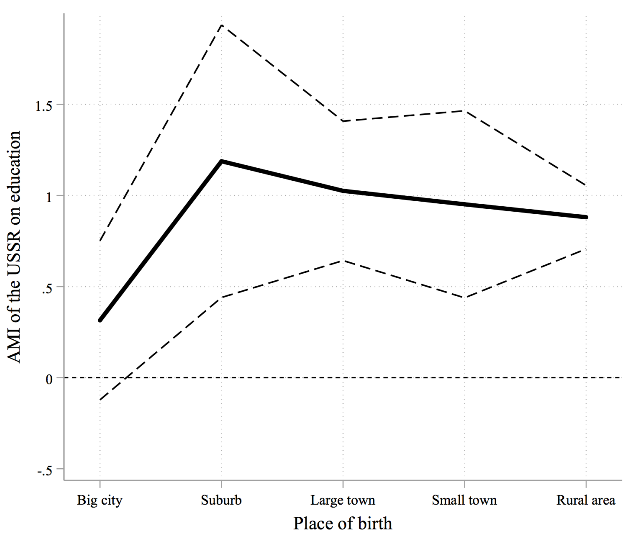
(a) Women



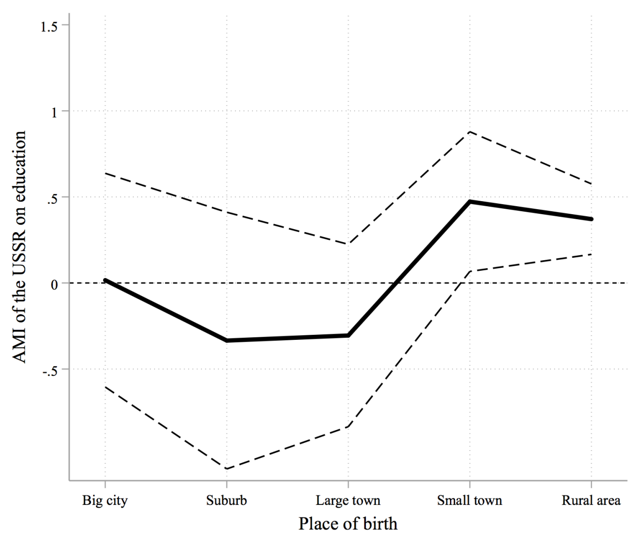
(b) Men

1.14.3 On Education Across the Place of Birth

Fig. A.14. AMI of the USSR on Education Across the Place of Birth by Gender



(a) Women

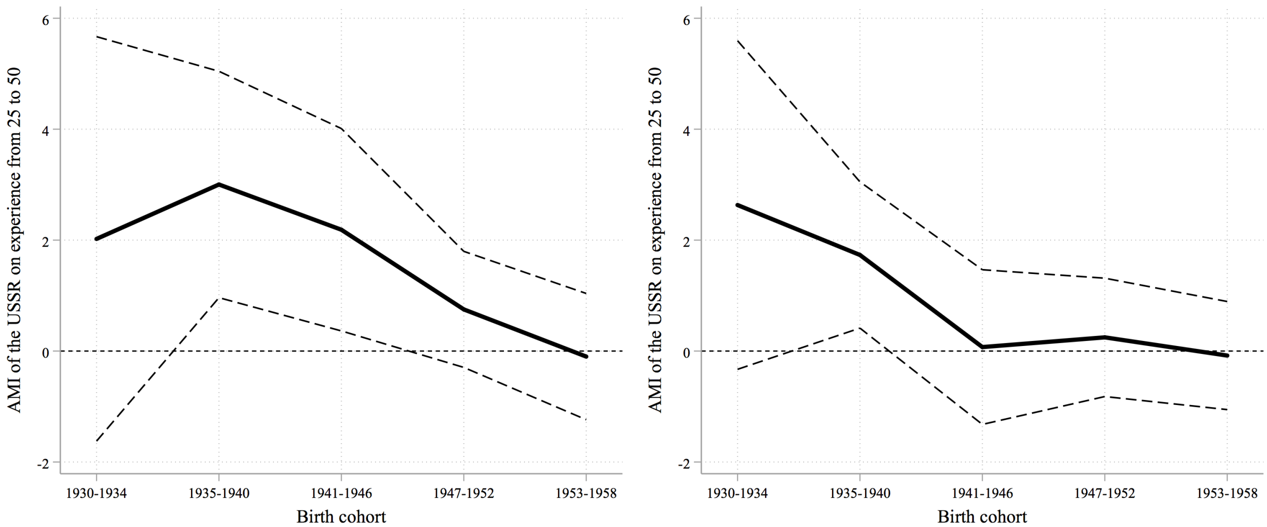


(b) Men

1.15 Appendix: Intensive Margin of the Impact of the USSR

1.15.1 On Experience from 25 to 50 Across Birth Cohorts

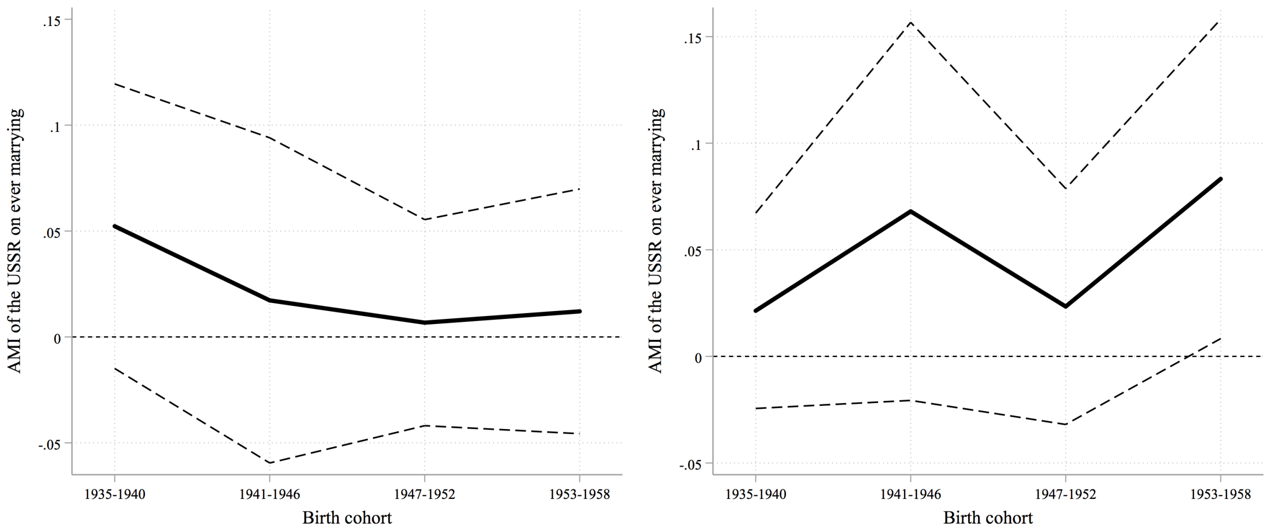
Fig. A.15. AMI of the USSR on Experience from 25 to 50 Across Birth Cohorts by Gender



(a) Women **(b)** Men

1.15.2 On Marrying Across the Year of Birth

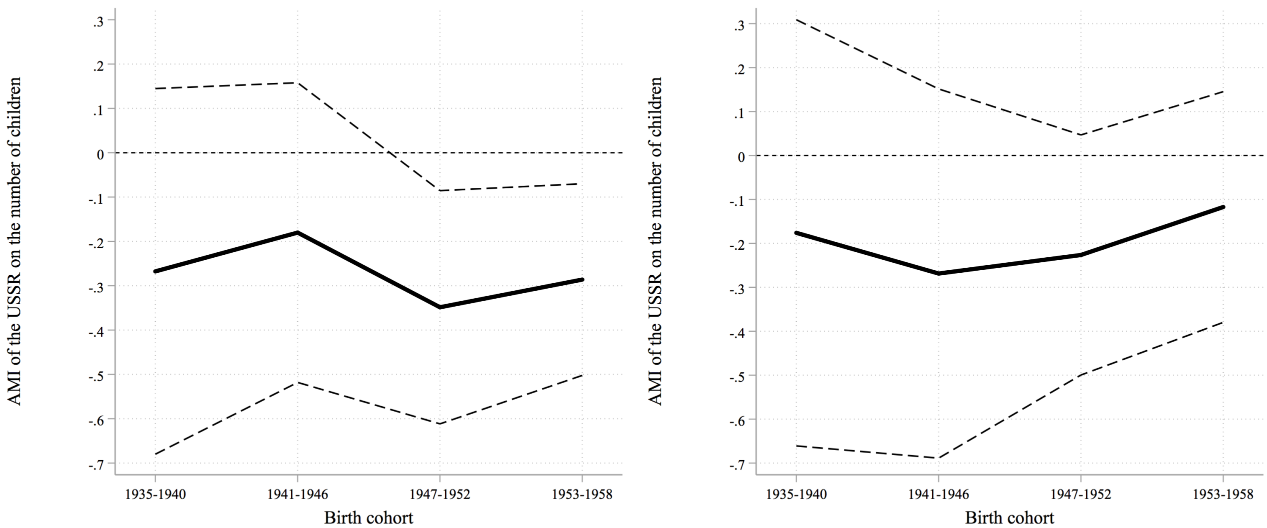
Fig. A.16. AMI of the USSR on Marrying Across Birth Cohorts by Gender



(a) Women **(b)** Men

1.15.3 On the Number of Children Across the Year of Birth

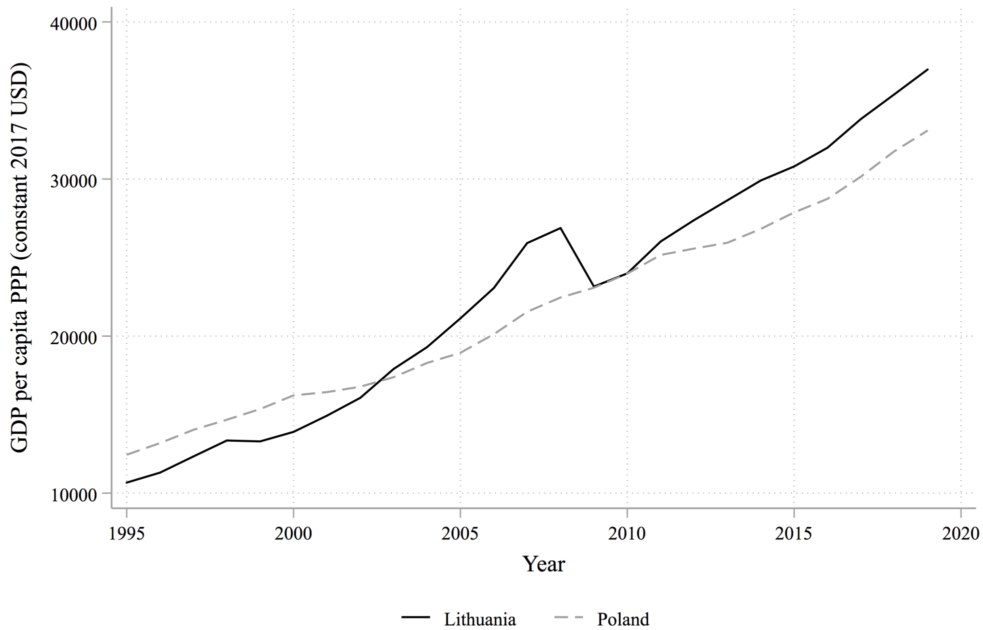
Fig. A.17. AMI of the USSR on the Number of Children Across Birth Cohorts by Gender



(a) Women **(b)** Men

1.16 Economic Conditions During the Transition Period

Fig. A.18. GDP Per Capita in Lithuania and Poland During the Transition Period



Source: World Development Indicators database, World Bank: a variable is GDP per capita, PPP (constant 2017 international $).

1.17 Appendix: Threats for Identification

1.17.1 Impact of WWII

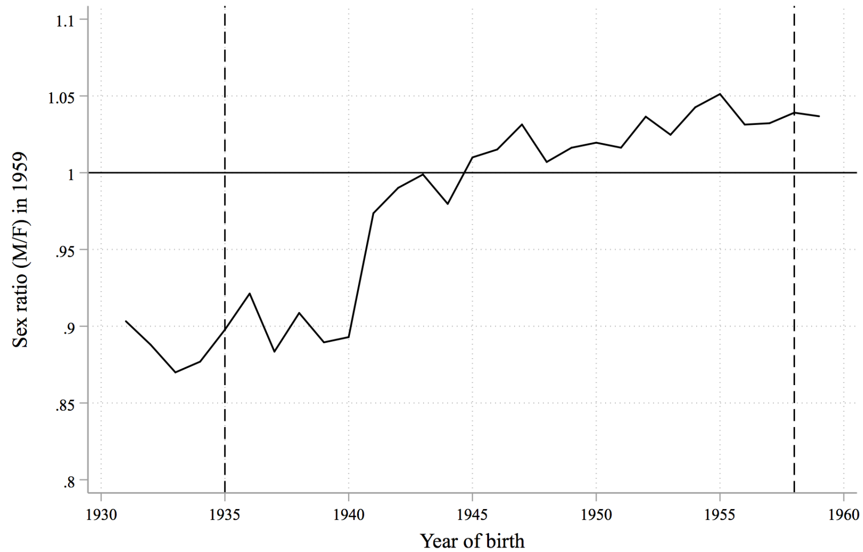
Table A.14: Impact of WWII

	Family killed or injured in WWII				
Region name	(1) No (%)	(2) Yes (%)	(3) Missing Info (%)	100 %	(4) Observations
Lithuania:					
Alytus county	86.49	10.81	2.70	100.00	37
Kaunas county	68.38	23.93	7.69	100.00	117
Klaipeda county	67.61	16.90	15.49	100.00	71
Marijampole county	68.00	28.00	4.00	100.00	25
Panevezys county	78.69	9.84	11.48	100.00	61
Siauliai county	75.00	18.75	6.25	100.00	48
Taurage county	89.47	10.53	0.00	100.00	19
Telsiai county	70.97	12.90	16.13	100.00	31
Utena county	78.26	19.57	2.17	100.00	46
Vilnius county	77.97	16.10	5.93	100.00	118
All regions	74.69	17.45	7.85	100.00	573
Poland:					
Greater Poland Voivodeship	14.29	61.90	23.81	100.00	21
Kuyavian-Pomeranian Voivodeship	51.61	48.39	0.00	100.00	31
Lesser Voivodeship	71.43	23.81	4.76	100.00	42
Lodz Voivodeship*	34.15	58.54	7.32	100.00	41
Lower Silesian Voivodeship	21.95	68.29	9.76	100.00	41
Lublin Voivodeship	32.08	62.26	5.66	100.00	53
Lubusz Voivodeship*	10.00	50.00	40.00	100.00	10
Masovian Voivodeship*	46.88	50.00	3.12	100.00	32
Opole Voivodeship	73.68	21.05	5.26	100.00	19
Podkarpackie Voivodeship	75.00	20.83	4.17	100.00	72
Podlaskie Voivodeship*	29.63	66.67	3.70	100.00	27
Pomeranian Voivodeship	67.50	30.00	2.50	100.00	40
Silesian Voivodeship	59.65	21.05	19.30	100.00	57
Swietokrzyskie Voivodeship*	16.67	83.33	0.00	100.00	6
Warmian-Masurian Voivodeship	84.62	15.38	0.00	100.00	13
West Pomeranian Voivodeship	33.33	59.26	7.41	100.00	27
All regions	49.44	42.86	7.71	100.00	532
Former Russian Partition regions	33.62	58.62	7.76	100.00	116

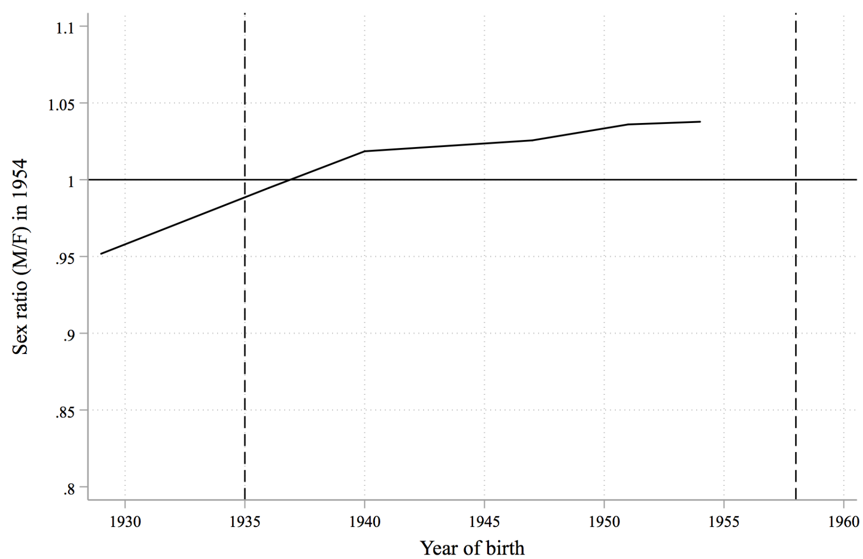
Note: Data come from the 2016 Life in Transition Survey (LiTS). Question 9.24a: *Were you, your parents or any of your grandparents physically injured or were your parents or any of your grandparents killed during the Second World War?* We restrict to respondents who were born from 1935 to 1958. The region name corresponds with the region of current residence.
* Region was in the former Russian Partition of Poland.

1.17.2 Demographic Characteristics: Sex Ratio

Fig. A.19. Sex Ratio Male/Female



(a) Lithuania

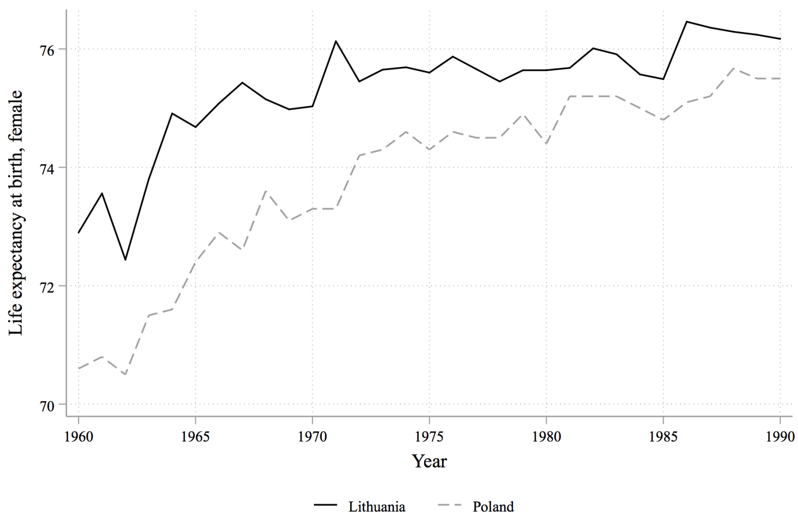


(b) Poland

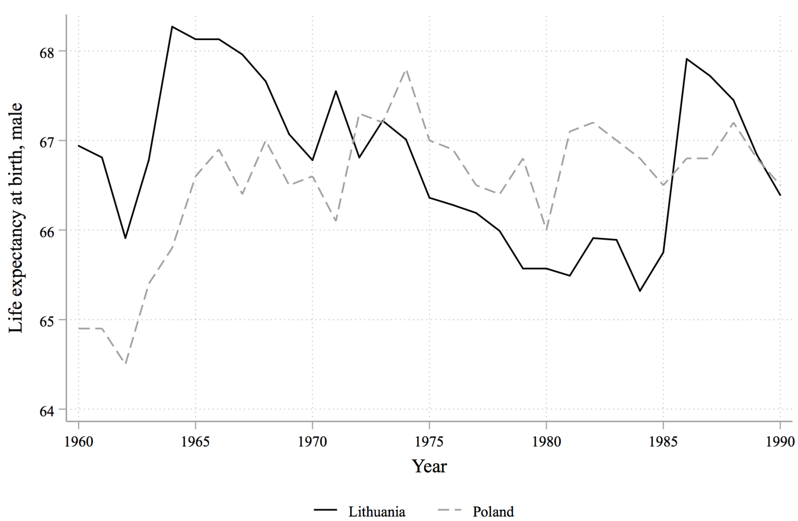
Source: 1959 Soviet Census.
http://www.demoscope.ru/weekly/ssp/sng_mar_59.php?reg=8&gor=3&Submit=OK
Polish Statistical Yearbook 1955. Table 6 p.38.
http://istmat.info/files/uploads/51389/rocznik_statystyczny_1955.pdf

1.17.3 Demographic Characteristics: Life Expectancy at Birth

Fig. A.20. Life Expectancy at Birth



(a) Women



(b) Men

Source: World Development Indicators database, World Bank: a variable is life expectancy at birth among female and male (years) in Lithuania and Poland

1.17.4 Out-Migration from Lithuania and Poland

Table A.15: Profile of Migrants from Lithuania and Poland

	Lithuania (1)	Poland (2)	(1) - (2)	P-value
Female	0.57	0.50	0.06	0.63
Health at age 10:				
Poor	0.00	0.07	-0.06	0.33
Fair	0.02	0.08	-0.07	0.33
Good	0.28	0.35	-0.18	0.18
Very good	0.70	0.50	0.31	0.02
Mental problem during childhood	0.01	0.00	0.00	0.78
A bad student at math at age 10	0.55	0.64	-0.18	0.17
The place of birth:				
A big city	0.54	0.18	0.17	0.10
The suburbs of a big city	0.00	0.01	-0.01	0.71
A large town	0.00	0.15	-0.11	0.20
A small town	0.09	0.18	-0.12	0.30
A rural area	0.36	0.48	0.07	0.61
Number of books at age 10:				
¡11 books	0.31	0.60	-0.21	0.14
11-25 books	0.11	0.17	0.04	0.74
¿26 books	0.58	0.23	0.17	0.17
Number of services of the individual's dwelling at age 10:				
No	0.45	0.29	0.24	0.04
1 service	0.09	0.30	-0.25	0.05
¿1 services	0.47	0.41	0.00	0.98
Number of bedrooms at age 10:				
¡2 bedrooms	0.12	0.20	-0.01	0.94
2 bedrooms	0.56	0.33	0.05	0.69
3 bedrooms	0.16	0.26	-0.10	0.45
¿3 bedrooms	0.17	0.20	0.05	0.65
Property was dispossessed	0.28	0.36	0.04	0.75
Observations	35.00	215.00		

Sources: We restrict to respondents who were born from 1935 and 1958 in Lithuania (Column 1) or Poland (Column 2) and currently reside in the other European country.

1.18 Appendix: Robustness Checks

1.18.1 Movements During Life

Table A.16: USSR, Education and Work Experience and Movements During Life

	Education	Cumulative work experience					
		No control for education			Controls for three education levels		
	(1)	(2) By 25	(3) By 50	(4) 25-50	(5) By 25	(6) By 50	(7) 25-50
Panel I: The region of birth							
Female	-0.332***	-0.585***	-2.287***	-1.702***	-0.715***	-2.238***	-1.523***
	(0.0656)	(0.213)	(0.463)	(0.322)	(0.206)	(0.444)	(0.307)
Female × USSR	0.554***	1.290***	2.344***	1.054***	1.598***	2.463***	0.866**
	(0.107)	(0.290)	(0.571)	(0.389)	(0.275)	(0.553)	(0.380)
USSR	0.256***	-0.734***	-0.319	0.415*	-0.468**	-0.0529	0.415*
	(0.0823)	(0.204)	(0.359)	(0.246)	(0.204)	(0.367)	(0.247)
Education:							
Secondary					-0.311	1.189**	1.500***
					(0.227)	(0.575)	(0.446)
High					-2.051***	-0.527	1.524***
					(0.329)	(0.639)	(0.431)
AMI of the USSR on women	0.809	0.556	2.025	1.468	1.130	2.410	1.281
P-value: AMI=0	0.000	0.009	0.000	0.000	0.000	0.000	0.000
R^2	0.328	0.0790	0.382	0.534	0.113	0.388	0.542
N	2250	2250	2250	2250	2250	2250	2250
Panel II: The region at age 18							
Female	-0.289***	-0.466**	-2.058***	-1.592***	-0.596***	-2.062***	-1.466***
	(0.0646)	(0.209)	(0.445)	(0.311)	(0.203)	(0.428)	(0.297)
Female × USSR	0.519***	1.143***	2.090***	0.948**	1.480***	2.299***	0.819**
	(0.106)	(0.280)	(0.553)	(0.386)	(0.265)	(0.539)	(0.380)
USSR	0.300***	-0.734***	-0.267	0.467*	-0.440**	-0.00417	0.436*
	(0.0852)	(0.204)	(0.355)	(0.242)	(0.204)	(0.367)	(0.251)
Education:							
Secondary					-0.312	0.998*	1.310***
					(0.231)	(0.575)	(0.439)
High					-2.191***	-0.924	1.267***
					(0.311)	(0.608)	(0.421)
AMI of USSR on women	0.819	0.408	1.823	1.415	1.039	2.294	1.255
P-value: AMI=0	0.000	0.060	0.000	0.000	0.000	0.000	0.000
R^2	0.338	0.0927	0.392	0.541	0.131	0.397	0.547
N	2190	2192	2192	2192	2190	2190	2190
Panel III: The region in which lived the most							
Female	-0.232***	-0.410*	-1.905***	-1.495***	-0.520**	-1.925***	-1.405***
	(0.0675)	(0.223)	(0.500)	(0.356)	(0.214)	(0.481)	(0.344)
Female × USSR	0.475***	1.086***	2.026***	0.940**	1.397***	2.223***	0.825**
	(0.103)	(0.292)	(0.593)	(0.416)	(0.274)	(0.571)	(0.407)
USSR	0.367***	-0.773***	-0.0789	0.694***	-0.432**	0.208	0.640**
	(0.0796)	(0.208)	(0.383)	(0.256)	(0.209)	(0.395)	(0.263)
Education:							
Secondary					-0.333	0.783	1.117***
					(0.230)	(0.565)	(0.428)
High					-2.174***	-0.974	1.199***
					(0.340)	(0.653)	(0.444)
AMI of USSR on women	0.842	0.313	1.947	1.634	0.965	2.430	1.465
P-value: AMI=0	0.000	0.158	0.000	0.000	0.000	0.000	0.000
R^2	0.335	0.0979	0.379	0.520	0.134	0.382	0.524
N	2183	2185	2185	2185	2183	2183	2183

Note: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. The cluster standard errors at the year of birth and the region of residence at age 18 are in parentheses. We restrict to individuals who were born from 1935 to 1958 in the former territories of Russian Empire in Lithuania and Poland. In *Panel (I)* we report the estimated coefficient γ_2 from Equation 1.4.1, and AMI from Equation 1.4.2. In *Panel (II)* and *Panel (III)*. All regressions control for constant, a four categories of health at age 10, a mental health problem dummy at age 10, to be a good student at math dummy; five place of birth dummies: a big city, the suburbs of a big city, a large town, a small town or rural area; the features of the individual's dwelling at age 10: the number of books by age 10, the number of services, the number of rooms, and the year fixed effects.

1.18.2 Interwar Borders

Table A.17: USSR, Education and Work Experience

		Cumulative work experience								
	Education	No control for education			Controls for three education levels			Heterogeneity with education		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Variables		By 25	By 50	25-50	By 25	By 50	25-50	By 25	By 50	25-50
Panel I: Both men and women										
Female	-0.289***	-0.469**	-2.065***	-1.596***	-0.595***	-2.059***	-1.464***			
	(0.0647)	(0.210)	(0.446)	(0.311)	(0.204)	(0.429)	(0.297)			
Female × USSR	0.564***	1.158***	2.249***	1.090***	1.521***	2.467***	0.946**			
	(0.118)	(0.306)	(0.587)	(0.413)	(0.288)	(0.576)	(0.407)			
USSR	0.267***	-0.790***	-0.372	0.418	-0.536**	-0.167	0.369			
	(0.0943)	(0.234)	(0.397)	(0.273)	(0.228)	(0.406)	(0.282)			
Education:										
Secondary					-0.273	1.074*	1.347***			
					(0.239)	(0.595)	(0.454)			
High					-2.157***	-0.809	1.348***			
					(0.332)	(0.644)	(0.443)			
AMI of USSR on women	0.831	0.369	1.876	1.508	0.985	2.300	1.315			
P-value: AMI=0	0.000	0.118	0.000	0.000	0.000	0.000	0.000			
R^2	0.338	0.0940	0.385	0.533	0.130	0.390	0.539			
N	1866	1868	1868	1868	1866	1866	1866			
Panel II: Women										
USSR	0.860***	0.365	2.035***	1.670***	0.994***	2.338***	1.344***	1.671***	3.980***	2.309**
	(0.0846)	(0.248)	(0.512)	(0.376)	(0.250)	(0.551)	(0.418)	(0.632)	(1.449)	(1.003)
Education:										
Secondary					-0.195	1.781**	1.976***	-0.146	1.889**	2.035***
					(0.339)	(0.895)	(0.703)	(0.348)	(0.932)	(0.734)
High					-2.149***	-0.0879	2.061***	-2.116***	0.0570	2.173***
					(0.387)	(0.834)	(0.636)	(0.435)	(0.914)	(0.688)
Education × USSR:										
Secondary × USSR								-1.060	-2.339	-1.280
								(0.664)	(1.522)	(1.080)
High × USSR								-0.686	-1.897	-1.211
								(0.699)	(1.490)	(1.021)
AMI of the USSR								1.080	2.622	1.541
P-value: AMI=0								0.000	0.000	0.002
R^2	0.408	0.0964	0.318	0.445	0.128	0.327	0.457	0.129	0.328	0.458
N	1075	1075	1075	1075	1075	1075	1075	1075	1075	1075
Panel III: Men										
USSR	0.250**	-0.748***	-0.294	0.454**	-0.511**	-0.0819	0.429*	0.268	1.528	1.260*
	(0.0999)	(0.264)	(0.377)	(0.227)	(0.256)	(0.377)	(0.229)	(0.576)	(1.039)	(0.665)
Education:										
Secondary					-0.451	-0.225	0.226	-0.402	-0.136	0.266
					(0.362)	(0.752)	(0.507)	(0.377)	(0.786)	(0.531)
High					-2.196***	-1.844*	0.352	-2.148***	-1.687	0.461
					(0.530)	(0.950)	(0.593)	(0.572)	(1.025)	(0.640)
Education × USSR:										
Secondary × USSR								-1.051	-1.922	-0.871
								(0.675)	(1.212)	(0.765)
High × USSR								-0.902	-2.237*	-1.336
								(0.795)	(1.351)	(0.825)
AMI of the USSR								-0.436	0.149	0.585
P-value: AMI=0								0.090	0.705	0.022
R^2	0.285	0.157	0.549	0.715	0.197	0.554	0.716	0.198	0.555	0.716
N	791	791	791	791	791	791	791	791	791	791

Note: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. The cluster standard errors at the year of birth and the region of residence at age 18 are in parentheses. We restrict to individuals who were born from 1935 to 1958 in the former territories of Russian Empire in Lithuania and Poland. In *Panel (I)* we report the estimated coefficient γ_2 from Equation 1.4.1, and *AMI* from Equation 1.4.2. In *Panel (II)* and *Panel (III)*, we report the estimated coefficient α^f, α^m from Equation 1.4.4, and *AMI* from Equation 1.4.8. All regressions control for constant, a four categories of health at age 10, a mental health problem dummy at age 10, to be a good student at math dummy; five place of birth dummies: a big city, the suburbs of a big city, a large town, a small town or rural area; the features of the individual's dwelling at age 10: the number of books by age 10, the number of services, the number of rooms, and the year fixed effects.

Table A.18: USSR and Marriage History, Children and Later-Life Well-Being

Variables	Marriage history and the number of children (1) Ever-married	(2) Number of marriages	(3) Age delivery	(4) Number of children	(5) Life satisfaction	(6) Life quality
Panel I: Both men and women						
Female	0.0235	0.00393	-3.069***	0.0528	-0.220	-1.322***
	(0.0173)	(0.0123)	(0.249)	(0.0749)	(0.138)	(0.448)
Female × USSR	-0.0285	0.0310	1.732***	-0.132	0.117	1.139*
	(0.0225)	(0.0274)	(0.380)	(0.121)	(0.238)	(0.681)
USSR	0.0599***	0.0476**	-0.952***	-0.0899	-0.602***	-3.021***
	(0.0202)	(0.0229)	(0.299)	(0.0958)	(0.198)	(0.534)
AMI of the USSR on women	0.031	0.079	0.779	-0.221	-0.485	-1.882
P-value: AMI=0	0.102	0.000	0.006	0.016	0.004	0.001
R^2	0.0534	0.0705	0.205	0.136	0.0925	0.130
N	1842	1781	1698	1698	1850	1823
Panel II: Women						
USSR	0.0281	0.0713***	0.906***	-0.246**	-0.380**	-1.630***
	(0.0218)	(0.0201)	(0.290)	(0.100)	(0.180)	(0.586)
R^2	0.0898	0.0769	0.148	0.166	0.136	0.169
N	1068	1038	995	995	1065	1049
Panel III: Men						
USSR	0.0635***	0.0464*	-0.927***	-0.0497	-0.799***	-3.376***
	(0.0208)	(0.0239)	(0.316)	(0.0991)	(0.207)	(0.585)
R^2	0.0987	0.220	0.140	0.193	0.135	0.145
N	774	743	703	703	785	774

Note: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. The cluster standard errors at the year of birth and the region of residence at age 18 are in parentheses. We restrict to individuals who were born from 1935 to 1958 in the former territories of Russian Empire in Lithuania and Poland. We consider only children born before 1990. In *Panel (I)* we report the estimated coefficient γ_2 from Equation 1.4.1, and AMI from Equation 1.4.2. In *Panel (II)* and *Panel (III)*, we report the estimated coefficient α^f, α^m from Equation 1.4.4. All regressions control for constant, a four categories of health at age 10, a mental health problem dummy at age 10, to be a good student at math dummy; five place of birth dummies: a big city, the suburbs of a big city, a large town, a small town or rural area; the features of the individual's dwelling at age 10: the number of books by age 10, the number of services, the number of rooms, and the year fixed effects.

1.18.3 The USSR and the Eastern Bloc

Table A.19: USSR, Education and Work Experience in the USSR and the Eastern Bloc

	Education	Cumulative work experience					
		No control for education			Controls for three education levels		
	(1)	(2) By 25	(3) By 50	(4) 25-50	(5) By 25	(6) By 50	(7) 25-50
Panel I: Former territories of the Russian Empire in Lithuania and Poland							
Female	-0.289***	-0.485**	-2.085***	-1.600***	-0.596***	-2.062***	-1.466***
	(0.0646)	(0.210)	(0.446)	(0.311)	(0.203)	(0.428)	(0.297)
Female × USSR	0.519***	1.163***	2.120***	0.956**	1.480***	2.299***	0.819**
	(0.106)	(0.280)	(0.554)	(0.387)	(0.265)	(0.539)	(0.380)
USSR	0.300***	-0.741***	-0.275	0.466*	-0.440**	-0.00417	0.436*
	(0.0852)	(0.205)	(0.356)	(0.242)	(0.204)	(0.367)	(0.251)
Education:							
Secondary					-0.312	0.998*	1.310***
					(0.231)	(0.575)	(0.439)
High					-2.191***	-0.924	1.267***
					(0.311)	(0.608)	(0.421)
AMI of the USSR on women	0.819	0.423	1.845	1.422	1.039	2.294	1.255
P-value: AMI=0	0.000	0.052	0.000	0.000	0.000	0.000	0.000
R^2	0.338	0.0925	0.391	0.540	0.131	0.397	0.547
N	2190	2190	2190	2190	2190	2190	2190
Panel II: Only Lithuania and Poland							
Female	-0.309***	-0.849***	-3.344***	-2.495***	-0.953***	-3.230***	-2.277***
	(0.0420)	(0.122)	(0.297)	(0.231)	(0.118)	(0.281)	(0.215)
Female × USSR	0.538***	1.501***	3.359***	1.858***	1.795***	3.390***	1.595***
	(0.0908)	(0.220)	(0.428)	(0.310)	(0.207)	(0.415)	(0.298)
USSR	0.327***	-1.057***	-0.437	0.621***	-0.781***	-0.245	0.536***
	(0.0706)	(0.167)	(0.294)	(0.202)	(0.164)	(0.297)	(0.206)
Education:							
Secondary					0.0195	1.815***	1.796***
					(0.161)	(0.381)	(0.289)
High					-1.797***	0.261	2.058***
					(0.208)	(0.446)	(0.326)
AMI of the USSR on women	0.865	0.444	2.922	2.479	1.014	3.144	2.130
P-value: AMI=0	0.000	0.008	0.000	0.000	0.000	0.000	0.000
R^2	0.277	0.0647	0.348	0.473	0.101	0.358	0.486
N	3980	3980	3980	3980	3980	3980	3980
Panel III: Former USSR and former Eastern Bloc							
Female	-0.319***	-0.537***	-2.864***	-2.327***	-0.660***	-2.740***	-2.079***
	(0.0265)	(0.0923)	(0.207)	(0.149)	(0.0890)	(0.193)	(0.138)
Female × USSR	0.517***	1.097***	3.022***	1.924***	1.401***	3.001***	1.601***
	(0.0561)	(0.151)	(0.293)	(0.202)	(0.140)	(0.279)	(0.193)
USSR	0.288***	-0.968***	-0.665***	0.303**	-0.665***	-0.373*	0.292**
	(0.0462)	(0.117)	(0.206)	(0.137)	(0.114)	(0.204)	(0.138)
Education:							
Secondary					-0.0609	1.712***	1.773***
					(0.131)	(0.256)	(0.172)
High					-1.994***	-0.180	1.814***
					(0.154)	(0.291)	(0.197)
AMI of the USSR on women	0.805	0.130	2.357	2.227	0.736	2.628	1.892
P-value: AMI=0	0.000	0.271	0.000	0.000	0.000	0.000	0.000
R^2	0.262	0.0460	0.333	0.497	0.0880	0.344	0.510
N	14615	14615	14615	14615	14615	14615	14615

Note: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. The cluster standard errors at the year of birth and the region of residence at age 18 are in parentheses. We restrict to individuals who were born from 1935 to 1958. We report the estimated coefficient γ_2 from Equation 1.4.1. All regression control for constant, a four categories of health at age 10, a mental health problem dummy at age 10, to be a good student at math dummy; five place of birth dummies: a big city, the suburbs of a big city, a large town, a small town or rural area; the features of the individual's dwelling at age 10: the number of books by age 10, the number of services (e.g., hot running water supply, inside toilet and others), the number of rooms, and the year fixed effects.

Table A.20: USSR and Marriage History, Children, Later-Life Well-Being in the USSR and the Eastern Bloc

Variables	Marriage history and the number of children (1) Ever-married	(2) Number of marriages	(3) Age delivery	(4) Number of children	(5) Life satisfaction	(6) Life quality
Panel I: Former territories of the Russian Empire in Lithuania and Poland						
Female	0.0236	0.00355	-3.066***	0.0594	-0.207	-1.270***
	(0.0154)	(0.0118)	(0.247)	(0.0752)	(0.135)	(0.438)
Female × USSR	-0.0300	0.0283	1.347***	-0.109	0.165	1.289**
	(0.0201)	(0.0244)	(0.352)	(0.109)	(0.207)	(0.612)
USSR	0.0423**	0.0457**	-0.392	-0.149*	-0.706***	-3.279***
	(0.0179)	(0.0203)	(0.284)	(0.0841)	(0.166)	(0.497)
AMI of the USSR on women	0.012	0.074	0.955	-0.258	-0.540	-1.991
P-value: AMI=0	0.443	0.000	0.000	0.001	0.001	0.000
R^2	0.0446	0.0558	0.186	0.134	0.0856	0.122
N	2214	2140	2040	2040	2220	2179
Panel II: Only Lithuania and Poland						
Female	0.0464***	0.0199**	-3.062***	0.144***	-0.114	-0.792***
	(0.0121)	(0.00900)	(0.157)	(0.0514)	(0.0837)	(0.286)
Female × USSR	-0.0531***	0.0153	1.389***	-0.186**	0.0651	0.840*
	(0.0173)	(0.0234)	(0.299)	(0.0903)	(0.170)	(0.501)
USSR	0.0487***	0.0533***	-0.283	-0.258***	-0.830***	-3.860***
	(0.0152)	(0.0192)	(0.235)	(0.0714)	(0.135)	(0.401)
AMI of the USSR on women	-0.004	0.069	1.106	-0.444	-0.765	-3.020
P-value: AMI=0	0.691	0.000	0.000	0.000	0.000	0.000
R^2	0.0387	0.0371	0.166	0.0936	0.0634	0.115
N	3920	3785	3633	3633	3939	3872
Panel III: Former USSR and former Eastern Bloc						
Female	0.0336***	0.0171***	-3.239***	0.119***	-0.287***	-1.135***
	(0.00715)	(0.00652)	(0.110)	(0.0365)	(0.0562)	(0.182)
Female × USSR	-0.0293**	0.00786	1.483***	-0.187***	0.249**	0.948***
	(0.0125)	(0.0169)	(0.217)	(0.0573)	(0.109)	(0.333)
USSR	-0.00140	0.0550***	-0.0486	-0.0950**	-0.684***	-1.390***
	(0.0111)	(0.0144)	(0.168)	(0.0461)	(0.0861)	(0.276)
AMI of the USSR on women	-0.031	0.063	1.435	-0.282	-0.435	-0.442
P-value: AMI=0	0.000	0.000	0.000	0.000	0.000	0.093
R^2	0.0220	0.0176	0.174	0.0801	0.0547	0.111
N	14371	13778	13319	13324	14426	14131

Note: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. The cluster standard errors at the year of birth and the region of residence at age 18 are in parentheses. We restrict to individuals who were born from 1935 to 1958. We report the estimated coefficient γ_2 from Equation 1.4.1. In *Panel (II)* and *Panel (III)*, we report the estimated coefficient α^f, α^m from Equation 1.4.4. All regression control for constant, a four categories of health at age 10, a mental health problem dummy at age 10, to be a good student at math dummy; five place of birth dummies: a big city, the suburbs of a big city, a large town, a small town or rural area; the features of the individual's dwelling at age 10: the number of books by age 10, the number of services (e.g., hot running water supply, inside toilet and others), the number of rooms, and the year fixed effects.

1.19 Appendix: Placebo Test

Table A.21: Placebo Test Similar to Table E.30 in Lippmann et al. (2020)

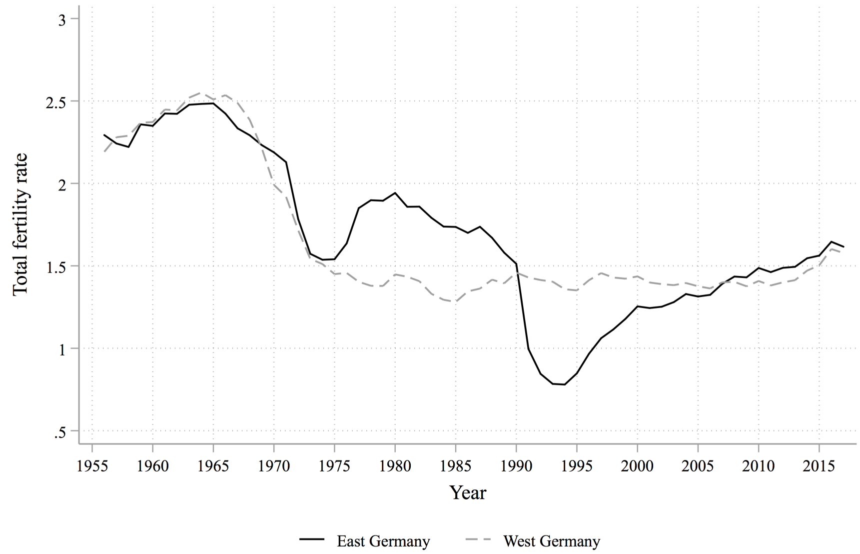
	Only interaction: "Female × USSR"			Only margin: "AMI of USSR on women"			Interaction and margin		
	(1) At 10%	(2) At 5%	(3) At 1%	(4) At 10%	(5) At 5%	(6) At 1%	(7) At 10%	(8) At 5%	(9) At 1%
6 USSR regions in Group 1	0.432***	0.390***	0.00952	0.380***	0.263***	0.110***	0.232***	0.200***	0.00952
	(0.0169)	(0.0158)	(0.00710)	(0.0161)	(0.0153)	(0.0125)	(0.0147)	(0.0137)	(0.00710)
7 USSR regions in Group 1	0.398***	0.272***	0.00667	0.316***	0.258***	0.0850***	0.216***	0.165***	0.00667
	(0.0158)	(0.0148)	(0.00664)	(0.0151)	(0.0143)	(0.0117)	(0.0137)	(0.0128)	(0.00664)
8 USSR regions in Group 1	0.373***	0.249***	0.0978***	0.111***	0.100***	0.100***	0.100***	0.100***	0.0978***
	(0.0258)	(0.0241)	(0.0108)	(0.0246)	(0.0234)	(0.0191)	(0.0224)	(0.0209)	(0.0108)
9 USSR regions in Group 1	0.520***	0.200***	0.200***	0.400***	0.400***	0.360***	0.400***	0.200***	0.200***
	(0.0775)	(0.0723)	(0.0325)	(0.0737)	(0.0701)	(0.0572)	(0.0673)	(0.0626)	(0.0325)
10 USSR regions in Group 1	1*	1*	1***	1*	1**	1**	1**	1**	1***
	(0.548)	(0.511)	(0.230)	(0.521)	(0.496)	(0.405)	(0.476)	(0.442)	(0.230)
R^2	0.340	0.262	0.0449	0.260	0.184	0.0634	0.156	0.123	0.0449
N	3003	3003	3003	3003	3003	3003	3003	3003	3003

Note: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. This table tests all of possible divisions of the 15 regions (10 in Lithuania and 5 in Poland) into two groups of respectively 10 (Group 1) and 5 (Group 2) regions. We estimate γ_2 from Equation 1.4.1, and AMI from Equation 1.4.2 as in Table 1.4 Column 4 (work experience from 25 to 50 years), changing the composition of the "USSR" dummy into a dummy for belonging in Group 1 rather than Group 2. We then define a dummy that equals 1 if the coefficients associated to "Female × USSR" and "AMI of USSR on women" are statistically significant at the relevant thresholds. We regress this dummy on the number of USSR regions in Group 1 as independent variable using Ordinary Least Squares. The omitted category is 5 USSR regions in Group 1. Column 7 displays the probability that the coefficients of interest are significant at the 10% level; column 8 at the 5% level and column 9 at the 1% level. For instance, the cell in the 8th column and 3rd row shows that with 8 USSR regions in Group 1 rather than zero, the probability that the coefficients of interest are statistically significant at the 5% level increases by 10 percentage points.

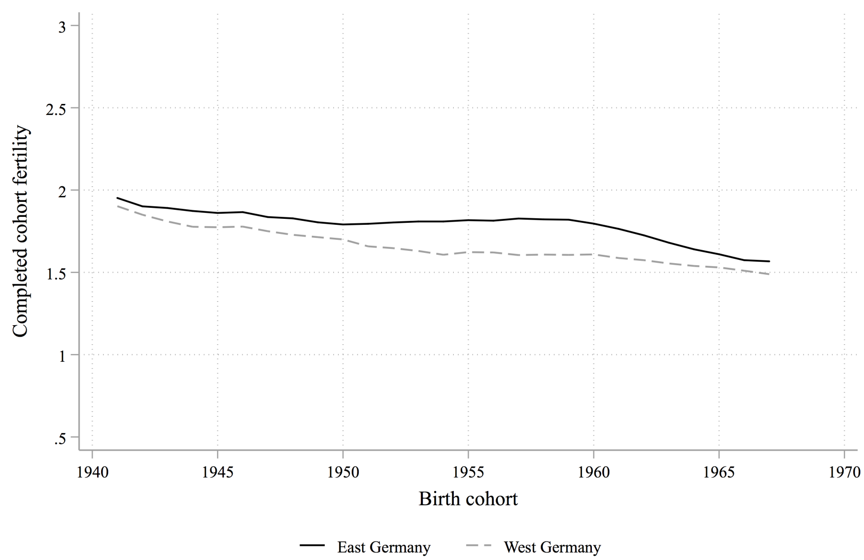
1.20 Appendix: East and West Germany

1.20.1 Fertility During Separation

Fig. A.21. Fertility Patterns in Germany



(a) Total Fertility Rate



(b) Completed Cohort Fertility

Note: Human fertility database (2019).

1.20.2 Results

Table A.22: Communism and Marriage History, Children, Later-Life Well-Being in East and West Germany

Variables	(1) Ever-married	(2) Number of marriages	(3) Age delivery	(4) Number of children	(5) Life satisfaction	(6) Life quality
	Marriage history and the number of children					
Panel I: Both men and women						
Female	0.0843***	0.0268	-3.275***	0.247***	-0.0181	-0.455
	(0.0182)	(0.0215)	(0.293)	(0.0550)	(0.106)	(0.305)
Female × East Germany	-0.0553*	-0.0214	0.280	-0.190**	-0.0146	0.159
	(0.0313)	(0.0457)	(0.474)	(0.0963)	(0.182)	(0.609)
East Germany	0.0559**	0.0636**	-2.016***	0.0781	-0.171	-0.137
	(0.0271)	(0.0315)	(0.412)	(0.0694)	(0.131)	(0.435)
AMI of East Germany on women	0.001	0.042	-1.736	-0.111	-0.186	0.022
P-value: AMI=0	0.969	0.163	0.000	0.137	0.141	0.958
R^2	0.0584	0.0424	0.207	0.0748	0.0473	0.0864
N	2139	2047	1895	1897	2226	2178
Panel II: Women						
East Germany	0.0107	0.0349	-1.786***	-0.0892	-0.243*	-0.170
	(0.0171)	(0.0293)	(0.324)	(0.0794)	(0.126)	(0.427)
R^2	0.0515	0.0576	0.138	0.0870	0.0826	0.0988
N	1129	1100	1025	1027	1152	1122
Panel III: Men						
East Germany	0.0507*	0.0662**	-2.005***	0.0810	-0.146	0.134
	(0.0284)	(0.0318)	(0.397)	(0.0686)	(0.129)	(0.420)
R^2	0.0772	0.0928	0.118	0.0867	0.0978	0.127
N	1010	947	870	870	1074	1056

Note: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. The cluster standard errors at the year of birth and the region of residence at age 18 are in parentheses. We restrict to individuals who were born from 1935 to 1958 in Germany excluding Berlin, and we consider only children born before 1990. East Germany is equal to one if a respondent was born in Brandenburg, Mecklenburg-Western Pomerania, Saarland, Saxonia, Saxonia-Anhalt, and Thuringia. It is equal to zero if Baden-Wuerttemberg, Bavaria, Bremen, Hamburg, Hesse, Lower Saxony, North Rhine-Westphalia, Rhineland-Palatinate, and Schleswig-Holstein. In *Panel (I)* we report the estimated coefficient γ_2 from Equation 1.4.1, and AMI from Equation 1.4.2. In *Panel (II)* and *Panel (III)*, we report the estimated coefficient α^f, α^m from Equation 1.4.4. All regressions control for constant, a four categories of health at age 10, a mental health problem dummy at age 10, to be a good student at math dummy; five place of birth dummies: a big city, the suburbs of a big city, a large town, a small town or rural area; the features of the individual's dwelling at age 10: the number of books by age 10, the number of services, the number of rooms, and the year fixed effects.

Chapter 2

Impact of Employment on Informal Caregiving to the Elderly Mothers in Europe

2.1 Introduction

In 2019, almost a quarter of European residents was 65 years old or more, and soon the proportion of elderly is expected to increase even further (Eurostat - European Commission, 2020). Long term care (LTC) expenditures do not necessarily cover all care loads, and elders often need to rely on informal caregivers, mainly their children. In many cases, potential caregivers work or are looking for a job. Thus, for many adults in Europe, the decisions to provide informal care and participate in the labor market influence each other: those with elder parents may have to quit their jobs to provide care, and those with less prosperous work opportunities may be more likely to become caregivers. Understanding how these two decisions are interconnected is critical to guide policy recommendations regarding formal care expenditures and more favorable labor conditions for potential care providers.

Until recently, scholars have mainly studied the impact of informal care provision on individuals' labor participation, whereas the opposite channel - how the work status affects the care choice - has been mostly neglected. Accordingly, this paper aims to close this gap in the literature and identify the causal impact of the working choice on care. Specifically, to solve the simultaneity problem between these two decisions and isolate the endogeneity in providing care, I exploit the shift in a macro-level variable (unemployment rate) due to the Great Recession that affects the individuals' propensity to care through their participation in the labor market. As a result, after controlling for observables, I find that being employed significantly reduces the probability of providing care to an elder.

This study uses the Survey of Health, Ageing, and Retirement in Europe (SHARE), retrospective SHARELIFE data, and the Job Episode Panel.[1],[2] This dataset provides a rich set of individual socio-demographic information of people above 50, including their working and care decisions. Further, I focus on the sample of twelve European countries that participate in the survey from 2004 to 2017. I model both endogenous processes jointly to capture the simultaneity of binary care and work decisions. Moreover, to identify the causal impact of work on care, I consider the following exclusion restriction for the work choice: the country's exposure to the Great Recession measured through changes in the unemployment rate. To quantify the impact of the crisis on labor participation, I follow Costa-Font, Karlsson, and Øien (2016) and Mommaerts and Truskinovsky (2020), and consider a change in the unemployment rate in a country. Further, I allow for a heterogeneous effect of the crisis across men and women in the same country. The identification strategy allows for the correlation of unobservables in work and care decisions. The analysis focuses on care provided to mothers as they are more likely to receive care from working-age individuals (Barczyk and Kredler (2019) and Bonsang and van Soest (2020) for the role of intra-household care and home production in couples). Moreover, this paper studies both extensive and intensive margins; the later is measured through care provision frequency.

The main identification strategy assumes that the Great Recession does not directly affect the care choice after controlling for an extensive set of observed variables, such as work experience, the health of caregiver and care receiver, residential proximity, and proxies for prices and availability of formal care. To construct work experience, I exploit the SHARELIFE data and the Job Episode Panel that were part of the SHARE survey in wave 3 (2009) and wave 7 (2017).

[1] This paper uses data from SHARE Waves 1, 2, 3, 4, 5, 6 and 7 (DOIs: 10.6103/SHARE.w1.700, 10.6103/SHARE.w2.700, 10.6103/SHARE.w3.700, 10.6103/SHARE.w4.700, 10.6103/SHARE.w5.700, 10.6103/SHARE.w6.700, 10.6103/SHARE.w7.700), see B "orsch-Supan et al. (2013) for methodological details.(1) The SHARE data collection has been funded by the European Commission through FP5 (QLK6-CT-2001-00360), FP6 (SHARE-I3: RII-CT-2006-062193, COMPARE: CIT5-CT-2005-028857, SHARELIFE: CIT4-CT-2006-028812), FP7 (SHARE-PREP: GA N°211909, SHARE-LEAP: GA N°227822, SHARE M4: GA N°261982) and Horizon 2020 (SHARE-DEV3: GA N°676536, SERISS: GA N°654221) and by DG Employment, Social Affairs & Inclusion. Additional funding from the German Ministry of Education and Research, the Max Planck Society for the Advancement of Science, the U.S. National Institute on Aging (U01_AG09740-13S2, P01_AG005842, P01_AG08291, P30_AG12815, R21_AG025169, Y1-AG-4553-01, IAG_BSR06-11, OGHA_04-064, HHSN271201300071C) and from various national funding sources is gratefully acknowledged(see `www.share-project.org`).

[2] See Brugiavini et al. (2019) for details about the Job Episode Panel.

This retrospective information allows for tracking life decisions from birth to the moment of the interview. Cumulative work experience is one of the key predictors of employment among elders, so I restrict the sample to individuals who participated in the SHARELIFE survey.

This study documents an increase by nine p.p. in the probability of providing informal care to a mother when an adult child does not participate in the labor market. Female caregivers mainly drive this result.

This paper contributes to the literature about informal care provision. Up to my knowledge, this paper is the first one to study the causal impact of labor participation on care provision in Europe. The majority of papers investigates the opposite channel - the impact of care on the employment status. Literature in this field often uses a two-stage least squares strategy and proposes an instrumental variable to correct potential endogeneity in the decision to give informal care. Early works used as an IV the living proximity between a potential caregiver and a carereciever (Bolin, Lindgren, and Lundborg, 2008a,b; Bonsang, 2009; Balia and Brau, 2014), the gender of the children of a carereciever (Bolin et al., 2008b; Bonsang, 2009), a carereciver's partner alive indicator (Bolin et al., 2008b; Van Houtven, Coe, and Skira, 2013), the share of disabled people in the household (Heitmueller, 2007), and others. However, due to the nature of these instruments, the endogeneity concern can potentially remain. More recent studies agree upon using a carereciever's poor health status as an instrument for informal care provision (Van Houtven et al., 2013; Crespo and Mira, 2014). Moreover, some scholars propose to exploit the dynamic framework and correct for unobserved characteristics directly (Michaud, Heitmueller, and Nazarov, 2010; Casado-Marín, García-Gómez, and López-Nicolás, 2011; Jiménez-Martín and Vilaplana Prieto, 2015; Ciccarelli and van Soest, 2018). Still, the current findings of the impact of care choice on employment decisions are mixed and vary across countries and age of potential caregivers' target sample. Specifically, they range from negative and significant to insignificant (Bolin et al. (2008b), Crespo and Mira (2014), Ciccarelli and van Soest (2018) for Europe, Heitmueller (2007) for the UK, and Van Houtven et al. (2013) for the USA).

Accordingly, this study adds to the literature by considering the impact of participation on care choice. Moreover, I directly exploit both dependent variables' binary nature and apply a bivariate model to compute the discrete change in work over care variable In this setting, the average partial effect more likely leads to an appropriate approximation comparing with the local linear approximation. Finally, I contribute to the list of articles based on the SHARE data by documenting the sample selection problem in conducting a panel data analysis. Specifically, in this case, the target sample should include respondents who are present at least twice in the survey: 1) being above 50, 2) being below statutory retirement age, and 3) having a mother alive in that periods. These restrictions lead to a positive selection of individuals regarding observable characteristics, such as age, health, and education. Even though controlling for permanent unobserved heterogeneity can be important in this context, any study related to informal care provision to the elderly should not neglect the sample selection problem, in particular, using the SHARE data.

The result of this analysis - a significant increase in care responsibilities in the absence of employment opportunities in Europe - has potential policy implications directly related to the trade-off between work and care to elders. In countries with a growing percent of elders, policymakers may be tempted to keep more individuals into the labor market to sustain a larger share of non-working elder population. However, if potential caregivers stay longer in the market, according to the results of this analysis, they will provide less care to elders. Keeping more individuals in the labor market may be beneficial for a variety of reasons, but policymakers should be conscious that it can reinforce the problem of unmet needs in elders' care when formal care provision is scarce.

The paper proceeds as follows. Section 2.2 describes the dataset and the exclusion restriction. In Section 2.3, I discuss the identification strategy and its limitations. Section 2.4 provides the main empirical findings. Section 2.5 shows robustness checks. Section 2.6 concludes. Additional tables and figures are in the Appendix.

2.2 Data and Descriptive Findings

In this Section, first, I discuss the SHARE data used in the paper. Next, I explain the impact of the Great Recession in Europe.

2.2.1 SHARE Data

This paper exploits the Survey of Health, Ageing and Retirement in Europe (SHARE), the SHARELIFE, and the SHARE Job Episodes Panel. Eligible participants in the SHARE survey are above 50 years old. The main survey provides socio-demographic, health, and financial information about individuals across European countries. It is a biannual survey, and this analysis uses wave 1 (2004/05), wave 2 (2006/07), wave 3 (2008/09), wave 4 (2010/11), wave 5 (2013), wave 6 (2015), and wave 7 (2017). To get a comparable number of periods before and after the crisis, I restrict to countries present before 2008 but Switzerland.[3] Following the literature (Barczyk and Kredler, 2018), I group countries according to LTC expenditures and established social norms: Denmark, the Netherlands, and Sweden (Northern); Austria, Belgium, Germany, and France (Central); Greece, Italy, and Spain (Southern); and the Czech Republic and Poland (Eastern).

Wave 3 and a part of wave 7, the so-called SHARELIFE survey, cover retrospective information about respondents. I exploit the Job Episode Panel based on the SHARELIFE survey to compute respondents' years of work experience. I focus on respondents below the country-specific statutory retirement age that varies across gender (see Appendix 2.7.1, Table B.1). Accordingly, I restrict further to respondents who participated either in wave 3 or wave 7.

In the main analysis, I study informal care to mothers because older women are the main carereceivers of working-age individuals. This is mainly due to their longer longevity and the age difference between spouses in the past. Accordingly, I restrict my sample to respondents whose mothers are alive. Considering also care to fathers does not change findings.

In total, there are 9982 individuals for whom I have 16811 observations, and on average, I observe each respondent 1.68 times. Many individuals appear only once in my sample.[4] Dynamic estimates should be taken with caution in this setting. Among respondents present for several periods, I document a positive selection in terms of the caregiver's socio-demographic variables and mother's health. Since these characteristics directly correlate with participation and care choices, a panel analysis suffers from the sample selection problem (Section 2.3.3 for details).

Employment variable. To get the main dependent variables of interest, I consider potential caregivers' current job situation, denoted as *work*. If the respondent answers to be employed or self-employed, then *work* equals one. If a respondent says to be unemployed, a homemaker, permanently sick, other, or retired, then it equals zero.[5] I do not distinguish between unemployment and retirement status since the principal interest of the paper is to determine the impact on care choice of non-participating in the labor market that is having more time.[6]

Care variables. The SHARE respondents answer a broad list of questions about given

[3] Bolin et al. (2008b), Bonsang (2009), and Barczyk and Kredler (2019) point out the problems in data collection in Switzerland.

[4] There is a considerable data loss because it is a biannual panel, and wave 3 lacks information about care provision.

[5] Van Houtven et al. (2013) define a similar classification.

[6] Michaud et al. (2010) and Crespo and Mira (2014) restrict the sample to respondents below 60 years, but their aim to measure the impact of care provision on the work decision. Van Houtven et al. (2013) consider both work and retirement decision, so they include respondents between 50 and 70 years old.

informal care. First, I examine the binary indicator, *care*, that equals one if informal care is provided either inside or outside the household to a mother and zero otherwise.[7] I pool together two types of informal care because only about 5 percent of respondents answer that they provide care inside the household. Table 2.1 shows the summary statistics of the extensive margin of informal care provision across the caregivers' gender. One-third of individuals in my sample reports to take care of their mothers at some point in time, and this share is larger among women.

Table 2.1: Care Choice Across Genders

	Care to a mother (%)		
	(1) No	(2) Yes	(3) Observations
Caregiver's gender:			
Female	6108 (63.76)	3471 (36.24)	9579 (100.00)
Male	5448 (75.33)	1784 (24.67)	7232 (100.00)
Total	11556	5255	16811

Note: Author's computations using the SHARE data.

Next, I exploit information about the frequency of care to get a proxy for the intensive margin of care provision. I look only at given care outside the household since there is no detailed information about care frequency for those who reported giving care inside the household.[8] Table 2.2 shows that one-fifth of female caregivers provide care on a daily basis. Unarguably, they are the ones who face the highest burden. Accordingly, I define the second dependent variable, *daily care*. The share of daily caregivers is smaller among men, which is in line with previous studies based on the SHARE (Crespo and Mira, 2014). In Appendix 2.7.2, Table B.2 shows the North-South gradient in daily caregiving across Europe: the intensive care provision is more common in Southern European countries than in Northern ones.

The SHARE data collection is different in waves 4 and 5: the question about care provision is asked only to a family respondent at an individual level. To overcome this obstacle, I restrict

[7] I get this information from two questions. SP008: *In the last twelve months, have you individually given any kind of help listed on this card to a family member from outside the household, a friend or neighbour?* SP018: *Is there someone living in this household whom you have helped regularly during the last twelve months with individual care, such as washing, getting out of bed, or dressing? Using followed up questions I can identify to whom given help.*

[8] I use the following question: SP011: *In the last twelve months, how often altogether have you given personal care or practical household help to this person? The answers are 1. Almost daily 2. Almost every week 3. Almost every month 4. Less often*

Table 2.2: Care Intensity Across Genders

	Frequency of care (%)				
	(1) Daily	(2) Weekly	(3) Monthly	(4) Less often	(5) Observations
Caregiver's gender:					
Female	737	1359	724	506	3326
	(22.16)	(40.86)	(21.77)	(15.21)	(100)
Male	258	592	490	380	1720
	(15.00)	(34.42)	(28.49)	(22.09)	(100)
Total	995	1951	1214	886	5046

Note: Author's computations using the SHARE data.

to family respondents in waves 4 and 5 and all respondents in the remaining waves. Ciccarelli and van Soest (2018) face the same challenge and restrict the analysis to family respondents in all waves. In Appendix 2.9.3, I repeat the original findings on the full sample and the family respondents sample.

Socio-economic variables. In the analysis, I control for a broad list of potential caregivers and their mothers' characteristics. Table 2.3 presents these variables, and columns 1 and 2 report the mean value of two subsamples of respondents: those who never provide daily care to a mother and those who do it at least once during the survey period. All the statistics correspond with the first wave an individual participates in the survey. It follows that caregivers are more likely to be women in good health with fewer siblings and kids; they have a slightly better education, and their mothers are in worse health. The father's alive indicator captures the importance of partner care among elders.

The target sample includes people above 50. At that age, labor supply choice is likely very persistent; thus, control variables include the education level and age. However, work history is also an important determinant of labor participation. Accordingly, using the Job Episodes Panel, I construct the cumulative years of work experience by 2005, the year before the crisis. Appendix 2.7.3, Fig. B.1 plots the distribution of work experience for men and women. The distribution for women is bimodal, and it suggests the positive mass of individuals who never worked at all. However, the profile of men varies, the distribution is right-skewed, meaning that often potential male-caregivers in my sample did work all the time.

Table 2.3: Descriptive Statistics of Respondents

	Never a daily caregiver (1)	SD	A daily caregiver at least once (2)	SD	Difference (1) - (2)	(p-value)
Socio-demographic characteristics						
Age	54.82	4.08	55.35	4.21	-0.53	0.00
Female	0.55	0.50	0.73	0.44	-0.18	0.00
Married	0.79	0.41	0.76	0.43	0.03	0.06
Number of children	2.08	1.19	1.82	1.04	0.27	0.00
Number of brothers	1.16	1.19	0.90	1.05	0.25	0.00
Number of sisters	1.12	1.19	0.84	1.03	0.28	0.00
Education						
Primary education or less	0.12	0.33	0.14	0.34	-0.01	0.25
Lower secondary education	0.16	0.37	0.20	0.40	-0.04	0.00
Upper secondary education	0.38	0.48	0.40	0.49	-0.02	0.17
Post-secondary education or above	0.34	0.47	0.26	0.44	0.08	0.00
Self-rated health						
Very Good	0.42	0.49	0.37	0.48	0.05	0.00
Good	0.36	0.48	0.39	0.49	-0.02	0.20
Fair	0.16	0.37	0.20	0.40	-0.04	0.01
Poor	0.05	0.22	0.04	0.20	0.01	0.27
Number of mobility limitations	0.82	1.61	0.87	1.54	-0.05	0.38
Born in the same country	0.94	0.25	0.96	0.20	-0.02	0.01
Owner	0.76	0.43	0.79	0.41	-0.03	0.11
Non-wage net income, K euro PPP	26.76	41.44	29.84	42.62	-3.08	0.05
Participation characteristics						
Work	0.64	0.48	0.52	0.50	0.12	0.00
Work experience by 2005	26.96	10.84	26.82	11.66	0.13	0.74
Care to a mother						
Care	0.24	0.43	0.71	0.45	-0.47	0.00
Outside the household	0.23	0.42	0.70	0.46	-0.47	0.00
Inside the household	0.01	0.11	0.04	0.19	-0.02	0.00
Daily care	0.00	0.00	0.52	0.50	-0.52	0.00
Mother's self-rated health						
Poor health	0.15	0.36	0.22	0.41	-0.07	0.00
Mother's characteristics						
Mother's age	80.71	5.83	82.17	5.53	-1.46	0.00
Father is alive	0.33	0.47	0.24	0.43	0.08	0.00
Residential proximity to mother						
Co-resident	0.06	0.25	0.20	0.40	-0.13	0.00
Less than 25 kms	0.57	0.50	0.74	0.44	-0.17	0.00
Less than 500 kms	0.29	0.46	0.05	0.22	0.24	0.00
More than 500 kms	0.08	0.26	0.01	0.11	0.06	0.00
Observations	6841		796			

Note: The list of observed characteristics in the analysis with the SHARE. All the characteristics are reported the first time an individual participates in the survey. Columns 1 and 2 correspond with individuals who never at least once provide daily care to a mother. The last column reports the P-value of the null hypothesis about the equalities of the two means.

2.2.2 Measure of the Great Recession

I exploit the country's exposure to the recent Great Recession as an exclusion restriction for labor participation in my identification strategy. This global macroeconomic downturn is unlikely correlated with an individual unobserved propensity to care after controlling for the observed variables. To measure the size of the impact of the Great Recession, I follow Costa-Font et al. (2016). I use quarterly data for GDP per capita and the quarterly gender-specific unemployment rate from Eurostat.[9]

First, I define the period of crisis as all the quarters with negative output growth (see Appendix 2.7.4, Fig. B.2; I consider the exposure to the crisis among men in Austria). Second, I map that time interval (the green vertical lines in Fig. B.2 in Appendix 2.7.4) to unemployment data and construct the percentage point change between the period with the last negative growth (the last crisis quarter) and the last positive growth (before the crisis). I repeat the same exercise separately for men and women and for each country in my sample. The summary statistics of the measure of crisis are in Table 2.4.

In waves 1 and 2 (before the crisis), I assign 0 to the measure of *crisis*. For the waves after the Great Recession, (wave 4 - 7), *crisis* varies across genders and countries. Accordingly, I assign the value shown in the corresponding cell in Table 2.4. There is significant difference across countries; however, in each country, women were less affected (see columns 3 and 4) than men. This observation is in line with Dolado, García-Penalosa, and Tarasonis (2020), who provide an excellent study of the impact of the Great Recession across Europe.

[9]From Eurostat, I consider the variable namq_10_gdp for quarterly GDP per capita, and une_rt_q for the quarterly gender-specific unemployment rate.

Table 2.4: Great Recession Across European Countries

	Pre-crisis unemployment rate		Crisis impact	
	(1) Men	(2) Women	(3) Men	(4) Women
Northern countries:				
Denmark	1.8	2.8	3.4	1.8
The Netherlands	1.8	3.8	0.3	-0.09
Sweden	3.8	4.4	2.6	1.3
Central countries:				
Austria	3.2	3.8	1.30	0.50
Belgium	5.3	6.3	0.89	0.59
Germany	7.2	7.7	-0.09	-.79
France	5.6	6.5	1.5	1.2
Southern countries:				
Greece	4.1	10.5	1.30	.60
Italy	4.2	7.2	1	.40
Spain	7.7	10.5	6.5	4.9
Eastern countries:				
Czech Republic	2.8	5.1	1.9	1.9
Poland	5	6.5	0	0

Note: Author's computations using data from Eurostat. Pre-crisis unemployment rate corresponds with 2007. Unemployment by sex and age - quarterly average (une_rt_q). Seasonally adjusted data, not calendar adjusted data. The percentage of the active population is aged 25 - 74 years. The pre-crisis value of unemployment in Poland for men and women corresponds with the third quarter of 2008, the same as in the Czech Republic.

2.3 Empirical Strategy

In this Section, I explain the identification strategy and discuss its potential failures. I conclude this Section with a discussion about the selection into the dynamic specification and why it can be misleading to use the panel analysis in this context.

The naive way to evaluate the impact of work decision on care is to estimate a single-equation model, where the decision to take care is regressed on the work variable. However, this approach leads to biased estimations due to omitted variable bias. Prior to the analysis, the direction of bias is not clear. Let me provide two examples. First, consider an individual whose mother devoted a lot of time to his education in early childhood. Then, one would expect this individual to find a job easily later in life due to larger human capital. Moreover, once his mother gets older, he may be more willing to take up the care of her due to a repayment motive for her time earlier. In this case, a probit regression would mistakenly attribute early life care to a work variable, and the coefficient would be *downward biased*.

However, the wealth of the family may affect at the same time the mother's propensity to devote time to child education and job prospect of her children. Accordingly, the coefficient of work on care could be *upward biased*.

Apart from the bias, the sign of the coefficient of work on care is ambiguous. We may consider a third time-spending activity on top of work and care, for example doing home chores (or leisure). Depending on whether care is a compliment or a substitute for the third activity, different scenarios arise. When care is provided at the cost of reduced work and not leisure, the coefficient becomes negative. Otherwise, an individual may decide to compensate extra time on care with a decrease in domestic chores hours to keep the same work.

2.3.1 Identification

To get a consistent estimate, I propose the following bivariate model:

$$\begin{aligned} care_{it} &= \mathbb{1}[\alpha_1 work_{it} + \alpha_2 mpoor_{it} + \beta x_{it} + \theta_t + \omega_{g(i)t} + u_{it} > 0] \\ work_{it} &= \mathbb{1}[\gamma_1 z_{g(i)t} + \gamma_2 mpoor_{it} + \delta x_{it} + \theta_t + \omega_{g(i)t} + v_{it} > 0], \end{aligned} \tag{2.3.1}$$

where $care_{it}$ is a dummy variable that equals 1 if an individual i at year t provides care to a mother. $work_{it}$ is a dummy variable if an individual i at year t works and 0 if he is unemployed, a homemaker, permanently, retired, or other. $mpoor_{it}$ equals 1 if a mother is in poor health. z_t is the measure of a country's exposure to the Great Recession that varies across genders, countries and periods (before or after 2008). x_{it} is a set of controls for individual i at year t: age, age squared, health status, married dummy, four health dummies, number of limitations with daily activities, four education dummies, number of children, number of brothers, number of sisters, cumulative years of work experience by 2005, non-wage income and a dummy for being a homeowner; and mother's characteristics: four dummies for residential proximity to a caregiver, mother's age, and partner alive dummy. θ_t is survey year dummies. Given the variation in state labor market conditions and state support for long term care, I control for country effects and country-specific linear time trends, $\omega_{g(i)t}$. Further, to allow for the common unobservables in both decisions, I assume that u_{it} and v_{it} are distributed as bivariate normal, with $\mathbb{E}(u_{it}|work_{it}, mpoor_{it}, x_{it}, \theta_t, \omega_{g(i)t}) = \mathbb{E}(v_{it}|work_{it}, mpoor_{it}, x_{it}, \theta_t, \omega_{g(i)t}) = 0$, $\text{var}(u_{it}|work_{it}, mpoor_{it}, x_{it}, \theta_t, \omega_{g(i)t}) = \text{var}(v_{it}|work_{it}, mpoor_{it}, x_{it}, \theta_t, \omega_{g(i)t}) = 1$ and

$\text{cov}(u_{it}, v_{it}|work_{it}, mpoor_{it}, x_{it}, \theta_t, \omega_{g(i)t}) = \rho$. The sign of this correlation coefficient sheds light on the type of selection in the endogenous equation when a participation variable is treated like exogenous.

This bivariate model allows to estimate the direct effect of work on care, α_1. Moreover, I control for mother's poor health status in care decision since it is a good predictor for care provision, α_2, and in the work equation, as it is an implicit measure of the impact of given care on work, γ_2. This model is properly identified if, after I control for the observed characteristics, the country's exposure to the Great Recession is uncorrelated with care choice.Following Costa-Font et al. (2016), I cluster standard errors at the country level to allow for correlation in job and care attachments due to social norms and labor market conditions across the respondents living in the same country.

In the article, first, I report the coefficients from 2.3.1. Next, for the qualitative importance of the work dummy variable in the care equation, I report the average partial effect (APE) that is the average difference between the probability of providing informal care if he works and the probability if he does not work. The average partial effect for the entire sample equals to $1/n \sum_i [\Phi(\alpha_1 + \alpha_2 mpoor_{it} + \beta x_{it} + \theta_t + \omega_{g(i)t}) - \Phi(\alpha_2 mpoor_{it} + \beta x_{it} + \theta_t + \omega_{g(i)t})]$. I use bootstrap standard errors with 50 draws to calculate the standard errors of the APE.

2.3.2 Potential Omitted Variables Bias

My analysis holds if the crisis impacts care choice only through the work choice and the other control variables.

Here I explain the potential scenarios in which the exclusion restriction can fail and how I overcome these limitations of the analysis.

First, the Great Recession can have an impact not only on employment choice but also on formal care provision, mainly paid care in nursing houses or at home. Then, to get consistent estimates using my identification, I need to control for formal care. To do so, I propose several values of formal care. First, I include country LTC expenditures in euro per inhabitant.[10] Then, I exploit the direct measure of the quantity of formal care: LTC beds in nursing and residential care facilities per hundred thousand inhabitants.[11] Next, I hypothesis if not only the opportunity cost of time of carerecievers decreases during the crisis but also the opportunity cost of paid care providers changes. In this case, not only the quantity of care matters but also the prices for purchased care. Unfortunately, there are no official statistics to make cross-country and overtime comparison. However, I use workers' earnings in the service sector as a

[10] Eurostat: ICHA11_HC
[11] OECD: variable hlth_rs_bdsns

proxy for earnings of paid care providers.[12] Barczyk and Kredler (2019) face a similar problem when they compute the market value of informal care and argue that the formal care provided at home requires the basic skills, and the official minimum wage can be considered as a relevant proxy. In general, the minimum wage does not decrease as a response to the crisis. Accordingly, in my case, I do not exploit it as a proxy for prices for formal care. In Section 2.5, I report the results for all four specifications and the main findings hold.

However, there is no consensus about the relationship between formal and informal care in the literature, if they are substitutes or complements. Depending on the dataset, authors analyze the paid care at home and/or staying in nursing houses. The findings are mixed. Using the SHARE data, Bonsang (2009) documents that formal home care and informal care are substitutes, whereas this pattern gets weaker if the carereciever's health worsens. Moreover, informal care is a complement to formal care provided at the nursing home. Moreover, Bolin et al. (2008a) and Balia and Brau (2014) show that informal and formal care are substitutes. In a recent study, Barczyk and Kredler (2019) point out that care tends to be concentrated on the one source. In particular, formal care in a nursing home is a clear substitute for informal care. In the US context, Van Houtven and Norton (2004) find evidence for substitutability.[13] Still, it is plausible to believe that the formal care availability correlates with its usage, and I can control for the different measures of formal care in a country in Section 1.7.

Second, the other problem for my identification may arise if the crisis affects the residential proximity directly (or through the employment decision) and an individual decides to provide care because he lives closer. Contrarily, people may move far to find a job in response to the Great Recession and not being able to care.[14] To argue that people do not move as a response to the crisis, I use the multinomial regression for the proximity and the same set of controls as in the main specification. *Panel I* in Table 2.5 reports no evidence that the crisis is significantly related to residential proximity. In Appendix 2.8.1, Table B.3 shows a significant persistence across periods, and very few individuals, whom I observe during two periods, co-reside with their elder parent. Moreover, there is no evidence for a change in this pattern after the Great

[12] ILO: ISCO-08 `https://www.ilo.org/shinyapps/bulkexplorer51/?lang=en&segment=indicator&id=EAR_4MTH_SEX_OCU_CUR_NB_A`

[13] It can be a promising avenue of the analysis to model work, informal and formal care decisions. However, due to data limitation, I cannot include the third simultaneous choice about formal care in my specification; this information is not available in caregivers' sample. Studying the trade-off between work and care choices is my primary interest; that is why I need to include the socio-demographic variables from their side. To include the formal care usage from the SHARE data, I need to look from a carereciever perspective, and it will lead to the loss of relevant information about caregivers.

[14] Due to data limitation I cannot identify the effects on the subsample of people whom I observe in pre-crisis waves and look at residential proximity before the Great Recession.

Recession.

Table 2.5: Impact of the Crisis on Controls

	Panel I: Impact of the Crisis on Residential Proximity			
	Co-reside	Less than 25 kms	Less than 500 kms	More than 500 kms
	(1)	(2)	(3)	(4)
Dependent variable: Residential proximity to a mother				
Crisis	-0.0148	-0.000665	0.00940	0.00607*
	(0.00920)	(0.00645)	(0.0118)	(0.00366)
Work	-0.0102	0.00901	0.0263	-0.0251***
	(0.0122)	(0.0355)	(0.0215)	(0.00935)
Mother is in poor health	0.0184	-0.0481*	0.00746	0.0222***
	(0.0142)	(0.0257)	(0.0141)	(0.00616)
Observations	16788	16788	16788	16788
	Panel II: Impact of the Crisis on Caregiver's Health			
	Very Good	Good	Fair	Poor
	(1)	(2)	(3)	(4)
Dependent variable: Caregiver's health				
Crisis	-0.00482	-0.00274	0.00746	0.000101
	(0.00515)	(0.00571)	(0.00483)	(0.00290)
Work	0.0648***	0.0126*	-0.0356**	-0.0418***
	(0.0150)	(0.00732)	(0.0168)	(0.00183)
Mother is in poor health	-0.0761***	-0.00679	0.0597***	0.0233*
	(0.0239)	(0.0158)	(0.0126)	(0.0130)
Observations	16788	16788	16788	16788

Note: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. The average marginal effects are reported. The cluster standard errors in the country of a caregiver are in parentheses. I restrict the sample to people between 50 years old and statutory retirement age in twelve European countries: Austria, Belgium, the Czech Republic, Denmark, Germany, Greece, France, Italy, the Netherlands, Poland, Spain, Sweden. The observation period includes six interview waves that span from 2004 to 2017. In *Panel I* and *Panel II*, the set of control variables includes caregivers' characteristics: age, age squared, married dummy, four health dummies, limitations with daily activities, four education dummies, years of work experience by 2005, non-wage income, number of children, number of brothers, and number of sisters; mother's characteristics: mother's age, and partner alive dummy; and interview year dummies, country-specific linear trend, and country dummies. Further, in *Panel II*, I control for four dummies for residential proximity to a caregiver.

Next, similar to the previous point, my identification becomes invalid if, as a result of the crisis, the carereciever's health worsens, leading to an increase in care needs. To rule out this scenario, I compute the average marginal impact of the Great Recession on the probability that caregiver's mother is in poor health after controlling for the same set of variables from the main specification. I find that AME is equal to 0.006 and statistically insignificant. Accordingly, it supports the proposed identification strategy that the Great Recession affects the care decision primarily through the caregiver's employment status.

Finally, I study the impact of the Great Recession on caregiver's health. *Panel II* in

Table 2.5 shows the results. Accordingly, the Great Recession does not impact statistically significant the caregiver's health, and so, its impact on care provision cannot be through this variable in the sample of this study.

2.3.3 Sample Selection in Dynamic Specification

In the main analysis, I control for age, educational attainment, work experience by 2005 as a determinant of labor participation. However, the working status in the previous wave can also matter. Excluding lagged values can lead to an omitted variable problem. On the other hand, the panel analysis requires selecting individuals who fulfill the criteria to be present at least two waves and have a mother alive in both periods. Below I explain the drawback of restricting the sample.

First, including lagged values in the model leads to a significant data loss: from 16788 to 9111 individuals-year observations. Second, this shrinkage in the sample size is non-random. I document that individuals who satisfy the panel analysis criteria are positively selected and sensible for labor participation and the mother's care needs.

To test it, I split the sample of respondents into two groups. The first group represents individuals who participate only in a cross-section and then exit the analysis: 3810 respondents of this type. The second group includes the initial wave of respondents who are a part of the panel sample: 3867 individuals. In Appendix 2.10, Table B.7 reports the mean values of caregivers and their mothers' observable characteristics and the P-value of the null hypothesis about the equality of means. Indeed, potential carerecievers differ, on average, across these two groups. Individuals who become present in the panel analysis are more likely to have a younger mother in better health, decreasing propensity to provide care. Their father is also more likely to be alive, and he can be a sensitive source of care provision. Moreover, individuals who can contribute to the dynamic specification are younger, less likely to be married, slightly in better health, and have fewer kids. All these determinants increase the probability of being in the labor market.

To reinforce the sample selection argument for the panel analysis, I run the probit regression about the chances to participate in the dynamic analysis versus exiting the sample (0 if a respondent belongs to the first group and 1 to the second one described above). I also include the same list of explanatory variables as in the other specifications. In Appendix 2.10, Table B.8 shows that the respondents do vary in essential characteristics. If a mother is in poor health or is older, then a respondent is less likely to be in the potential panel analysis.

Furthermore, individuals who are present in the panel study are significantly more educated, which is an important driver of labor participation.

Even though care and participation decisions are likely to be dynamic decisions, the panel data is inherently selected. The work-care trade-off can be more sensitive for individuals who are excluded from the panel study. Accordingly, the dynamic analysis by including lagged values can lead to problematic results based on a non-representative sample. Therefore, I prefer the cross-section specification given the vast list of explanatory variables and, importantly, the usage of the cumulative years of work experience before the crisis.

2.4 Results

In this Section, first, I report the main findings for care and daily care provision. Next, I discuss the heterogeneity across genders.

2.4.1 Main Findings

Table 2.6 reports the main results of the employment status on providing informal care and potentially more demanding daily informal care. Columns 1 and 5 show the results without taking into account endogeneity. From the one-equation model, it follows that being employed is not associated with informal care provision and negatively associated with daily care provision. In line with the literature, a mother in poor health is a strong predictor of the care provision. The estimates of Equation 2.3.1 are in columns 2 and 6. Once I correct the endogeneity of care choice, then work has a negative and statistically significant impact on the extensive margin of informal care and daily care conditional on observables (see columns 3 and 6). To control for the possible adverse impact of care on work, I include the mother's poor health status in the work equation, and the negative sign of the coefficient is consistent with previous studies based on the SHARE. In Appendix 2.8.2, Table B.4 reports the same regressions with the full set of controls. Moreover, I also show that the coefficients in a bivariate model of daily care do not change across the sample of family respondents and all respondents (see Appendix 2.9.3, Table B.6).

The correlation of unobservables is positive for occasional and daily informal care (see the rho coefficient in Table 2.6). It implies that there are common factors in errors that change the propensity to work and give care to the mother in the same direction. The unobservable drivers make a person more likely to participate and to take up care. In the endogenous specification, the positive selection of individuals in terms of their own characteristics to participate in the labor market is attributed to the impact of work on care. However, once I control for this endogeneity using the bivariate model, the impact is measured for all individuals.

The magnitude of coefficients in Table 2.6 cannot be interpreted as the size of the effect because it shows the results of the bivariate model. That is why, at the bottom of Table Table 2.6, I report the average partial effects of work status in the care decision. Participation in the labor market decreases the probability of providing care by about 9 percentage points. The impact is significant for both care and daily care provision. Next, I study the heterogeneity of this effect across genders.

Table 2.6: Maximum Likelihood Estimation of Work and Care Choices

	Informal care				Daily informal care			
	Ignoring endogeneity		Bivariate probit		Ignoring endogeneity		Bivariate probit	
	(1) Care	(2) Work	(3) Care	(4) Work	(5) Care	(6) Work	(7) Care	(8) Work
Bivariate probit coefficients								
Work	0.00911		-0.298**		-0.180***		-0.716***	
	(0.0864)		(0.121)		(0.0481)		(0.0641)	
Crisis		-0.0613***		-0.0514**		-0.0599***		-0.0521**
		(0.0218)		(0.0219)		(0.0218)		(0.0216)
Mother is in poor health	0.352***	-0.188***	0.332***	-0.188***	0.628***	-0.187***	0.585***	-0.194***
	(0.0669)	(0.0607)	(0.0548)	(0.0619)	(0.0710)	(0.0606)	(0.0550)	(0.0638)
rho			0.183				0.321	
Likelihood	-35384757.9	-33871651.4	-69245879.8		-12991181.3	-33837772.3	-46805006.8	
Observations	16811	16811	16811		16788	16788	16788	
Average partial effect of work decision on care choice								
Average partial effect	.00276		-.0934*		-.0186***		-.0862***	
95 % bootstrap s.e.	(.0255)		(.0525)		(.00508)		(.0316)	

Note: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. The cluster standard errors in the country of a caregiver are in parentheses. I restrict the sample to individuals between 50 years old and statutory retirement age in twelve European countries: Austria, Belgium, the Czech Republic, Denmark, Germany, Greece, France, Italy, the Netherlands, Poland, Spain, Sweden. The observation period includes six interview waves that span from 2004 to 2017. The dependent variable varies across columns. In columns (1) and (5), it is a decision to care outside the household and to provide daily care respectively; in columns (2) and (6), it is a decision to participate; and in columns (3)-(4) and (7)-(8), I estimate a bivariate model of both decisions to work and to care. The set of control variables in all regressions includes caregivers' characteristics: age, age squared, married dummy, four health dummies, limitations with daily activities, four education dummies, years of work experience by 2005, non-wage income, number of children, number of brothers, and number of sisters; mother's characteristics: four dummies for residential proximity to a caregiver, mother's age, and partner alive dummy; and interview year dummies, country-specific linear trend, and country dummies. In the bivariate model, the average partial effect of work decision on care choice is equal to $1/n \sum_i [\Phi(\alpha_1 + \alpha_2 mpoor_{it} + \beta x_{it} + \theta_t + \omega_{g(i)t}) - \Phi(\alpha_2 mpoor_{it} + \beta x_{it} + \theta_t + \omega_{g(i)t})]$.

2.4.2 Heterogeneity Across Genders

First, I report the results from Table 2.6 for the sample of men and women in columns 1 and 2 in Table 2.7. Next, I repeat the analysis separately by gender of potential caregivers. Interestingly, the Great Recession has a positive impact on female participation in line with Dolado et al. (2020), in which they explain an increase in women's labor participation during the last crisis in Europe through an added worker effect.

Table 2.7: Maximum Likelihood Estimation of Work and Care Choices Across Genders

	All		Male		Female	
	Bivariate probit		Bivariate probit		Bivariate probit	
	(1) Care	(2) Work	(3) Care	(4) Work	(5) Care	(6) Work
Bivariate probit coefficients						
Work	-0.715***		-0.631		-0.652***	
	(0.0638)		(0.616)		(0.191)	
Crisis		-0.0520**		-0.176***		0.0695***
		(0.0216)		(0.0339)		(0.0176)
Mother is in poor health	0.585***	-0.193***	0.739***	-0.235**	0.570***	-0.133***
	(0.0551)	(0.0638)	(0.0932)	(0.119)	(0.0814)	(0.0468)
rho	0.321		0.114		0.342	
Likelihood	-46805006.8		-17035516.8		-28783875.1	
Observations	16788		7220		9568	
Average partial effect of work decision on care choice						
Average partial effect	-.0862***		-.0037		-.0897**	
95 % bootstrap s.e.	(.0316)		(.102)		(.0392)	

Note: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. The cluster standard errors in the country of a caregiver are in parentheses. I restrict the sample to individuals between 50 years old and statutory retirement age in twelve European countries: Austria, Belgium, the Czech Republic, Denmark, Germany, Greece, France, Italy, the Netherlands, Poland, Spain, Sweden. The observation period includes six interview waves that span from 2004 to 2017. The dependent variable varies across columns. The set of control variables in all regressions includes caregivers' characteristics: age, age squared, married dummy, four health dummies, limitations with daily activities, four education dummies, years of work experience by 2005, non-wage income, number of children, number of brothers, and number of sisters; mother's characteristics: four dummies for residential proximity to a caregiver, mother's age, and partner alive dummy; and interview year dummies, country-specific linear trend, and country dummies. In the bivariate model, the average partial effect of work decision on care choice is equal to $1/n \sum_i [\Phi(\alpha_1 + \alpha_2 mpoor_{it} + \beta x_{it} + \theta_t + \omega_{g(i)t}) - \Phi(\alpha_2 mpoor_{it} + \beta x_{it} + \theta_t + \omega_{g(i)t})]$.

Columns 3 and 4 Table 2.7 show that after correcting for endogeneity, the impact of labor status on men's daily care choice is not small and not statistically significant. However, for female daily caregivers, the average partial effect is statistically significant at a 5 percent level (see columns 5 and 6). If a woman participates in the labor market, her probability of providing daily care decreases almost by 9 p.p. This drop is quantitatively and statistically significant. Due to data shortage, I cannot study heterogeneity across countries or country groups because the number of observations shrinks.

2.5 Robustness Checks

Now, I show the robustness check of my analysis to the inclusion of formal care proxies.

One threat to the study is the omitted bias because of formal care usage. If the Great Recession leads to changes in the LTC provision, and I do not control for it, then the identification is challenging. As many scholars report, there is little data on formal care availability; accordingly, I use only macro characteristics relevant for respondents. Since my analysis covers twelve countries over six waves, it is important to have comparable information across time and states. I use three different proxies: LTC expenditures, the number of available nursing beds, and potential nursers' income for elders. The last measure is of particular importance because it is a proxy for formal care prices, which can be affected by the Great Recession. The choice to take care of the mother potentially can depend on the caregiver's opportunity cost and the cost of paid formal care. In Appendix 2.9.2, column 1 in Table B.5 repeats the main specification for daily care provision. Columns 2-4 report the results with further controls for formal care. In each case, the sample size gets smaller, but the main findings remain the same.[15]

2.6 Conclusion

This paper studies the decision to provide informal care among working-age individuals above 50 years old in twelve European countries. Specifically, it analyzes the impact of employment status on caregiving to the elderly mothers. I develop a simultaneous choice model of labor participation and informal care decision. The main identification strategy exploits the exogenous shift in working status due to the Great Recession to correct for endogeneity. After controlling for an extensive set of observed variables, such as work experience, the health of caregiver and care receiver, residential proximity, and proxies for prices and availability of formal care, I find that staying outside of the labor market increases by nine p.p. in the probability of providing informal care to a mother. Female caregivers mainly drive this result.

This paper adds to the literature about informal care provision and employment status. To my knowledge, it is the first one to estimate the causal impact of labor participation on informal care in Europe. Apart from closing the literature gap, there is a potentially important policy

[15] Due to the convergence problem, I cannot include all proxies in one specification.

implication of the study. In particular, with more elderly people, authorities may encourage individuals to stay in the labor market. However, these policies could have a negative impact on care to elders. More individuals in the labor market may result to be beneficial for the society. Yet, policymakers should be aware about the problem of unmet needs in elders' care when formal care provision is scarce.

Appendix B

2.7 Appendix: SHARE Data and External Statistics

2.7.1 Statutory Retirement Age

Table B.1: Statutory Retirement Age Across Genders and Countries in 2014

	(1) Men	(2) Women
Northern countries:		
Denmark	65	65
The Netherlands	65	65
Sweden	67	67
Central countries:		
Austria	65	65
Belgium	65	65
Germany	65	65
France	66	66
Southern countries:		
Greece	65	65
Italy	66	65
Spain	65	65
Eastern countries:		
Czech Republic	63	60
Poland	65	60

Note: OECD report 2014.

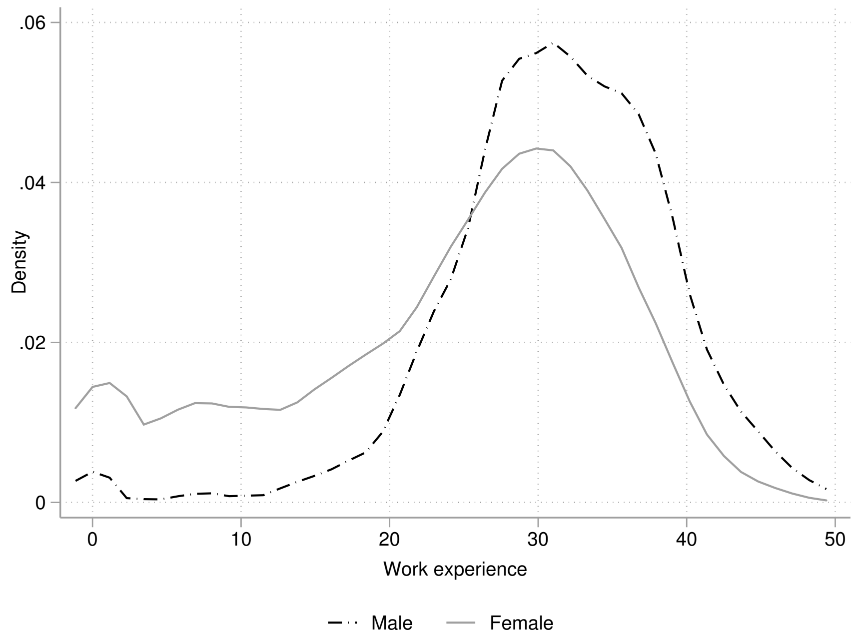
2.7.2 Frequency of Care Across Country Groups

Table B.2: Care Intensity Across Country Groups

	Frequency of care (%)				
	(1) Daily	(2) Weekly	(3) Monthly	(4) Less often	(5) Observations
Country groups:					
Northern	92	556	544	409	1601
	(5.75)	(34.73)	(33.98)	(25.55)	(100)
Central	442	922	472	333	2168
	(20.38)	(42.51)	(21.76)	(15.35)	(100)
Southern	340	305	103	89	837
	(40.62)	(36.44)	(12.31)	(10.63)	(100)
Eastern	121	168	95	55	439
	(27.56)	(38.27)	(21.64)	(12.53)	(100)
Total	995	1951	1214	886	5046

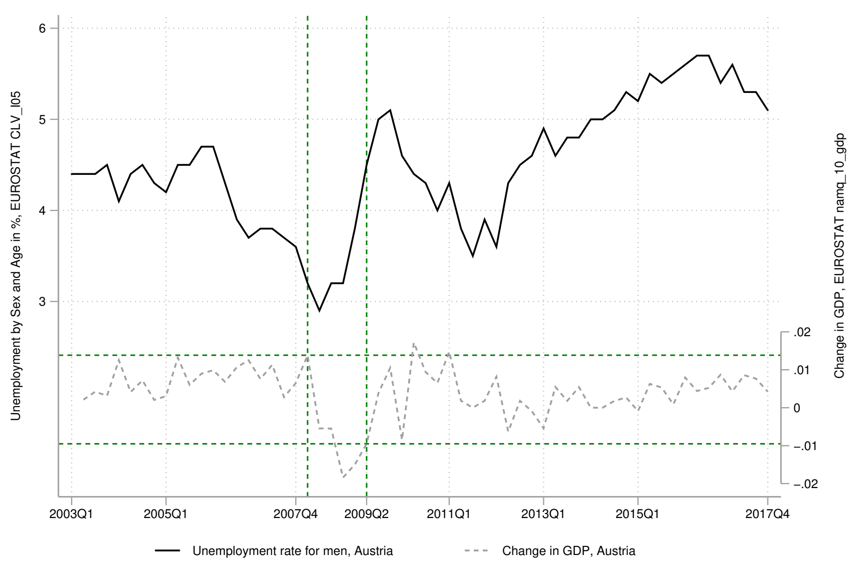
2.7.3 Work Experience Before 2005 Across Genders

Fig. B.1. Years of Work Experience by 2005



2.7.4 Measure of the Great Recession

Fig. B.2. Impact of the Great Recession on Men in Austria



2.8 Appendix: Results

2.8.1 Change in Mother's Residential Proximity

Table B.3: Change in the Transition of the Mother's Residential Proximity before and after the Crisis

	Residential proximity to mother					
	Co-reside (%)	Less than 25 kms (%)	Less than 500 kms (%)	More than 500 kms (%)	Total (%)	N
			Before the crisis			
Lagged value:						
Co-resident	87.80	5.22	6.69	0.28	100.00	116
Less than 25 kms	3.63	92.20	4.17	0.00	100.00	1,213
Less than 500 kms	0.77	10.93	86.96	1.35	100.00	616
More than 500 kms	0.49	10.46	9.13	79.93	100.00	123
Total	9.78	57.63	27.13	5.46	100.00	2,068
			After the crisis			
Lagged value:						
Co-resident	80.15	15.76	3.87	0.23	100.00	532
Less than 25 kms	3.64	89.10	6.74	0.53	100.00	4006
Less than 500 kms	1.38	13.08	82.30	3.24	100.00	1914
More than 500 kms	0.79	12.83	15.05	71.33	100.00	466
Total	11.07	54.71	26.15	8.07	100.00	6918

2.8.2 Full Specification

Table B.4: Maximum Likelihood Estimation of Work and Care choices

	Informal care				Daily informal care			
	Ignoring endogeneity		Bivariate probit		Ignoring endogeneity		Bivariate probit	
	(1) Care	(2) Work	(3) Care	(4) Work	(5) Care	(6) Work	(7) Care	(8) Work
Bivariate probit coefficients								
Work	0.00911		-0.298**		-0.180***		-0.716***	
	(0.0864)		(0.121)		(0.0481)		(0.0641)	
Crisis		-0.0613***		-0.0514**		-0.0599***		-0.0521**
		(0.0218)		(0.0219)		(0.0218)		(0.0216)
Mother is in poor health	0.352***	-0.188***	0.332***	-0.188***	0.628***	-0.187***	0.585***	-0.194***
	(0.0669)	(0.0607)	(0.0548)	(0.0619)	(0.0710)	(0.0606)	(0.0550)	(0.0638)
Born in the same country	0.105	-0.164***	0.0906	-0.164***	-0.00119	-0.165***	-0.0202	-0.168***
	(0.135)	(0.0299)	(0.129)	(0.0293)	(0.167)	(0.0300)	(0.158)	(0.0291)
Property owner	0.0651	0.0628	0.0705	0.0649	-0.0147	0.0640	-0.00488	0.0639
	(0.0563)	(0.120)	(0.0469)	(0.121)	(0.0937)	(0.119)	(0.0869)	(0.120)
Female	0.442***	-0.117	0.430***	-0.112	0.448***	-0.118	0.424***	-0.114
	(0.0616)	(0.0745)	(0.0552)	(0.0735)	(0.0450)	(0.0755)	(0.0349)	(0.0751)
Age	-0.0617	1.182***	0.0432	1.178***	-0.278**	1.179***	-0.101	1.176***
	(0.0594)	(0.242)	(0.109)	(0.238)	(0.119)	(0.242)	(0.147)	(0.243)
Age squared	0.000518	-0.0118***	-0.000537	-0.0118***	0.00252**	-0.0118***	0.000727	-0.0118***
	(0.000465)	(0.00204)	(0.000969)	(0.00201)	(0.00108)	(0.00205)	(0.00131)	(0.00205)
Married	-0.0620	0.0652	-0.0565	0.0682	0.0391	0.0643	0.0493	0.0673
	(0.0608)	(0.162)	(0.0531)	(0.161)	(0.0691)	(0.163)	(0.0538)	(0.161)
Number of children	-0.0613***	0.0478**	-0.0564***	0.0488**	-0.0606***	0.0478**	-0.0513***	0.0483**
	(0.0122)	(0.0202)	(0.0145)	(0.0214)	(0.0122)	(0.0203)	(0.0127)	(0.0208)
Number of mobility limitations	-0.0257	-0.0850***	-0.0325	-0.0859***	0.0291	-0.0851***	0.0168	-0.0875***
	(0.0211)	(0.0227)	(0.0208)	(0.0227)	(0.0515)	(0.0226)	(0.0491)	(0.0228)
Self-rated health								
Good	-0.148***	-0.137***	-0.158***	-0.135***	0.0441	-0.139***	0.0223	-0.132***
	(0.0412)	(0.0374)	(0.0386)	(0.0380)	(0.0750)	(0.0369)	(0.0754)	(0.0364)
Fair	-0.125***	-0.342***	-0.156***	-0.342***	-0.0268	-0.342***	-0.0819**	-0.340***
	(0.0301)	(0.106)	(0.0351)	(0.107)	(0.0246)	(0.106)	(0.0339)	(0.105)
Poor	-0.405***	-0.950***	-0.486***	-0.946***	-0.285	-0.951***	-0.421	-0.939***
	(0.107)	(0.0533)	(0.111)	(0.0493)	(0.314)	(0.0530)	(0.295)	(0.0520)
Number of brothers	-0.0644***	0.0104	-0.0635***	0.0110	-0.0439	0.00989	-0.0418	0.00918
	(0.00738)	(0.0342)	(0.00794)	(0.0338)	(0.0355)	(0.0347)	(0.0297)	(0.0347)
Number of sisters	-0.0445***	-0.0104	-0.0450***	-0.0107	-0.0608**	-0.00982	-0.0595**	-0.0103
	(0.00991)	(0.00865)	(0.0101)	(0.00895)	(0.0245)	(0.00833)	(0.0242)	(0.00842)
Education								
Lower secondary education	-0.0402	0.297***	-0.0148	0.299***	-0.0312	0.297***	0.0162	0.299***
	(0.155)	(0.0573)	(0.151)	(0.0574)	(0.0989)	(0.0574)	(0.0950)	(0.0553)
Upper secondary education	0.00844	0.449***	0.0490	0.449***	0.0237	0.448***	0.0959	0.449***
	(0.125)	(0.100)	(0.121)	(0.101)	(0.0685)	(0.103)	(0.0664)	(0.104)
Post-secondary education or above	0.232*	0.751***	0.298**	0.750***	0.123*	0.748***	0.241***	0.747***
	(0.131)	(0.122)	(0.126)	(0.123)	(0.0684)	(0.126)	(0.0499)	(0.129)
Experience by 2005	0.00474**	0.0314***	0.00748***	0.0314***	0.00179	0.0312***	0.00644***	0.0314***
	(0.00206)	(0.00345)	(0.00170)	(0.00340)	(0.00288)	(0.00352)	(0.00243)	(0.00343)
Non-wage net income, K euro PPP	0.000469	-0.00378**	0.000132	-0.00381**	0.000805	-0.00377**	0.000223	-0.00377**
	(0.000821)	(0.00168)	(0.000748)	(0.00168)	(0.000634)	(0.00169)	(0.000642)	(0.00168)
Father is alive	-0.239***	0.0653	-0.231***	0.0651	-0.0896	0.0619	-0.0765	0.0618
	(0.0833)	(0.0743)	(0.0795)	(0.0747)	(0.0596)	(0.0767)	(0.0564)	(0.0792)
Mother's age	0.0254***	0.00333	0.0255***	0.00369	0.0277***	0.00298	0.0274***	0.00343
	(0.00693)	(0.00247)	(0.00697)	(0.00238)	(0.0103)	(0.00266)	(0.0103)	(0.00262)
Residential proximity								
Less than 25 kms	-0.322***	0.0434	-0.315***	0.0429	-0.363***	0.0416	-0.347***	0.0447
	(0.0697)	(0.0959)	(0.0646)	(0.0961)	(0.101)	(0.0947)	(0.107)	(0.0942)
Less than 500 kms	-0.808***	0.0927	-0.793***	0.0897	-1.443***	0.0921	-1.394***	0.0961
	(0.0865)	(0.0578)	(0.0881)	(0.0575)	(0.234)	(0.0574)	(0.231)	(0.0605)
More than 500 kms	-1.234***	-0.171***	-1.239***	-0.168***	-2.483***	-0.171***	-2.465***	-0.165***
	(0.0826)	(0.0573)	(0.0841)	(0.0577)	(0.340)	(0.0576)	(0.336)	(0.0562)
Constant	-0.575	-30.71***	-3.111	-30.63***	4.471	-30.61***	0.256	-30.57***
	(1.770)	(7.140)	(3.090)	(7.021)	(3.235)	(7.145)	(4.095)	(7.168)
rho			0.183				0.321	
Likelihood	-35384757.9	-33871651.4	-69245879.8		-12991181.3	-33837772.3	-46805006.8	
Observations	16811	16811	16811		16788	16788	16788	
Average partial effect of work decision on care choice								
Average partial effect			-.0934*				-.0862***	
Bootstrap s.e			(.0525)				(.0316)	

Note: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. The cluster standard errors in the country of a caregiver are in parentheses. I restrict the sample to individuals between 50 years old and statutory retirement age in twelve European countries: Austria, Belgium, the Czech Republic, Denmark, Germany, Greece, France, Italy, the Netherlands, Poland, Spain, Sweden. The observation period includes six interview waves that span from 2004 to 2017. The dependent variable varies across columns. In columns (1) and (5) it is decision to care outside the household and to provide daily care respectively, in column (2) and (6) it is decision to participate, and in columns (3)-(4) and (7)-(8) I estimate a bivariate model of both decisions to work and to care. The set of control variables in all regressions includes interview year dummies, country specific linear trend, and country dummies.

2.9 Appendix: Robustness Checks

2.9.1 Formal Care Controls

Nursing beds per inhabitant: https://appsso.eurostat.ec.europa.eu/nui/show.do?query=BOOKMARK_DS-565686_QID_143A72EF_UID_-3F171EB0&layout=TIME,C,X,0;GEO,L,Y,0;UNIT,L,Z,0;ICHA11_HC,L,Z,1;INDICATORS,C,Z,2;&zSelection=DS-565686INDICATORS,OBS_FLAG;DS-565686ICHA11_HC,HC3;DS-565686UNIT,PPS_HAB;&rankName1=ICHA11-HC_1_2_-1_2&rankName2=UNIT_1_2_-1_2&rankName3=INDICATORS_1_2_-1_2&rankName4=TIME_1_0_0_0&rankName5=GEO_1_2_0_1&ppcRK=FIRST&ppcSO=ASC&sortC=ASC_-1_FIRST&rStp=&cStp=&rDCh=&cDCh=&rDM=true&cDM=true&footnes=false&empty=false&wai=false&time_mode=NONE&time_most_recent=false&lang=EN&cfo=%23%23%23,%23%23%23.%23%23%23

Unemployment rate: https://appsso.eurostat.ec.europa.eu/nui/show.do?query=BOOKMARK_DS-055628_QID_555D27C2_UID_-3F171EB0&layout=TIME,C,X,0;GEO,L,Y,0;S_ADJ,L,Z,0;AGE,L,Z,1;UNIT,L,Z,2;SEX,L,Z,3;INDICATORS,C,Z,4;&zSelection=DS-055628INDICATORS,OBS_FLAG;DS-055628S_ADJ,SA;DS-055628AGE,Y25-74;DS-055628SEX,F;DS-055628UNIT,PC_ACT;&rankName1=UNIT_1_2_-1_2&rankName2=AGE_1_2_-1_2&rankName3=INDICATORS_1_2_-1_2&rankName4=SEX_1_2_-1_2&rankName5=S-ADJ_1_2_-1_2&rankName6=TIME_1_0_0_0&rankName7=GEO_1_2_0_1&sortC=ASC_-1_FIRST&rStp=&cStp=&rDCh=&cDCh=&rDM=true&cDM=true&footnes=false&empty=false&wai=false&time_mode=NONE&time_most_recent=false&lang=EN&cfo=%23%23%23%2C%23%23%23.%23%23%23

Quarterly GDP: https://appsso.eurostat.ec.europa.eu/nui/show.do?query=BOOKMARK_DS-406779_QID_-4FDBF075_UID_-3F171EB0&layout=TIME,C,X,0;GEO,L,Y,0;UNIT,L,Z,0;S_ADJ,L,Z,1;NA_ITEM,L,Z,2;INDICATORS,C,Z,3;&zSelection=DS-406779UNIT,CLV_I05;DS-406779INDI OBS_FLAG;DS-406779S_ADJ,SCA;DS-406779NA_ITEM,B1GQ;&rankName1=UNIT_1_2_-1_2&rankName2=INDICATORS_1_2_-1_2&rankName3=NA-ITEM_1_2_-1_2&rankName4=S-ADJ_1_2_-1_2&rankName5=TIME_1_0_0_0&rankName6=GEO_1_2_0_1&sortC=ASC_-1_FIRST&rStp=&cStp=&rDCh=&cDCh=&rDM=true&cDM=true&footnes=false&empty=false&wai=false&time_mode=NONE&time_most_recent=false&lang=EN&cfo=%23%23%23%2C%23%23%23.%23%23%23

2.9.2 Controls for Formal Care

Table B.5: Maximum Likelihood Estimation of Work and Care Choices with Additional Controls

	Main	Additional controls for formal care		
	(1)	(2)	(3)	(4)
Dependent variable: Daily care				
Work	-0.716***	-0.592***	-0.714***	-1.287***
	(0.064)	(0.079)	(0.082)	(0.489)
Mother is in poor health	0.585***	0.564***	0.597***	0.484***
	(0.055)	(0.090)	(0.060)	(0.076)
LTC, PPP euros per inhabitant		-0.000		
		(0.001)		
Available beds in nursing houses			-0.002***	
			(0.000)	
Earnings ISCO-08, $ 2011				0.015***
				(0.001)
Dependent variable: Work				
Crisis	-0.052**	-0.033*	-0.053*	-0.189**
	(0.022)	(0.019)	(0.028)	(0.083)
Mother is in poor health	-0.194***	-0.171**	-0.188***	-0.433***
	(0.064)	(0.081)	(0.067)	(0.164)
LTC, PPP euros per inhabitant		-0.000		
		(0.001)		
Available beds in nursing houses			-0.000	
			(0.000)	
Earnings ISCO-08, $ 2011				0.002
				(0.002)
rho	0.321	0.250	0.329	0.631
Likelihood	-46805006.8	-34793000.2	-42834671.4	-7009708.73
Observations	16788	11765	12874	3703

Note: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. The cluster standard errors in the country of a caregiver are in parentheses. I restrict the sample to individuals between 50 years old and statutory retirement age in twelve European countries: Austria, Belgium, the Czech Republic, Denmark, Germany, Greece, France, Italy, the Netherlands, Poland, Spain, Sweden. The observation period includes six interview waves that span from 2004 to 2017. I estimate a joint model of both decisions to work and to care. Column 2 controls for LTC expenditure per inhabitation in a country, column 3 includes the number of available nursing beds; and column 4 controls for ILO income in a country. The set of control variables in all regressions includes caregivers' characteristics: age, age squared, married dummy, four health dummies, limitations with daily activities, four education dummies, years of work experience by 2005, non-wage income, number of children, number of brothers, and number of sisters; mother's characteristics: four dummies for residential proximity to a caregiver, mother's age, and partner alive dummy; and interview year dummies, country-specific linear trend, and country dummies.

2.9.3 Sample of Family Respondents and All Respondents

Table B.6: Maximum Likelihood Estimation of Work and Care Choices

	Informal care			Daily informal care		
	Original sample (1)	Only family respondents (2)	All respondents (3)	Original sample (4)	Only family respondents (5)	All respondents (6)
Dependent variable: Care						
Work	-0.298**	-0.167	-0.396***			
	(0.121)	(0.159)	(0.113)			
Mother is in poor health	0.332***	0.350***	0.326***			
	(0.0548)	(0.0919)	(0.0513)			
Dependent variable: Daily care						
Work				-0.716***	-0.519***	-0.716***
				(0.0641)	(0.165)	(0.0637)
Mother is in poor health				0.585***	0.628***	0.585***
				(0.0550)	(0.0493)	(0.0549)
Dependent variable: Work						
Crisis	-0.0514**	-0.0693**	-0.0223	-0.0521**	-0.0690**	-0.0483**
	(0.0219)	(0.0284)	(0.0230)	(0.0216)	(0.0276)	(0.0217)
Mother is in poor health	-0.188***	-0.137	-0.156**	-0.194***	-0.140	-0.192***
	(0.0619)	(0.0845)	(0.0620)	(0.0638)	(0.0860)	(0.0633)
rho	0.183	0.127	0.224	0.321	0.231	0.320
Likelihood	-69245879.8	-52171988.5	-76094535.9	-46805006.8	-35210071.6	-46822876.3
Observations	16811	12976	19693	16788	12962	16802

Note: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. The cluster standard errors in the country of a caregiver are in parentheses. I restrict the sample to individuals between 50 years old and statutory retirement age in twelve European countries: Austria, Belgium, the Czech Republic, Denmark, Germany, Greece, France, Italy, the Netherlands, Poland, Spain, Sweden. The observation period includes six interview waves that span from 2004 to 2017. The set of control variables in all regressions includes caregivers' characteristics: age, age squared, married dummy, four health dummies, limitations with daily activities, four education dummies, years of work experience by 2005, non-wage income, number of children, number of brothers, and number of sisters; mother's characteristics: four dummies for residential proximity to a caregiver, mother's age, and partner alive dummy; and interview year dummies, country specific linear trend, and country dummies.

2.10 Appendix: Sample Selection in the Panel Analysis

Table B.7: Descriptive Statistics of Respondents Across Cross-section and Dynamic Specifications

	Participate only in a cross-section (1)	SD	Participate in a panel (2)	SD	Difference (1) - (2)	P-value
Socio-demographic characteristics						
Age	55.56	4.51	54.24	3.53	1.32	0.00
Female	0.56	0.50	0.57	0.50	-0.00	0.72
Married	0.81	0.39	0.76	0.43	0.05	0.00
Number of children	2.09	1.23	2.02	1.12	0.08	0.00
Number of brothers	1.13	1.19	1.13	1.17	-0.00	0.96
Number of sisters	1.10	1.19	1.08	1.15	0.02	0.56
Education						
Primary education or less	0.15	0.35	0.10	0.30	0.04	0.00
Lower secondary education	0.17	0.38	0.17	0.37	0.01	0.55
Upper secondary education	0.38	0.49	0.38	0.49	-0.00	0.83
Post-secondary education or above	0.30	0.46	0.35	0.48	-0.05	0.00
Self-rated health						
Very Good	0.40	0.49	0.44	0.50	-0.04	0.00
Good	0.38	0.48	0.36	0.48	0.02	0.14
Fair	0.17	0.38	0.16	0.37	0.01	0.12
Poor	0.05	0.22	0.04	0.21	0.01	0.11
Number of mobility limitations	0.86	1.62	0.80	1.59	0.06	0.12
Born in the same country	0.94	0.24	0.93	0.25	0.01	0.31
Owner	0.78	0.41	0.75	0.43	0.03	0.00
Non-wage net income, K euro PPP	26.61	42.73	27.44	40.22	-0.83	0.38
Participation characteristics						
Work	0.58	0.49	0.67	0.47	-0.09	0.00
Experience by 2005	26.72	11.61	27.18	10.18	-0.46	0.07
Care to a mother						
Mother's characteristics						
Father is alive	0.29	0.46	0.34	0.47	-0.05	0.00
Mother's age	81.97	6.08	79.80	5.36	2.17	0.00
Mother's self-rated health						
Poor health	0.19	0.40	0.12	0.32	0.08	0.00
Residential proximity to mother						
Co-resident	0.08	0.28	0.07	0.26	0.01	0.10
Less than 25 kms	0.59	0.49	0.58	0.49	0.02	0.10
Less than 500 kms	0.25	0.44	0.28	0.45	-0.03	0.01
More than 500 kms	0.07	0.25	0.07	0.25	-0.00	1.00
Observations	3810		3867			

Note: All the characteristics are reported the first time an individual participates in the survey. Column 1 corresponds with individuals who contribute to a cross-section analysis and have a mother alive. Column 2 corresponds with individuals who can contribute to the dynamic specification because their mothers are alive at least in two waves. The last column reports the P-value of the null hypothesis about the equality of means.

Table B.8: Selection into the Panel Analysis

	(1) To Be Present in Panel Analysis
Probit coefficients	
Crisis	-0.0612
	(0.0608)
Mother is in poor health	-0.325***
	(0.0835)
Born in the same country	0.0326
	(0.118)
Property owner	0.0253
	(0.0484)
Female	-0.0125
	(0.0508)
Age	1.056***
	(0.0942)
Age squared	-0.00991***
	(0.000845)
Married	-0.118**
	(0.0591)
Number of children	-0.0338
	(0.0231)
Number of mobility limitations	-0.0132
	(0.0154)
Self-rated health	
Good	-0.00385
	(0.0510)
Fair	-0.0253
	(0.0618)
Poor	-0.120
	(0.153)
Number of brothers	-0.0337**
	(0.0160)
Number of sisters	0.0000814
	(0.0306)
Education	
Lower secondary education	0.135*
	(0.0791)
Upper secondary education	0.122**
	(0.0528)
Post-secondary education or above	0.257***
	(0.0790)
Experience by 2005	-0.00218
	(0.00223)
Non-wage net income, K euro PPP	-0.000652
	(0.000521)
Father is alive	-0.0249
	(0.0515)
Mother's age	-0.0414***
	(0.00802)
Residential proximity	
Less than 25 kms	-0.0192
	(0.0600)
Less than 500 kms	-0.0165
	(0.0780)
More than 500 kms	-0.0692
	(0.134)
Constant	-24.24***
	(2.921)
Likelihood	-13052626.9
Observations	7662

Note: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. The cluster standard errors in the country of a caregiver are in parentheses. I restrict the sample to individuals between 50 years old and statutory retirement age in twelve European countries: Austria, Belgium, the Czech Republic, Denmark, Germany, Greece, France, Italy, the Netherlands, Poland, Spain, Sweden. The observation period includes six interview waves that span from 2004 to 2017.

Chapter 3

Online Discrimination and (Self) Regulation: Evaluating the Airbnb's Nondiscrimination Policy (with Michelangelo Rossi)

3.1 Introduction

During the last two decades, the volume of online transactions has enormously risen across different sectors and industries. Platforms such as Airbnb and Uber changed the traditional ways of doing business in the hospitality and transportation markets, respectively. Thanks to these websites, a great number of non-professional sellers entered those markets. However, some characteristics of the transactions that now occur on these digital marketplaces are unaltered relative to traditional off-line settings; unfortunately, one of them regards the discrimination of minorities by sellers and buyers.

In the most recent years, evidence shows the difficulties by non-white users to fully integrate into digital platforms. Regarding Airbnb, discrimination is present on both sides of the market (guests and hosts): Edelman, Luca, and Svirsky (2017) show that booking requests sent by users with African American sounding names are less likely to be accepted relative to users with white-sounding names. On the other side, Asian and Hispanic Airbnb's hosts charge lower prices with respect to white hosts with similar properties (Kakar, Voelz, Wu, and Franco, 2018).

Although the issues related to differences in treatment on the grounds of race are evidential and have been fairly documented, the present laws (such as the Title II of the Civil Rights Act of 1964) are not fully able to protect digital users because of the nature of digital transactions. The so-called sharing economy is constituted by a myriad of private small, often non-professional, sellers. Accordingly, as it is pointed out by Todisco (2015) and Leong and Belzer (2017), Airbnb's hosts may refuse to rent on the grounds of race and respect the Civil Right Act once they live in the rented building with less than five rooms to rent. Not the entire universe of Airbnb listings satisfies the requirement, but this exception captures the difficulty to frame digital platforms into the classical perimeters of traditional laws.[1]

To reduce discrimination and partially remedy the absence of clear laws, platforms often implement self-regulation to establish users' conduct rules. Among others, Airbnb launched a Nondiscrimination Policy at the end of 2016, prohibiting any discriminatory behavior by hosts and setting up several objectives to be reached in the next years in terms of inclusiveness.

This paper studies this policy in four US cities: New Orleans, New York, Portland, and San Francisco.[2] In the two years after the policy, we document an increase in the share of hosts activating a setting called "instant book" that automatically accepts any guests' requests. Still, the number of non-white guests on the platform only slightly increase.

With a difference-in-differences and an event study approach, we estimate variations in the number of non-white guests before and after the policy comparing hosts who can choose to reject a guest request (and potentially discriminate) with hosts who already activated the "instant book" setting before the policy.

We find that the number of hosts who decide to activate the instant booking option and accept all booking requests gradually increased after the policy. Further, results show hosts who can choose to reject a guest request do not rent more to non-white guests after the policy. Accordingly, non-instant bookable hosts (the majority on the platform) did not adjust their behavior.

We conclude that Airbnb's Nondiscrimination Policy may be considered as the first step toward inclusiveness. This self-regulation increases renting opportunities for non-white guests thanks to expanding the "instant book" setting among Airbnb's hosts. Still, further measures have to be taken to encourage hosts who do not want or cannot activate this setting.

The paper proceeds as follows. Section 3.2 provides some background context regarding

[1]A further relevant point regards the platform responsibility of illegal, discriminatory behaviors (see Todisco (2015) and Leong and Belzer (2017).

[2]These are the only US cities for which data is available before and after the policy (2016).

the Nondiscrimination Policy implemented by Airbnb in 2016. We describe the dataset in Section 3.3 and we discuss the identification strategy in Section 3.4. Section 3.5 provides the main empirical findings, and we extend these results in 3.5.2 and 3.5.3. Additional robustness checks are in Sections 3.6. Section 3.7 concludes. Additional tables and figures are in the Appendix.

3.2 The Airbnb's Nondiscrimination Policy

Airbnb is one of the leading digital platforms operating in the hospitality market since 2008. It connects hosts and guests in several countries around the world and, since 2014, many journalists and scholars reported evidence of discrimination against minorities. In particular, it has proven for non-white guests to be more challenging to book a place on Airbnb relative to white guests. Among other researchers, Edelman and Luca (2014) suggested mitigating discriminatory behavior among users by taking an example from the hotel industry where the Title II of the Civil Rights Act forbids lodging providers to reject booking requests based on racial preferences.

In June 2016, as a response to the evidence of unfair treatment, Airbnb committed to updating its policies against discrimination, consulting the civil rights activist Laura Murphy. Later, on September 8, 2016, all users (hosts and guests) received an email in which Airbnb admitted being slow in developing a well-functioning anti-discriminatory policy and explained the coming changes. The new terms of the policy are the following:

- By November 1, 2016, all the registered hosts and guests have to agree with the "Airbnb Community Commitment" if they want to host and travel using the platform:

  I agree to treat everyone in the Airbnb community—regardless of their race, religion, national origin, ethnicity, disability, sex, gender identity, sexual orientation, or age—with respect, and without judgment or bias;[3],[4]

- If a host rejects guests based on racial preferences, using language that shows his or her motivations, Airbnb may suspend the host from the platform;

[3] Airbnb Community Commitment: https://www.airbnb.com/help/article/1523/general-questions-about-the-airbnb-community-commitment

[4] Official email by CEO Brian Chesky: https://blog.atairbnb.com/fighting-discrimination-and-creating-a-world-where-anyone-can-belong-anywhere/?af=14383374&c=GD_us_gen_pub

- Starting from the first half of 2017, Airbnb will develop a feature that automatically blocks the calendar after a host rejects a guest claiming that the dwelling is not available. This measure wants to hold hosts accountable, and it prevents them from making available their spaces depending on the guests' race;

- Finally, Airbnb set up a target of making one million listings bookable via an instant book by January 2017. Hosts can choose the instant book option to allow guests to book their listings without prior host approval.[5]

The policy got immediate media coverage: on September 8, 2016, The New York Times[6] and The Wall Street Journal[7] published two articles about the policy and argued its potential impact. Users got emails and started to discuss how to interpret the new rules and their future applications on the platform.[8] In the email, Airbnb also encourages guests not to be silent if they feel any discrimination and guarantee to find alternative accommodations.

The policy directly impacts the interactions between hosts and guests via the booking process on the platform. Here, we describe this process with a scheme of the decisions taken by hosts and guests; then, we illustrate how the policy could affect users' behavior.

First, to travel with Airbnb, a guest (she) chooses which listing to send a booking request. At this stage, she has two different options: to make a request to a listing with the instant book option on or off. An instant book is a free option that each host (he) can activate or deactivate at any time. If a listing has an instant book on, then guests' requests are automatically accepted.[9] With instant book off, then the host can decide to accept or reject a request. Finally, conditional on the host accepting the request (or having instant booking on), the guest rents the host's dwelling, and the two parties can leave feedback about the transaction.

Fig. 3.1 illustrates this procedure with a graph.

Discriminatory behavior can arise at any stage of this process. Guests may prefer white hosts over non-white hosts leading the latter to charge a lower price and attract more guests. On the other side, hosts may refuse requests by non-white guests or prefer white guests over

[5]For all the details of the policy: https://www.airbnb.com/help/article/1405/airbnb-s-nondiscrimination-policy--our-commitment-to-inclusion-and-respect

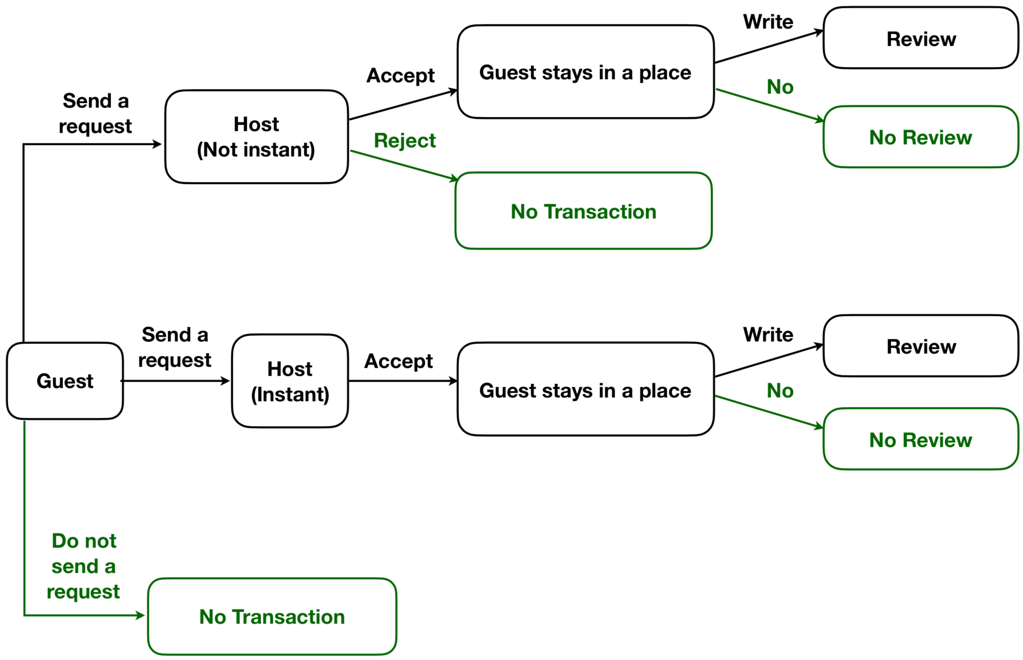
[6]https://www.nytimes.com/2016/09/09/technology/airbnb-anti-discrimination-rules.html

[7]https://www.wsj.com/articles/airbnb-promotes-diversity-to-prevent-booking-discrimination-by-hosts-14

[8]Blog: https://community.withairbnb.com/t5/Help/Age-Discrimination-and-new-Air-BNB-Policy/td-p/191482

[9]All hosts in Airbnb can require a guest to provide a national ID, phone number or others. Hence, to send a request, guests need to fulfill the conditions imposed by hosts. This procedure is not restrictive, and it applies independently of the instant book option.

Fig. 3.1. Booking Process on Airbnb



non-whites.

In this paper, we look at the guests' ethnicity by studying the names of guests who report a review. Specifically, we compare the ethnicity of guests who rent from hosts with the instant book option on and off. We take advantage of the fact that, if hosts choose an instant book, they cannot reject any booking request and manage their listings as a regular hotel. With this strategy, we can estimate whether non-white hosts become more inclusive after the policy. However, we cannot identify which stage of the booking process is affected: non-white guests may send more requests to white hosts; or, white hosts may accept more requests by non-white guests. Moreover, although less likely, the policy may affect the guests' decision to review. Here we assume that the propensity to leave a review is not affected by the policy and does not differ between white and non-white guests.

3.3 Data Description

We use data from *Inside Airbnb*[10] for four US cities with available data before and after the Nondiscrimination policy: New Orleans, New York, Portland, and San Francisco. The number

[10] http://insideairbnb.com/index.html

of available snapshots slightly varies across cities (40, on average). We choose these four cities since it is key for our identification strategy to have well-defined pre-periods, and these cities are the only ones whose snapshots date back to 2015. In this Section, when we report the descriptive statistics over time, we exploit all the available periods until March 2019. However, starting from Section 3.4, we analyze a shorter window to remove potential confounders: from April 1, 2016, to April 1, 2017. We merge the universe of listings present in each snapshot with corresponding reviews to identify the hosts' name, the reviewer's name, and the day of a written comment to construct the panel data. In total, during all the periods, there are 3,979,530 listing-reviewer observations and 225,762 unique listings.

The crucial part of our analysis regards the identification of the hosts' and guests' ethnicities. We use hosts' and guests' names and apply the NamePrism algorithm to identify the probability a name to belong to one of the following six ethnicities: White, Black, API (Asian and Pacific Islander), AIAN (American Indian and Alaska Native), 2PRACE (more than two race) and Hispanic.[11] In our main analysis, we divide users into white and non-white, so we pool together several ethnic groups to increase the number of observations in a minority group. In the extension of the analysis, we check separately the impact on Black, Asian, Pacific Islander, and Hispanic (directly from NamePrism) users. We consider only the first listed name when users include several names in their titles (less than 5 percent of all users include several names). We construct for each host and guest the probability of being non-white as one minus probability to be white (directly comes from NamePrism).

Although we focus on racial discrimination in the main text, in the Appendix 3.10.3 we look at the effect of the policy on gender discrimination. To identify female names we use the algorithm provided by Genderize.io[12]. Fig. C.7a and Fig. C.7b show the density of host's and guest's names respectively.

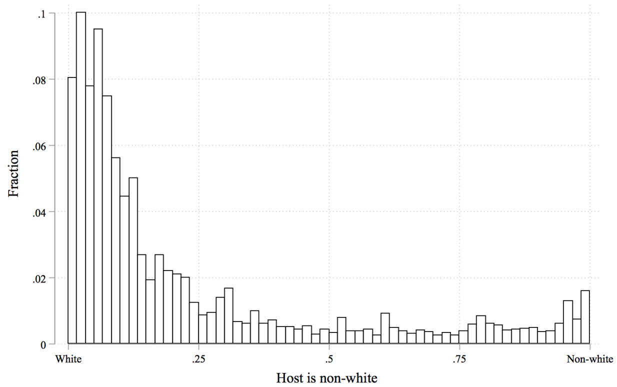
The density of host and guest probability to be non-white are in Fig. 3.2a and 3.2b, respectively. From these two Figures, we can observe that Airbnb's users are predominately white and the empirical densities for name's ethnicities are heavily left-skewed. We find a similar pattern looking at each city separately (see Fig. C.1).

The share of white and non-white Airbnb users varies over time, together with the total number of hosts and guests present on the platform. Fig. 3.3a and 3.3b show the evolution of

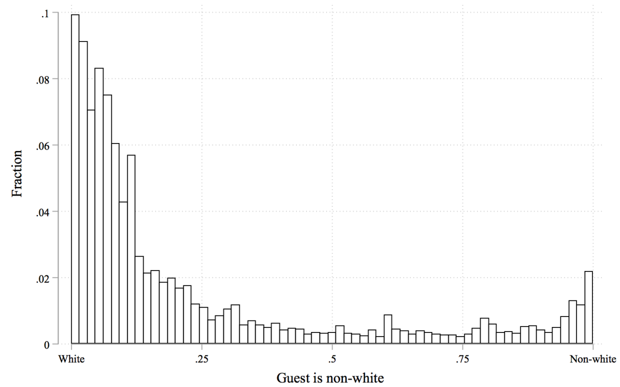
[11] We highly appreciate the access to NamePrism classifications provided by Professor Steven Skiena: `http://name-prism.com/about`. See Ye, Han, Hu, Coskun, Liu, Qin, and Skiena (2017) and Ye and Skiena (2019) for details.

[12] `https://store.genderize.io`.

Fig. 3.2. Ethnicity of Users on Airbnb



(a) Host



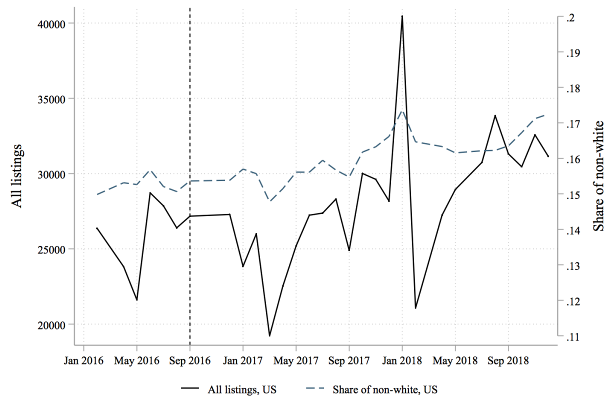
(b) Guest

the number of listings and the written comments; and the shares of non-white hosts and guests over time. Accordingly, we observe a seasonal pattern in the number of listings and comments with picks during and after the summer season. Conversely, the share of non-white hosts and guests are relatively stable over time. Both shares gradually increase after 2016 showing progressive improvement in the platform inclusiveness. These findings are stable across cities (see Fig. C.2 and C.3 for listings and guests respectively).

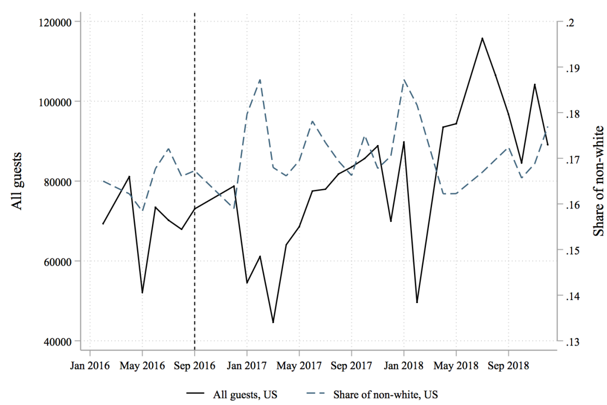
With a similar approach, we can check the share of Airbnb listings with instant booking on overtime reported in Fig. 3.3c.

As mentioned in the previous Section, Airbnb promoted the instant booking option among their hosts after 2016 in relation to its Nondiscrimination policy. In this line, the share of listings with an automatic acceptance of requests increases sharply over time, and the start of growth coincides with the policy moment. In March 2019, more than half of all listings active on Airbnb have the instant booking option on. This translates into partial improvement concerning the platform's inclusiveness since instant booking hosts cannot reject guests' requests. The explanation to such a dramatic increase in the share of instant bookable host could be due to significant media coverage of the Nondiscrimination Policy from the very first days after its announcement. Accordingly, hosts could perhaps decide to activate instant booking to avoid being under suspicion regardless of their discriminatory behavior. In the next Section, we will provide evidence in line with the greater inclusiveness of hosts with instant booking on compared to those with instant booking off.

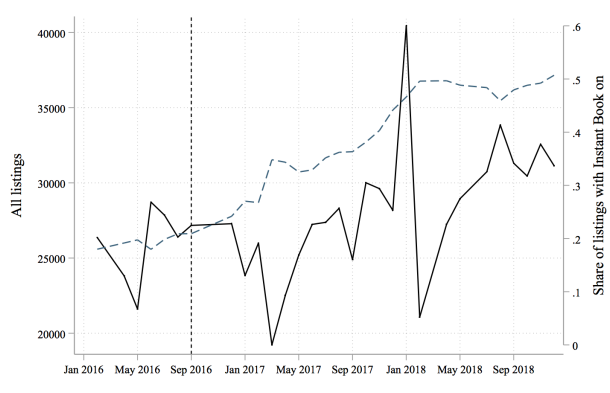
Fig. 3.3. Change in Users' Ethnicity and Instant Booking Option on Airbnb



(a) Number and Share of Non-White Hosts Over Time



(b) Number and Share of Non-White Guests Over Time



(c) Number and Share of Instant Booking Listings Over Time

Note: In these Figures, to simplify the representation, we assume that a user is white if the probability to be white is higher than 0.5. We consider only the first listed name when users include several names in their titles (less than 5 percent of all users include several names).

3.4 Identification Strategy

In this Section, we discuss two identification strategies to evaluate the policy's impact on those hosts who do not activate instant booking. First, we describe the difference-in-differences design. Next, and related to it, we implement an event study.

3.4.1 Difference-in-Differences Design

Our first identification strategy is *a difference-in-differences design* that compares the probability to have non-white guests before and after the Nondiscrimination policy across listings with and without an instant booking option. Since the hosts who have instant booking on cannot reject the booking requests, they cannot discriminate. Accordingly, in our analysis we use them as a control group. Only listings without instant booking could be affected by the change in the platform since, after the policy, hosts face higher costs to reject booking requests motivated by ethnic preferences. Thus, we consider them as a treated group. The main specification is:

$$\textit{Non-white guest}_{ijt} = \alpha_1 + \alpha_2 \textit{Not instant}_i + \alpha_3 \textit{After}_t + \delta \textit{Not instant}_i \times \textit{After}_t + \beta \boldsymbol{X}_{it} + \epsilon_{ijt}, \quad [3.4.1]$$

where $\textit{Non-white guest}_{ijt}$ is the probability to be non-white for a guest j who leaves a written comment about listing i at date t. $\textit{Not instant}_i$ is equal to 1 if a listing i has instant booking off in all the periods from the beginning of April 2016 to April 2017, it is equal to 0 if a listing i has an instant booking on during the same periods. $After_t$ is equal to 1 if a guest wrote a comment after October 8, 2016, but before April 1, 2017, it is equal to 0 if a comment was written from May 1, 2016, to October 8, 2016. We use several specifications to control for observable variables and possible unobserved characteristics.

The least restrictive specification does not include any other control except for the first three listed above. Also, in our second specification, we control for a week, month, city, neighborhood, month-neighborhood effects, $\boldsymbol{X}_{it}$. Further, we include observable characteristics: host's ethnicity, how long the listing is present on Airbnb, the total number of reviews displayed on the web page before a written comment was posted, total listing rating, the room type (entire places or shared apartments including shared rooms), rewarded superhost status[13],

[13] Airbnb rewards top-rated hosts with a superhost badge that it is re-rewarded every three months. To get and keep this status, hosts need to get a high rating, host at least ten stays, and have a low cancellation rate and a high response rate. For further details, you can see the Airbnb webpage: `https://www.airbnb.com/superhost`.

the price per night, availability of listings in the next 30 days. The most restrictive specification exploits the panel data structure and includes listing fixed effects. We use cluster standard errors at the listing level to allow for serial correlation within the listing. Our results remain unchanged if we cluster standard errors at the neighborhood level or the combination of time and neighborhood.

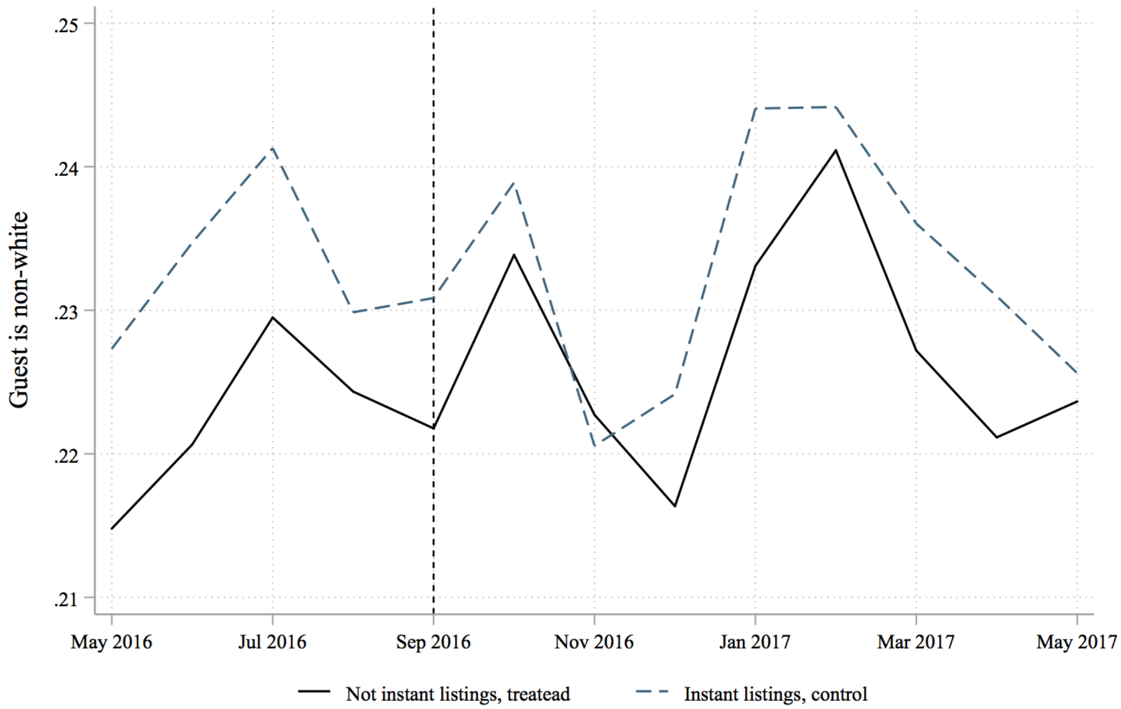
In the analysis, we consider October 8, 2016, as the starting moment from which policy has an effect. Users received the email about the policy on September 8, 2016; then, they had to comply with the proposed changes to be a part of the community. Accordingly, the first bookings under the new policy can appear from that day on. However, we see that 44 percent of listings that are present consecutive months on the platform is fully booked for the next 30 days. Consequently, we assume that the first reviewed transactions affected by the policy happen after October 8, 2016.

The identification design relies on the parallel trend assumption across treated and control groups. Fig. 3.4 confirms its validity: before the policy, the probability of having a transaction with non-white guests follows the same trend across instant booking or not instant booking listings. Hence, the difference in trends, δ, represents the policy's causal effect. Since October 8, 2016, it is more costly for all hosts to reject non-white guests; we expect hosts are more likely to have non-white guests in not instant booking listings, implying that δ is positive. Moreover, from Fig. 3.4 it is possible to see that non-white guests more often stay in listings with instant booking on relative to listings with instant booking off. This finding suggests that the spread of the instant booking option may improve the inclusiveness of the platform.

Now, we discuss the potential threats to our strategy and how we overcome them. First, since all the listings are eligible to use the instant booking option, hosts decide to turn it on. If the probability of activating instant booking changes simultaneously with the policy, it can fail our identification. In this case, δ will capture any innate difference among hosts' attitude to non-white guests. To check how the probability of changing the instant booking status, turning it on or off, depends on the policy date, we recall Fig. 3.3c. Indeed, it suggests that the share of instant booking listings increases after October 8, 2016. To abstract from this problem, we restrict to the listings that either had instant booking always on or off from April 2016 to April 2017. In the Subsection 3.5.3 we repeat the same identification strategy by using the same control group, but with two other treated groups: listings that changed their instant booking status and all listings that are present on the platform.

The other potential issue in our strategy could be due to changes in the probability of

Fig. 3.4. Parallel Trend Assumption



Note: The graph shows the evolution the average probability for a guest to be non-white across treated and control groups. We use all cities and restrict on listings entered before August 2016 and exited after October 2016.

entering or exit the platform. Accordingly, along with the article, we focus only on the listings that entered before August 2016 and exited after October 2016. We exclude any listings that exited the platform in October 2016 to get a more precise estimate in our main specification.

Further failure of our research design can be due to a change in the likelihood of writing a comment by guests who rent to hosts with instant booking on and off before and after September 8, 2016. To test it formally, it is necessary to use all the universe of transactions and see how the number of reviewed transactions evolves on the platform depending on the policy time. Looking at the number of written reviews in our sample, we see that the pattern between instant and non-instant booking changes before and after the policy (see Table 3.1). This could be potentially related to the treatment, but it could also be due to something unrelated to the treatment, like seasonality affecting both groups differently. We will address this point in more detail in the future version.

Now, we check how treated and control groups vary across observable characteristics: see Table 3.1. The same listings are present in the top and bottom panels. *Panel (I)* restricts all listings from April 1, 2016, to October 8, 2016. *Panel (II)* includes all listings from October 8, 2016, to March 31, 2017. All the statistics in *Panel (I)* correspond with the last observation before the policy and in *Panel (II)* - with March 31, 2017, the observation before each listing exits the platform. Information about always instant booking listings (control group) is in column (1) and about listings that are never instant booking (treated group) in column (2). We can see that these two groups are statistically different across almost all the variables, but superhost status. Listings with always instant booking are more likely to be managed by non-white hosts, have more reviews, and more available days. Accordingly, the average trip length may be shorter among their guests. Likewise, the prices are lower, and they offer less likely entire places. Despite this statistically significant difference between the two groups, the difference's relative size is not large.

Finally, the representativeness of our estimates can be of concern. Although we include many controls, the listings with many or few transactions have the same weight in our specification. It can be alarming because the distribution of the total number of guests per month is highly left-skewed (see Fig. C.5). Thus, we propose two types of weights to account for this heterogeneity across observations. The first weights are $\frac{\log(N)}{N}$, where N is the total number of guests per month. When N is very small or very large, we assign similar weights to observations, whereas if N is around 4, we give the highest weight to observations. Next, we construct the optimal weights following Solon, Haider, and Wooldridge (2015). Since these weights are linear in N, the corresponding estimates are very close to those for an unweighted sample.[14] Fig. C.6 illustrates both weighting schemes. In Appendix 3.10.1, we show that our results are robust for weighted and optimally weighted samples.

3.4.2 Event Study Set-Up

Our second specification is *an event study set-up* in which we estimate the effect of the policy relative to September 2016. To define correctly the observation window and end points, we closely follow Schmidheiny and Siegloch (2019). Since we observe several written reviews per month, we use a day-month-year format of a review date to define a time unit in a panel data. The main equation is:

[14]The P-value of a modified Breusch-Pagan test is 0.000, so we confirm heteroskedasticity in a sample.

Table 3.1: Descriptive Statistics of Control and Treated Groups

	Always Instant (1)	SD	Never Instant (2)	SD	(1) - (2)	[p-value]
Panel (I): Before September 8, 2016						
Host is non-white	0.25	0.29	0.21	0.26	0.03	0.00
Days on Airbnb	339.99	221.38	366.44	223.18	-26.45	0.00
Number of reviews	40.31	48.52	26.72	36.64	13.59	0.00
Total rating	91.50	7.69	93.86	6.61	-2.36	0.00
Entire apartments	0.50	0.50	0.60	0.49	-0.10	0.00
Superhost	0.19	0.39	0.18	0.38	0.01	0.12
Price per night	136.01	112.89	153.13	118.59	-17.13	0.00
Availability next 30 days	9.69	9.05	9.25	9.70	0.44	0.01
Number of listings	3469		22764			
Panel (II): After September 8, 2016						
Host is non-white	0.25	0.29	0.21	0.26	0.03	0.00
Days on Airbnb	855.07	414.50	872.88	400.46	-17.81	0.02
Number of reviews	50.06	56.16	25.32	36.47	24.74	0.00
Total rating	91.72	6.75	94.19	5.57	-2.48	0.00
Entire apartments	0.50	0.50	0.60	0.49	-0.10	0.00
Superhost	0.28	0.45	0.29	0.45	-0.01	0.43
Price per night	134.60	113.78	151.91	121.27	-17.30	0.00
Availability next 30 days	11.62	10.04	10.78	10.44	0.84	0.00
Number of listings	3469		22758			

Note: Both Panels restrict on listings entered before August, 2016 and exited after October, 2016. In addition, *Panel (I)* restricts on all listings from April 1, 2016 to October 8, 2016. *Panel (II)* includes all listings from October 8, 2016 to March 31, 2017. The same listing presents in both panels. The number of reviews corresponds with the number of written reviews that a listing received from April 1, 2016 to October 8, 2016, or from October 8, 2016 to March 31, 2017. All the statistics in *Panel (I)* correspond with the last observation before policy and in *Panel (II)* - with March 31, 2017 or the last snap before listing exits.

$$\textit{Non-white guest}_{ijt} = \beta_1 + \beta_2 \textit{Not instant}_i + \textit{Not instant}_i \sum_{\substack{k=-5 \\ k\neq -1}}^{6} \gamma_k b_{it}^k + \beta \boldsymbol{X}_{it} + \theta_i + \epsilon_{ijt}, \qquad [3.4.2]$$

where θ_i is a listing fixed effect; b_{it}^j is a dummy indicating the number of months relative to policy implementation. To build this variable, we change the format of a review date to month-year, $\tilde{t}$:

$$b_{it}^{k} = \begin{cases} \mathbb{1}[k \geq \tilde{t} - \textit{October 2016}] & \text{if } k = 6 \text{ and } k \leq 8 \\ \mathbb{1}[k = \tilde{t} - \textit{October 2016}] & \text{if } -5 < k < 5 \\ \mathbb{1}[k \leq \tilde{t} - \textit{October 2016}] & \text{if } k = -6 \text{ and } k \geq -8. \end{cases} \quad [3.4.3]$$

We bin b_{it}^{k} at the endpoints to increase the precision of the estimates far from the policy. In particular, we consider that the effect of February 2016, March 2016, April 2016 coincides and can be aggregated as the impact of April 2016 on a transaction with non-white guests. Similarly, April 2017, May 2017, and June 2017 are used as of April 2017. $\boldsymbol{X}_{it}$ includes the same controls as we use in a difference-in-differences design in the most restrictive specification. We cluster standard errors at the listing level.

The control group in our analysis includes listings that did not receive any treatment. That is why the identification comes from the difference among listings with and without instant booking. We normalize the coefficients to September 2016 to simplify the interpretation of the results. By plotting γ_k, we can directly validate the parallel trend assumption and see the effect of the policy. Besides, the two identification strategies should lead to similar findings and suffer similar failures.

3.5 Results

The implemented policy should change users' behavior and make the platform more inclusive for non-white guests. First, we show findings in the difference-in-differences and event study settings. Next, we repeat the same analysis, but for two other groups that may be considered as treated. Finally, we study the potential impact on prices.

3.5.1 Main Findings

In general, an event study set-up implies the inclusion of listing fixed effects. Still, it can be too restrictive in Airbnb's setting since listings can share similar patterns. Accordingly, we also show the results of the first identification strategy using both repeated cross-section and individual fixed effects.

Table 3.2 shows the difference-in-differences coefficient, δ, from Equation 3.4.1. Each panel differs in covariates, from the least restrictive to the most restrictive specification, which we describe in Section 3.4.1. *Panels (I) - (III)* exploit repeated cross-section whereas *Panels (IV)* allows for listing unobserved fixed effects. In the first Column, we pool together all cities; Columns 2 - 5 correspond with San Francisco, New York, Portland, and New Orleans. There is no statistically significant effect of the policy on the probability of having a non-white guest in any specification. The magnitude of the potential impact is also small. Accordingly, it implies that the policy does not make the platform more inclusive for non-white guests trying to book to Airbnb hosts without instant booking.

The number of observations varies considerably across cities, together with the average probability of having a non-white guest. For example, the number of observations in New York is more than twice the total in all other cities under analysis. Accordingly, along with all the findings, we also report separate evidence for each city.

Next, we use an event study framework and Fig. 3.5 shows estimates γ_k of Equation 3.4.2 normalized to September 2016. The dashed lines are the 95 percent standard intervals using the cluster standard errors at the listing level. The probability to have a non-white guest does not change after the policy, which confirms the statistically insignificant effect of the previous analysis. The magnitude of δ is not directly comparable with the one of γ_k in the event study set-ups since by estimating δ we do not allow for a heterogeneous impact of each period before and after the policy. Still, it is clear from the graph that all γ_k are small in line with the size of δ.

We can confirm the parallel trend assumption by looking at the periods before the policy. Indeed, the treated and control groups follow the same trend. The small deviation from zero disappears when we consider 90 percent confidence intervals or look separately at each city.

In Fig. 3.6 we present the same specification as before restricting on the subsample of each city: parallel trend assumption holds in all the cases, and the policy does not have a significant effect in any of those cities. The standard errors of the estimates get larger by looking separately at each city comparing with Fig. 3.5 because sample sizes decrease.

3.5.2 Impact on a Per-Night Price

Next, we check the effect on prices. As a result of the policy, hosts who stay on the platform may decide to charge more to have a selected sample of guests coming. In this Subsection, using the same sample of listing as in the previous analysis, we study a potential impact on prices per night.

Table 3.2: Difference-in-Differences Set-Up

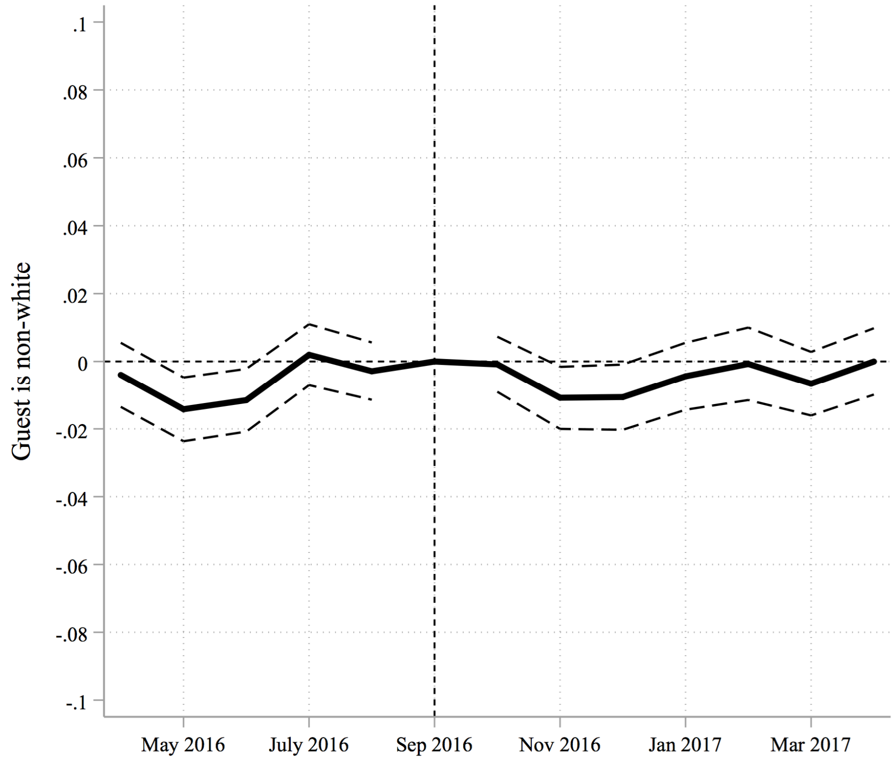
	All cities (1)	San Francisco (2)	New York (3)	Portland (4)	New Orleans (5)
Dependent variable: Guest is non-white					
Mean[Guest is non-white]	.227	.227	.247	.187	.185
Panel (I): No fixed effects, no controls					
Not instant × After	-0.000371	0.000542	-0.00116	-0.00308	0.00141
	(0.00202)	(0.00460)	(0.00301)	(0.00405)	(0.00449)
R^2	0.000132	0.000958	0.000370	0.000181	0.000232
Observations	542190	93960	310138	74511	63581
Panel (II): Fixed effects, no controls					
Not instant × After	0.000157	0.00151	-0.00108	-0.00162	0.00304
	(0.00199)	(0.00453)	(0.00298)	(0.00422)	(0.00464)
R^2	0.0273	0.0144	0.0213	0.0220	0.0210
Observations	542190	93960	310138	74511	63581
Panel (III): Fixed effects, controls					
Not instant × After	-0.000414	-0.00117	-0.000945	-0.000980	0.00307
	(0.00199)	(0.00451)	(0.00298)	(0.00421)	(0.00466)
R^2	0.0306	0.0171	0.0256	0.0243	0.0231
Observations	530560	92417	301941	73413	62789
Panel (IV): Individual fixed effects, controls					
Not instant × After	-0.00194	-0.00305	-0.00207	-0.00491	0.00239
	(0.00214)	(0.00471)	(0.00325)	(0.00448)	(0.00486)
R^2	0.0930	0.0714	0.0974	0.0610	0.0658
Observations	527155	92009	299380	73232	62534

Note: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. The cluster standard errors at the listing level are in parentheses. We restrict on listings entered before August 2016 and exited after October 2016. We restrict the period from May 1, 2016 to April 1, 2017. We report the estimated coefficient δ from Equation 3.4.1. In *Panel (I)*, we control for three dummy variables: Not Instant, After and Not Instant × After. In *Panel (II)*, we add further controls for week, month, city, neighborhood, month-neighborhood fixed effects. In *Panel (III)*, we include observable characteristics: host's ethnicity, how long the listing is present on Airbnb, the total number of reviews, total listing rating, the room type, superhost status, price per night, availability of listings in the next 30 days. *Panel (IV)* controls for all the variables in previous regressions and listing fixed effects.

Identification Strategy

When we study the impact of this policy on the guest's ethnicity, we use each transaction occurred during the snapshot. However, to estimate the effect on prices, we need to use a

Fig. 3.5. Impact of Airbnb's Nondiscrimination Policy



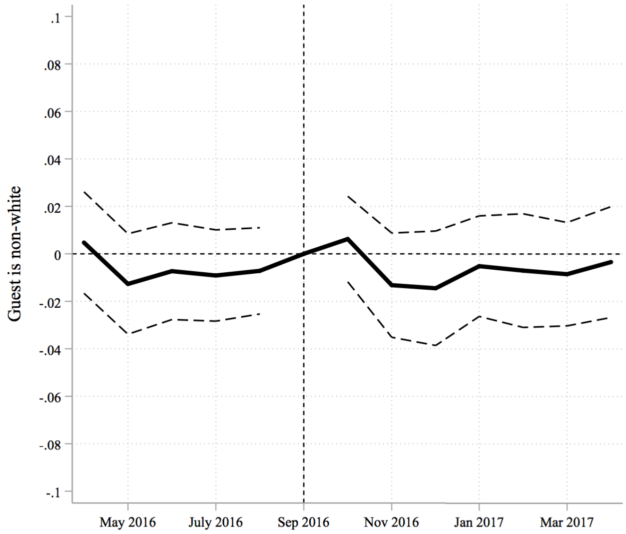
Note: The graph shows the estimated coefficients γ_k from Equation 3.4.2 using all cities. We restrict to listings entered before August 2016 and exited after October 2016. We restrict the period from February 1, 2016, to July 1, 2017. We do not label the endpoints since they are binned. In the regression, we control for observable characteristics: host's ethnicity, how long the listing is present on Airbnb, the total number of reviews, total listing rating, the room type, superhost status, price per night, availability of listings in the next 30 days, and week, month, city, neighborhood, month-neighborhood, listing fixed effects. The dashed lines correspond with 95 percent standard intervals and are based on cluster standard errors at the listing level.

listing-snap dataset since prices vary across snaps but not within. As before, we restrict on listings entered before August 2016 and exited after October 2016, to focus on changes due to policy rather than new entrants. The main equation of *the difference-in-differences set-up* is:

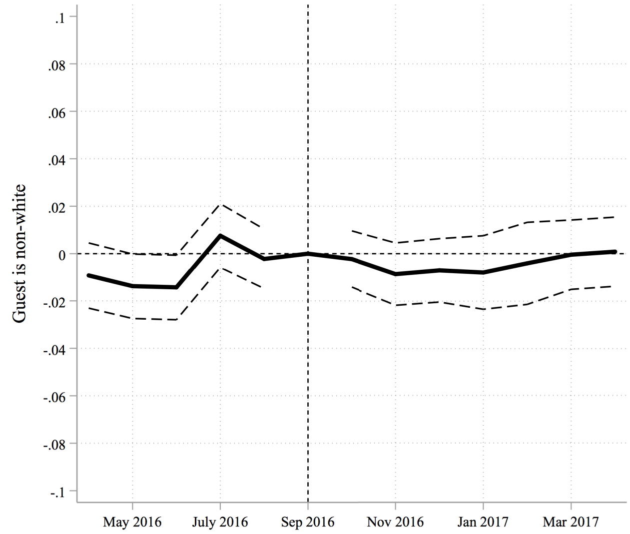
$$Price_{is} = \alpha_1 + \alpha_2 Not\ instant_i + \alpha_3 Snap\ after_s + \delta Not\ instant_i \times Snap\ after_s + \beta \boldsymbol{X}_{is} + \epsilon_{is}, \tag{3.5.1}$$

where $Price_{is}$ is a per-night price for listing i in snap s. $Not\ instant_i$ is equal to 1 if a listing i has instant booking off in all snaps from the beginning of April 2016 to April 2017, it is equal to 0 if a listing i has an instant booking on during the same periods. $Snap\ after_s$ is equal to 1 if a snap is after October 8, 2016, but before April 1, 2017, it is equal to 0 if it is from April 1, 2016, to October 8, 2016. As before $\boldsymbol{X}_{is}$ controls for observable variables and

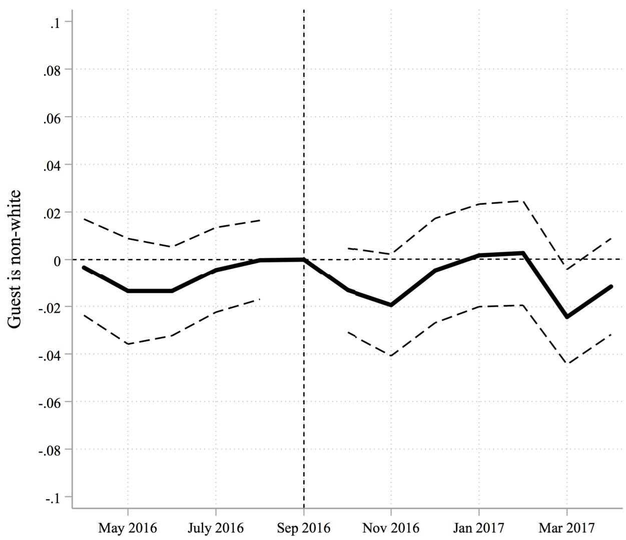
Fig. 3.6. Impact of Airbnb's Nondiscrimination Policy Across Cities



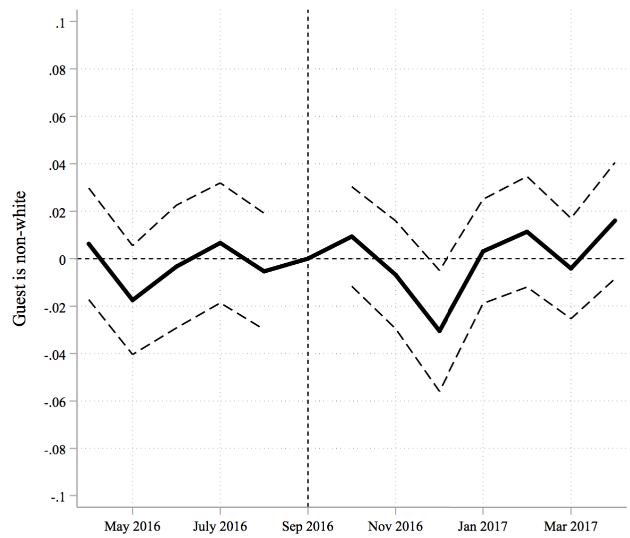
(a) San Francisco



(b) New York



(c) Portland



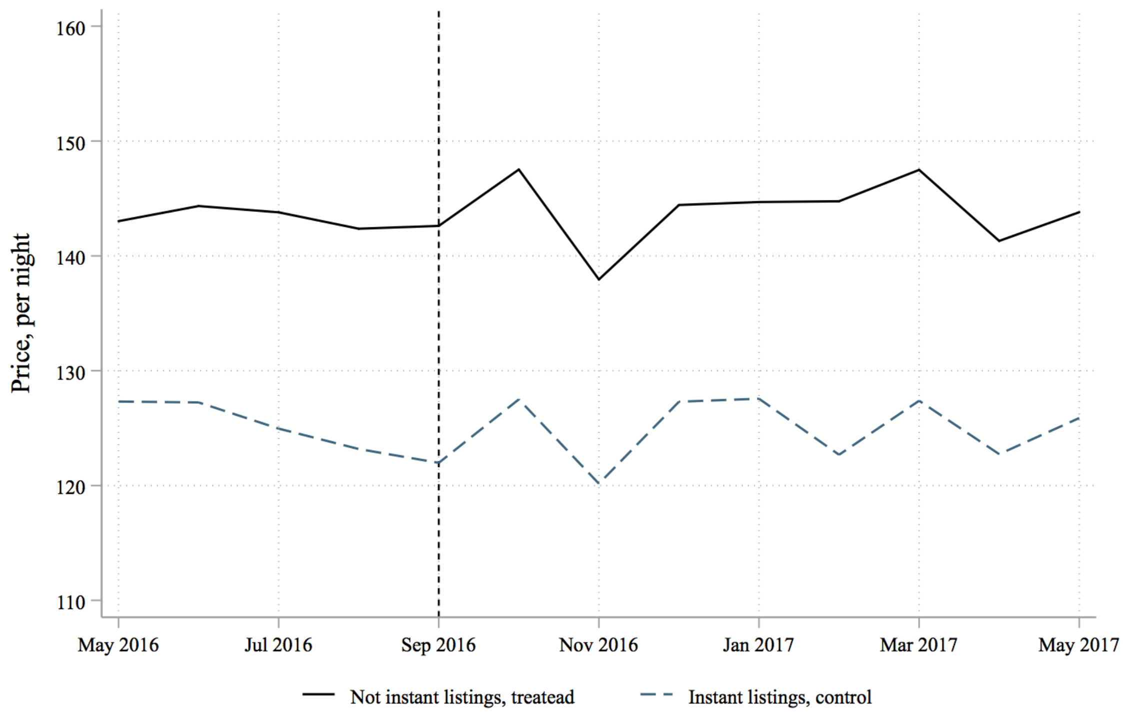
(d) New Orleans

Note: Each subgraph shows the estimated coefficients γ_k from Equation 3.4.2 using each city separately. We restrict to listings entered before August 2016 and exited after October 2016. We restrict the period from February 1, 2016, to July 1, 2017. We do not label the endpoints since they are binned. In the regressions we control for observable characteristics: host's ethnicity, how long the listing is present on Airbnb, the total number of reviews, total listing rating, the room type, superhost status, the price per night, availability of listings in the next 30 days, and listing, week, month, city, neighborhood, month-neighborhood fixed effects. The dashed lines correspond with 95 percent standard intervals and are based on cluster standard errors at the listing level.

unobserved characteristics: host's ethnicity, how long the listing is present on Airbnb, the total number of reviews displayed on the web page, total listing rating, the room type (an entire place or shared apartments including shared rooms), rewarded superhost status, availability of listings in the next 30 days. We also include the price per night in the specification about entry and exit. We also control for time, city, neighborhood, neighborhood × city, and neighborhood × city × time fixed effects. The most restrictive specification exploits the panel data structure of our data and includes listing fixed effects. We use cluster standard errors at the listing level to allow for serial correlation within the listing.

This identification strategy relies on a parallel trend assumption. Specifically, it assumes no difference in charged prices over time before policy between treated and control groups. Fig. 3.7 shows no evidence for the failure of this assumption.

Fig. 3.7. Parallel Trend in Per-Night Prices



As before, to capture the heterogeneous impact of each period, we make an *event study analysis*. The main equation is:

$$Price_{is} = \beta_1 + \beta_2 Not\ instant_i + Not\ instant_i \sum_{\substack{k=-5 \\ k\neq -1}}^{6} \gamma_k b_{is}^k + \beta \boldsymbol{X}_{is} + \theta_i + \epsilon_{is}, \qquad [3.5.2]$$

where θ_i is a listing fixed effect. b_{is}^k is a dummy indicating the number of snaps relative to policy implementation.

Findings

Table 3.3 shows the result of difference-in-differences analysis. The average price per night over the period under analysis is US$147. Columns 1-2 exploit the repeated cross-section, and we see that listings without an instant booking option charge more. However, the policy does not affect prices for listings without instant booking relative to listings with instant booking. The magnitude of δ is not sensitive to the inclusion of listing fixed effects in Column 3.

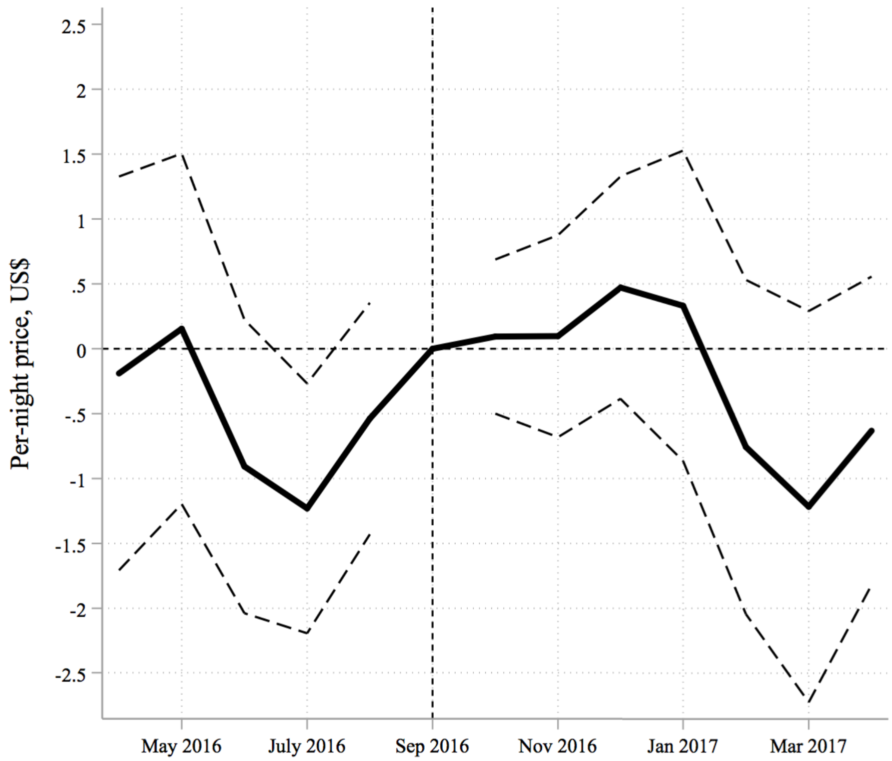
Table 3.3: Difference-in-Differences Set-Pp: Impact on Per-Night Prices

	(1)	(2)	(3)
Mean [Price per night, US$]	147.3	147.3	147.3
Not instant	11.45***	11.45***	
	(1.859)	(1.859)	
Not instant × Snap after	1.144	1.144	0.155
	(1.118)	(1.118)	(0.452)
Constant	138.3***	138.3***	148.2***
	(1.688)	(1.688)	(0.167)
FE	✓	✓	✓
Extra controls		✓	✓
Individual FE			✓
R^2	0.182	0.182	0.977
Observations	207534	207534	203643

Note: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. The cluster standard errors at the listing level are in parentheses. We restrict on listings entered before August 2016 and exited after October 2016. We restrict the period from May 1, 2016 to April 1, 2017. We report the estimated coefficient δ from Equation 3.5.1 changing the dependent variable. In Columns 1 we control for three dummy variables: Not Instant, Snap After and Not Instant × After, host's ethnicity, how long the listing is present on Airbnb, total number of reviews displayed on web page, total listing rating, the room type (entire place or shared apartments including shared rooms), rewarded superhost status, availability of listings in the next 30 days. Column 2 includes further time, city, neighborhood, neighborhood × city, and neighborhood × city × time fixed effects. In addition, Column 3 controls for listing fixed effects.

Next, we plot the event study coefficients γ_j in Fig. 3.8. The results coincide with a difference-in-differences setting since there is no impact of the policy on the prices per night.

Fig. 3.8. Impact of Airbnb's Nondiscrimination Policy on Per-Night Prices



Note: The graph shows the estimated coefficients γ_k from Equation 3.5.2 if the dependent variable is price per night using all cities. We restrict to listings entered before August 2016 and exited after October 2016. We restrict the period from February 1, 2016, to July 1, 2017. We do not label the endpoints since they are binned. In the regression, we control for observable characteristics: host's ethnicity, how long the listing is present on Airbnb, the total number of reviews displayed on the web page, total listing rating, the room type (entire places or shared apartments including shared rooms), rewarded superhost status, availability of listings in the next 30 days, and time, city, neighborhood, neighborhood × city, and neighborhood × city × time listing fixed effects. The dashed lines correspond with 95 percent standard intervals and are based on cluster standard errors at the listing level.

3.5.3 Other Treated Groups

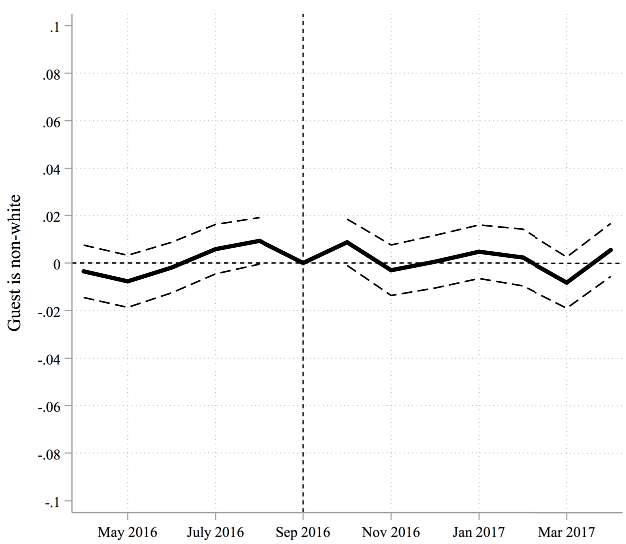
So far, we check the policy's impact on the probability of having a transaction with a non-white guest comparing hosts who either always had instant booking option on or off from April 2016 to April 2017. In doing so, we use 80 percent of all transactions that took place during the period under analysis. The remained 20 percent of transactions occur with listings in which hosts changed their instant booking status at least once between April 2016 and April 2017. Accordingly, this group of hosts who change their status can be of interest themselves, in particular, if they change as the result of the policy. To address this question, in this Subsection, we define two other treated groups by keeping the same control group as in the previous identification.

First, we consider listings that changed their booking status at least once from April 2016 to April 2017, the *switchers*, as a treated group. The identifying strategy remains the same. To ease representation we plot directly the coefficients from the event study set-up in Fig. 3.9a. We see no trend in the pre-period that supports the validity of our strategy. Still, the policy does not have a significant impact on the probability of having a non-white guest.

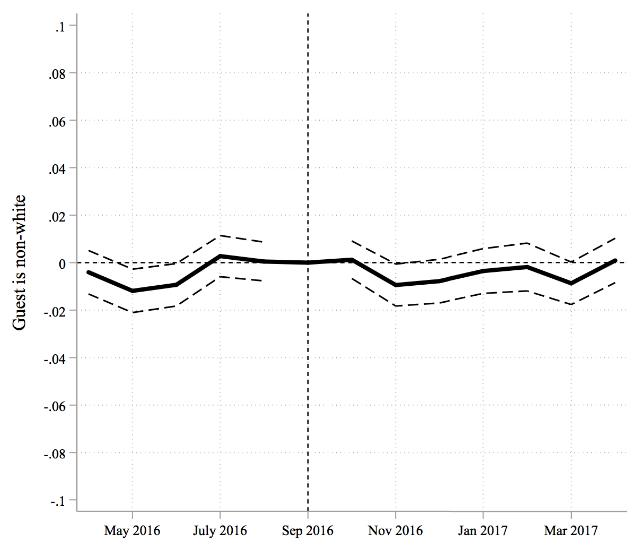
Second, we pool together all transactions and add switchers to the original treated group. This specification covers all universe of hosts and guests on the platform during the period under consideration. Fig. 3.9b shows that even for all transactions there is no effect of the policy.

Since the policy does not impact any group of listings, in the next Section, we show the robustness checks using only our main specification.

Fig. 3.9. Impact of Nondiscrimination Policy using Different Treated Groups



(a) Treated Group Includes Switchers



(b) Treated Group Includes All Listings

Note: Each subgraph shows the estimated coefficients γ_k from Equation 3.4.2 using a different treated group. In the left figure, we define a treated group as listings that changed the instant booking status at least once. In the right figure, the treated groups include listings that have instant booking off and those who switch from April 2016 to April 2017. The control group in both cases is instant booking listings between April 2016 and April 2017. We restrict on listings entered before August 2016 and exited after October 2016. We restrict the period from February 1, 2016, to July 1, 2017. We do not label the endpoints since they are binned. In the regression, we control for observable characteristics: host's ethnicity, how long the listing is present on Airbnb, the total number of reviews, total listing rating, the room type, superhost status, price per night, availability of listings in the next 30 days, and week, month, city, neighborhood, month-neighborhood, listing fixed effects. The dashed lines correspond with 95 percent standard intervals and are based on cluster standard errors at the listing level.

3.6 Robustness Checks

In this Section, we verify whether our results are robust to other specifications. First, we check the impact of the policy on each minority group separately. Then, we consider if the impact differs across hosts or neighborhoods. Further, in Appendix, we discuss the results for the weighted sample, the effect on gender discrimination, and how the effect of the policy can be influenced by the Black Lives Matter movement or Trump's popularity in the country.

3.6.1 Minority Groups

So far, we have pooled together all non-white ethnic groups because of the sample size concern. However, in this part, we repeat the same analysis studying separately the probability of having a transaction with a Black, Hispanic, or Asian guest. Fig. 3.10 shows that policy does not have any impact on any minority guest. In each specification, we use the most restrictive set of controls and replace the variable "host is non-white" with "host is the same ethnicity as a guest."

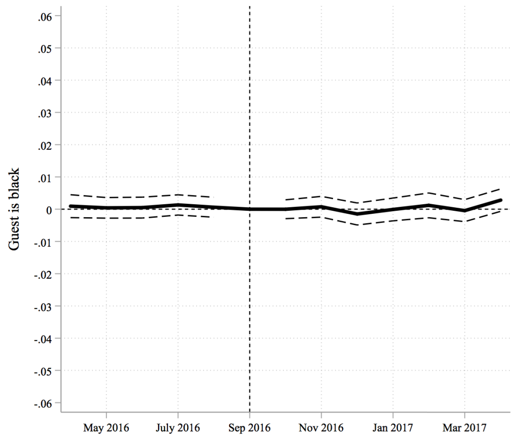
Subsection Host's Ethnicity, Experience and Ethnic Composition of Neighborhood

The other plausible channel of heterogeneity through which the policy affects users can be a change in personal preference across the host's ethnicity, the host's experience, or the ethnic composition of the neighborhood.

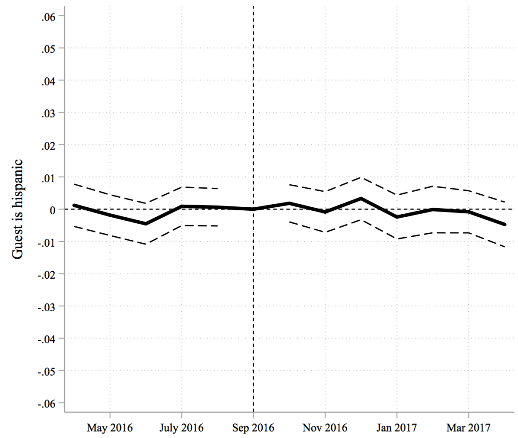
The first potentially important variable to study can be the host's ethnicity. If white and non-white hosts respond differently to the policy, then we should see a different impact restricting to one of them. Fig. 3.11a - 3.11b reports the event study set-up for white and non-white hosts respectively. The share of minority users is small on Airbnb; that is why confidence intervals are particularly large when we restrict to non-white hosts. Still, there is no impact of the policy on guest's inclusiveness on the platform for any host's ethnicity.

Next, we check how the host's experience with non-white guests can change the attitude and the impact of the policy. We consider the hosts with at least 10 reviews before February 1, 2016, and compute the share of non-white guests to all guests before that moment. This is a very restrictive condition since only a few hosts in our sample have an established history almost 10 months before the policy got implemented. Fig. 3.11c - 3.11d confirm that there is not a statistically significant effect on the probability to have a white guest for hosts with an extensive experience with non-white guests due to the policy.

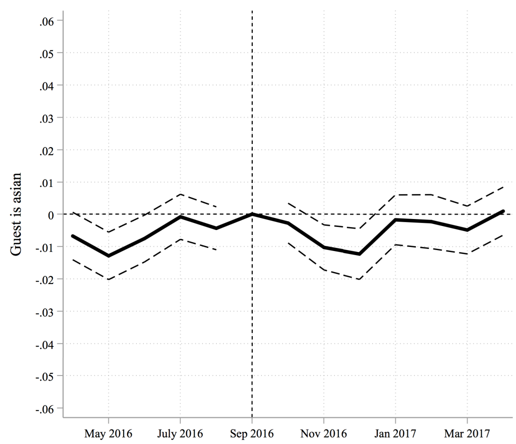
Fig. 3.10. Impact of Nondiscrimination Policy on Each Minority Group



(a) Guest is black



(b) Guest is hispanic



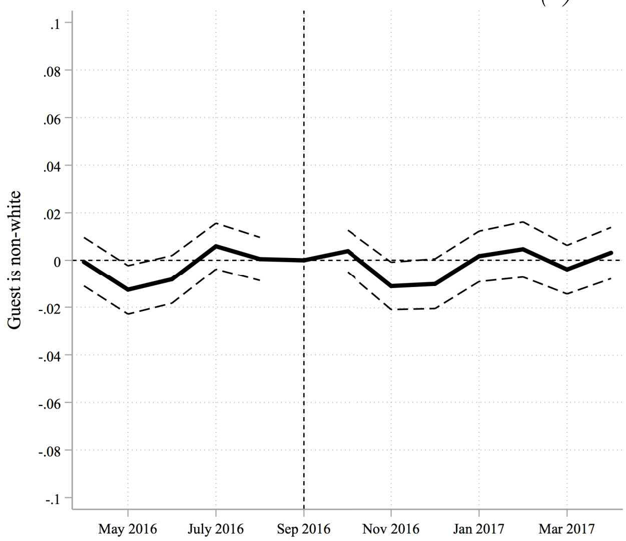
(c) Guest is asian

Note: Each subgraph shows the estimated coefficients γ_k from Equation 3.4.2 using the probability for a guest to be Black, Hispanic, or Asian as the dependent variable. We use all cities and restrict to listings entered before August 2016 and exited after October 2016. We restrict the period from February 1, 2016, to July 1, 2017. We do not label the endpoints since they are binned. In the regression, we control for observable characteristics: host's ethnicity, how long the listing is present on Airbnb, the total number of reviews, total listing rating, the room type, superhost status, price per night, availability of listings in the next 30 days, and week, month, city, neighborhood, month-neighborhood, listing fixed effects. The dashed lines correspond with 95 percent standard intervals and are based on cluster standard errors at the listing level.

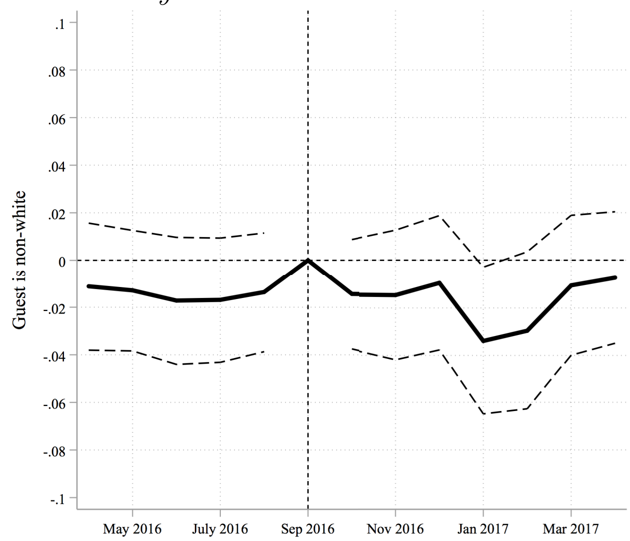
Further, Ewens, Tomlin, and Wang (2014) provide suggestive evidence that discrimination varies across the ethnic neighborhood composition. It means that also, the impact of policy can be heterogeneous across neighborhoods. To do that, we exploit listings latitude and longitude from *Inside Airbnb*, and we identify the 2010 Census Tract to which they belong. It is key to associate listings with the official census boundaries to study the impact of the policy within each part of the city. We merge each listing characteristics with ethnic and economic characteristics

Fig. 3.11. Impact of Nondiscrimination Policy Across Hosts

Panel (I): Host's ethnicity

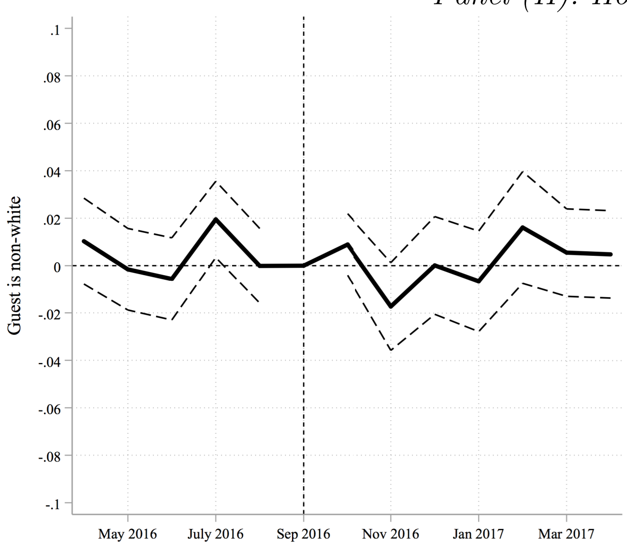


(a) White Hosts

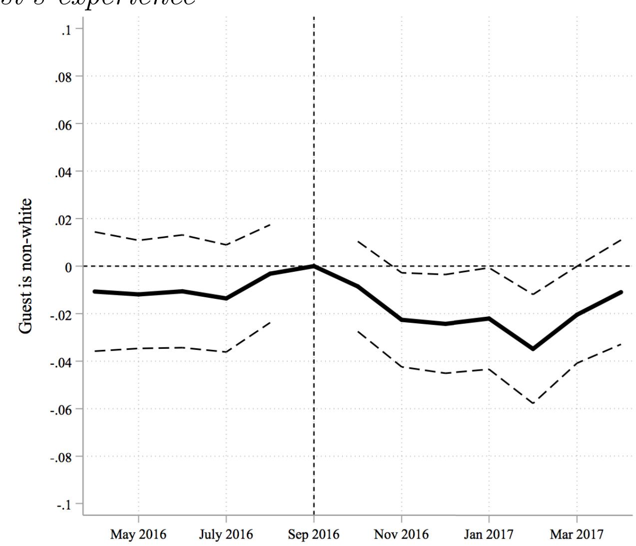


(b) Non-white Hosts

Panel (II): Host's experience



(c) Few non-white guests by February 2016



(d) Many non-white guests by February 2016

Note: Each subgraph shows the estimated coefficients γ_k from Equation 3.4.2 using each city separately. We restrict to listings entered before August 2016 and exited after October 2016. We restrict the period from February 1, 2016, to July 1, 2017. We do not label the endpoints since they are binned. In the regressions we control for observable characteristics: host's ethnicity, how long the listing is present on Airbnb, the total number of reviews, total listing rating, the room type, superhost status, the price per night, availability of listings in the next 30 days, and listing, week, month, city, neighborhood, month-neighborhood fixed effects. The dashed lines correspond with 95 percent standard intervals and are based on cluster standard errors at the listing level.

of the Census Tract using data from the *2012-2016 American Community Survey (ACS) 5-Year Estimates*.[15] For each 2010 Census Tract, we use two variables from ACS: the percent of the white population and the median income. To visualize our procedure, Fig. C.4 shows the administrative boundaries of all cities excluding water territories. Each census tract is colored according to the percentile of the white population in a tract. Yellow dots correspond to listings managed by white hosts in September 2016, whereas dark blue dots - by non-white hosts in September 2016. We see that cities and neighborhoods differ dramatically in the distribution of the non-white population across the 2010 Census Tracts. A similar finding is true for economic characteristics (for the sake of brevity, we do not show the maps with economic variables).

Similarly to Ewens et al. (2014), we divide listings located in census tracts in which the percent of the non-white population is below the 25th percentile and above the 75th percentile of the distribution of non-white in each city. Fig. 3.12a - 3.12b show the event study graphs for each subgroup. They confirm no effect of the policy independently of the racial composition of areas. Despite small sample sizes, there is not any clear pattern across categories.

Finally, we repeat the same exercise studying the neighborhoods' income. We check the 25th and 75th percentiles of the distributions of the median income. These distributions vary a lot across cities: we include the relevant percentiles in each city in Table C.1. Then, we split the sample into three groups: rich, middle, and poor. The sample of rich neighborhoods includes all the listings allocated in the census tracts, where the median income is larger than the 75th percentile of the median income distribution. Similarly, we construct two other subgroups. In terms of both characteristics, the most unequal city is New York.[16] The average probability to have a transaction with non-white guests is slightly higher in poor areas. This pattern is uniform across cities but Portland. In line with the event study results, there is not a statistically significant effect.

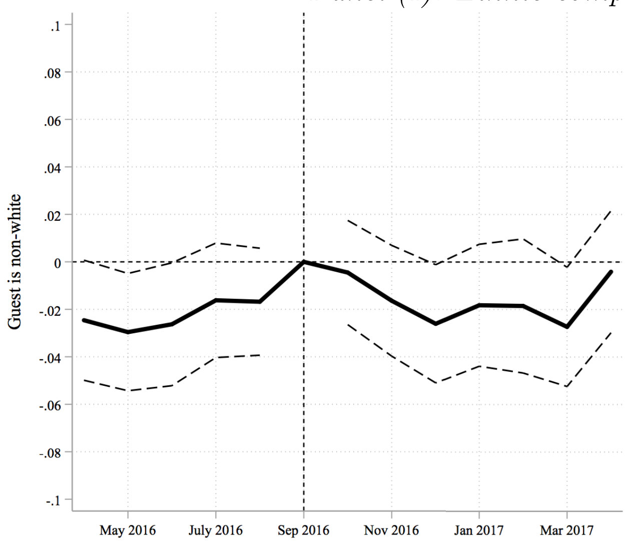
In all the previous specifications, we assign equal weights to all observations in a sample. Still, there is heterogeneity in the number of transactions that listings have. We can assign low weights to listings with many transactions and high weights to listings with few transactions. Appendix 3.10.1 shows the results. Once we apply the logarithmic weights, we lose roughly 10 percent of the sample because we discard all transactions that correspond with the hosts with only one transaction in a snapshot (their weights are zero). Table C.2 reports the findings for

[15] Further details are available upon request. We use a similar approach to San Francisco Planning Department (2012).

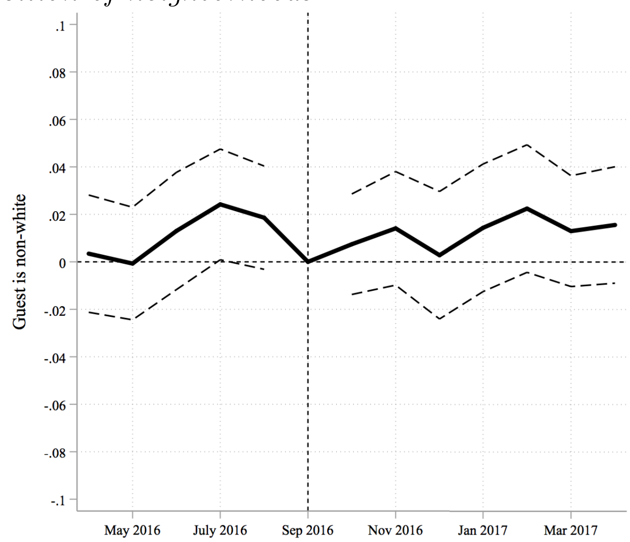
[16] We consider that a city is more unequal if the ratio of the 75th percentile to the 25th percentile is larger. For example, in New Orleans, the median income in rich areas is 1.5 times larger than the median income in poor areas Table C.1

Fig. 3.12. Impact of Nondiscrimination Policy Across Neighborhoods

Panel (I): Ethnic composition of neighborhoods

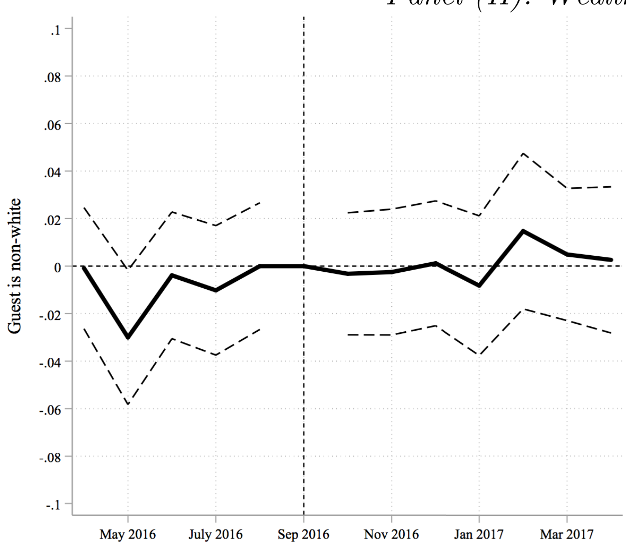


(a) Predominantly White Neighborhoods

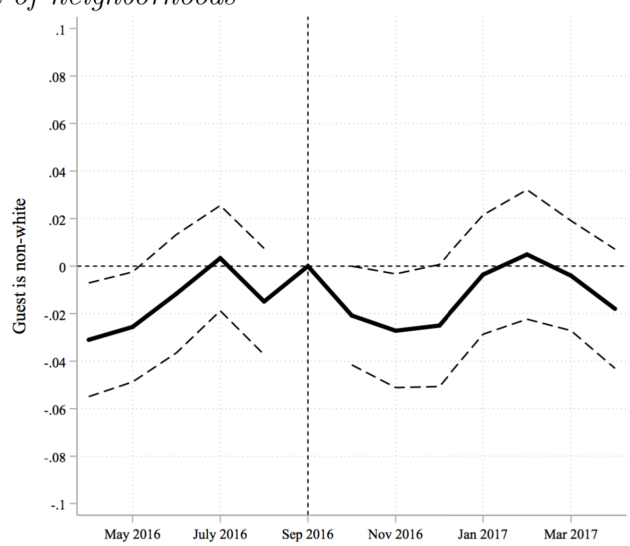


(b) Predominantly Non-White Neighborhoods

Panel (II): Wealth of neighborhoods



(c) Rich Neighborhoods



(d) Poor Neighborhoods

Note: Each subgraph shows the estimated coefficients γ_k from Equation 3.4.2 using each city separately. We restrict to listings entered before August 2016 and exited after October 2016. We restrict the period from February 1, 2016, to July 1, 2017. We do not label the endpoints since they are binned. In the regressions, we control for observable characteristics: host's ethnicity, how long the listing is present on Airbnb, the total number of reviews, total listing rating, the room type, superhost status, the price per night, availability of listings in the next 30 days, and listing, week, month, city, neighborhood, month-neighborhood fixed effects. The dashed lines correspond with 95 percent standard intervals and are based on cluster standard errors at the listing level.

repeated cross-sections. Although the magnitude of the coefficients slightly changes, the impact of the policy is not statistically significant in any subsample. Table C.3 reports the optimal weights. Findings for weighted and unweighted samples are similar.

3.7 Conclusion

This paper studies the impact that the Airbnb Nondiscrimination Policy has on the platform's inclusiveness. We find that, after the policy, the share of non-white guests slightly increases. Moreover, the promotion of the instant booking option leads to a higher share of hosts who accept all booking requests and cannot select guests.

Still, the policy was not able to improve the proportion of non-white guests accepted by those hosts who can discriminate among guests (instant booking off). This may signal the reluctance of certain hosts to accept non-white guests in the absence of a proper nudge by the platform.

In line with the remedies proposed by Edelman et al. (2017), Airbnb increased the number of hosts who behave as regular hotels. However, we believe the platform's next step should regard further effort in assuring an equal and inclusive environment for ethnic minorities by those hosts who do not want or cannot use instant booking. The instant booking option has been criticized by many Airbnb hosts who see it as extraneous from the platform's original nonprofessional, share-economy spirit. Accordingly, a sizable number of Airbnb hosts may never activate the instant booking option, which could jeopardize the creation of a welcoming platform for all minorities.

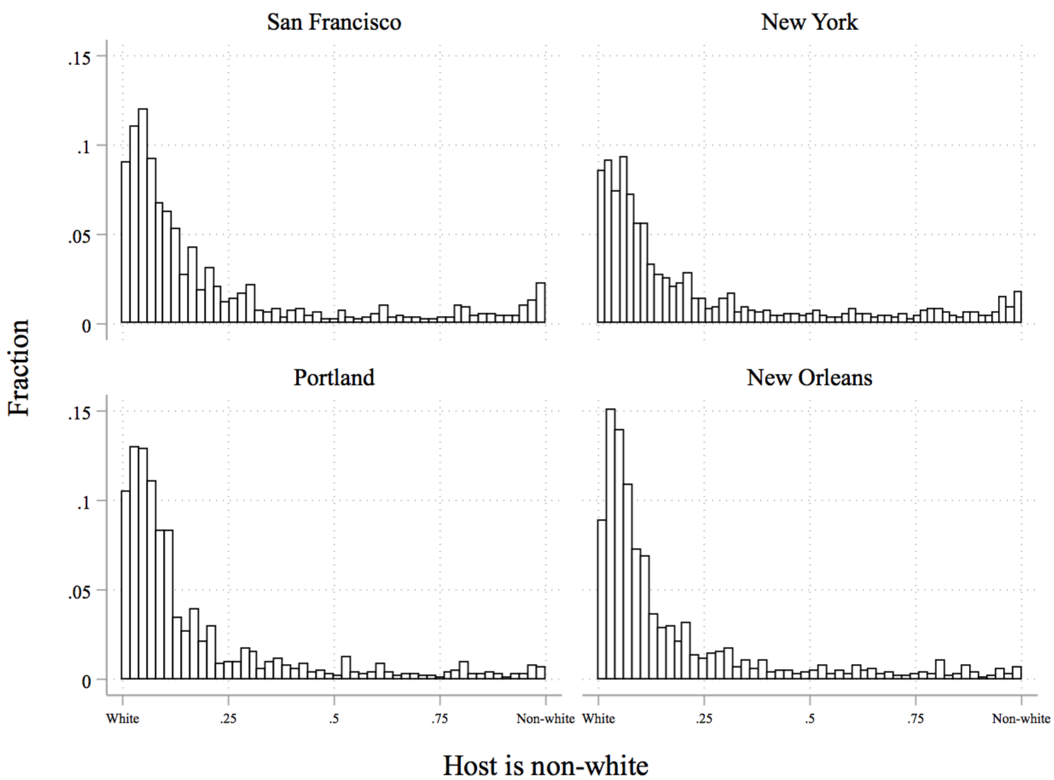
However, we have to acknowledge that this policy was the first attempt at self-regulation by the platform to fight discrimination. Recently, Airbnb launched further changes to become a more inclusive platform and to guarantee the same rights of access to all users.

Appendix C

3.8 Appendix: Data Description

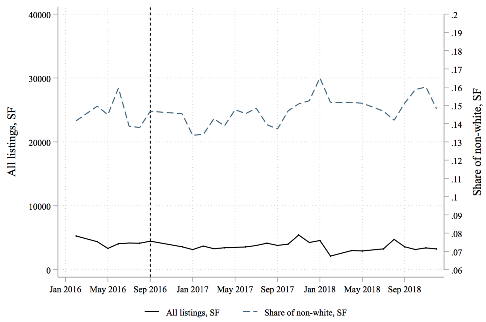
3.8.1 Host's Ethnicity Across Cities

Fig. C.1. Host's Ethnicity Across Cities

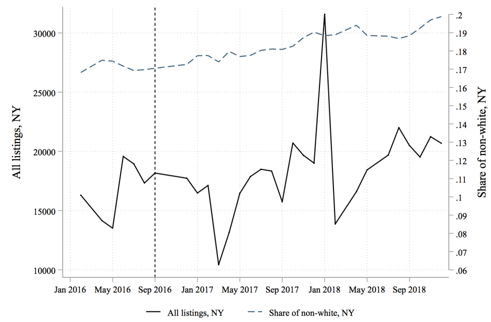


3.8.2 Share of Non-White Hosts and Guests Across Cities

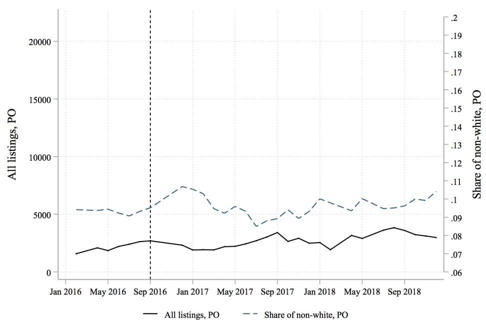
Fig. C.2. Number and Share of Non-White Hosts Across Cities



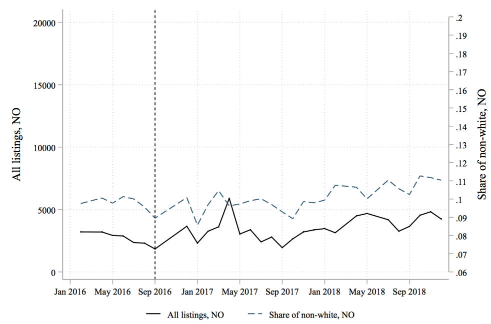
(a) San Francisco



(b) New York

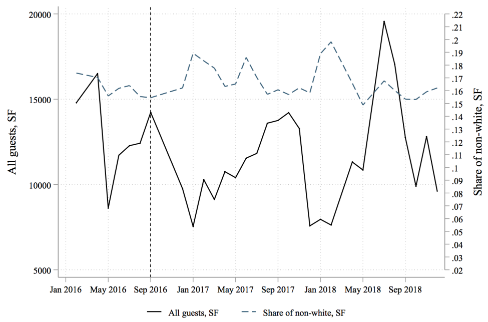


(c) Portland

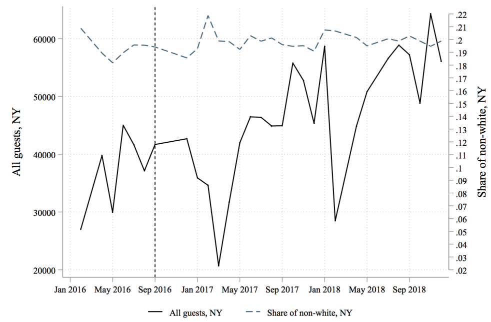


(d) New Orleans

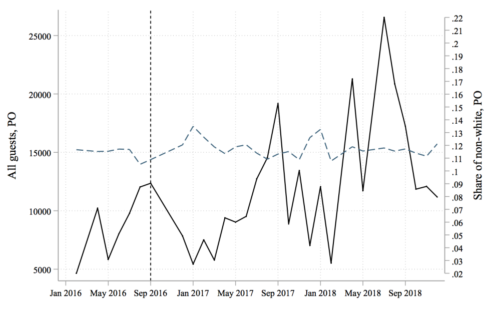
Fig. C.3. Number and Share of Non-White Guests Across Cities



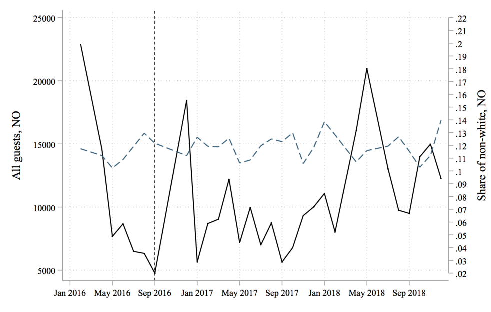
(a) San Francisco



(b) New York



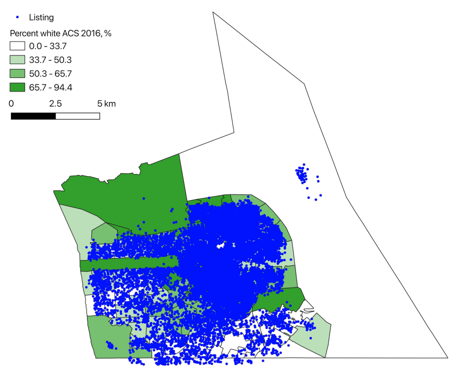
(c) Portland



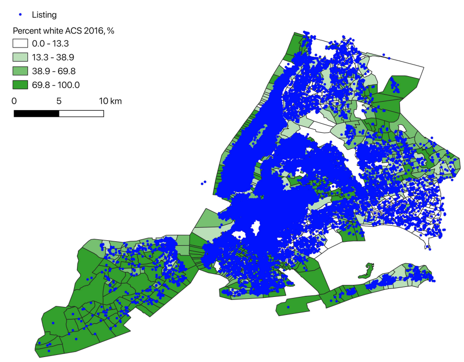
(d) New Orleans

3.8.3 2010 Census Tract Boundaries and Percent of the White Population Across Cities

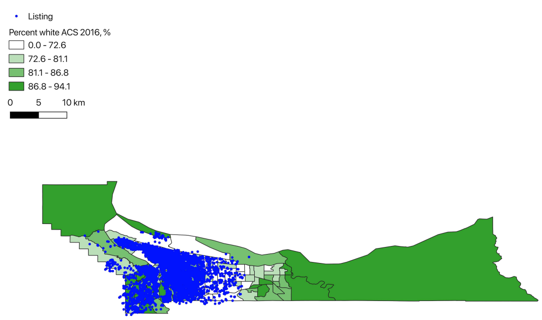
Fig. C.4. 2010 Census Tract Boundaries and Percent of the White Population Across Cities



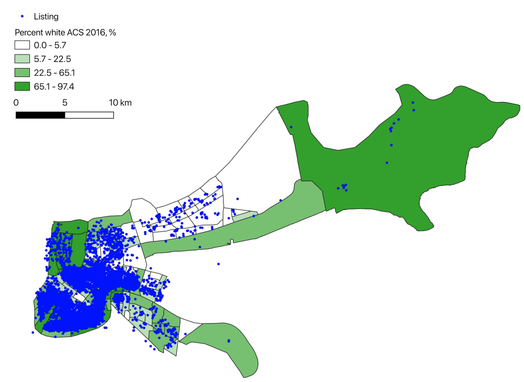
(a) San Francisco



(b) New York



(c) Portland



(d) New Orleans

Note: Each subgraph shows a city's administrative boundary. Each cell corresponds with the 2010 Census Tract and each blue dot represents the listing' geolocation. We plot all the unique listings present on Airbnb during available periods. We color cells according to the percentile of the percent of the white population in a tract using data from 2012-2016 American Community Survey 5-Year Estimates.

3.9 Appendix: Heterogeneity Across the Income of Neighborhoods

Table C.1: Difference-in-Differences Set-Up Across Neighborhoods

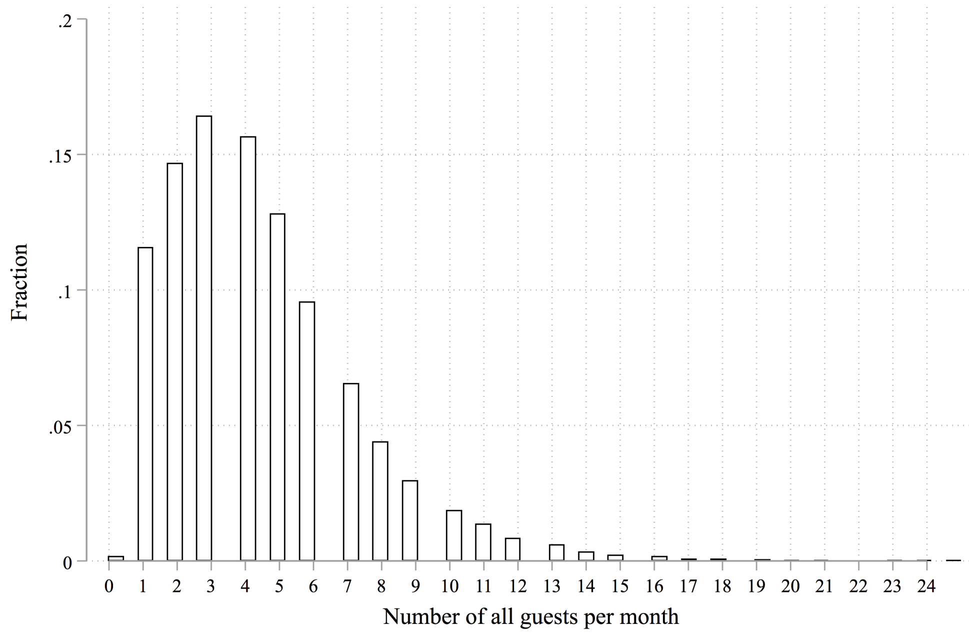
	Neighborhood		
	(1) Rich	(2) Middle	(3) Poor
Dependent variable: Guest is non-white			
Panel (I): All			
Not instant × After	0.00469	0.00144	-0.00173
	(0.00431)	(0.00348)	(0.00932)
Mean[Guest is non-white]	.219		.239
R^2	0.111	0.104	0.114
Observations	77858	180909	84287
Panel (II): San Francisco			
Not instant × After	0.0120	-0.00226	-0.00124
	(0.00944)	(0.00604)	(0.0115)
Mean[Guest is non-white]	.213	.219	.238
25th or 75th percentile	$\geq US\$124,063$		$\leq US\$66,698$
R^2	0.0904	0.0801	0.0926
Observations	16226	29051	12791
Panel (III): New York			
Not instant × After	0.00458	0.00418	-0.00880
	(0.00734)	(0.00385)	(0.00938)
Mean[Guest is non-white]	.234	.242	.256
25th or 75th percentile	$\geq US\$101,554$		$\leq US\$46,140$
R^2	0.117	0.110	0.117
Observations	44659	104268	53025
Panel (IV): Portland			
Not instant × After	-0.0158	-0.00377	0.00656
	(0.0120)	(0.00767)	(0.0146)
Mean[Guest is non-white]	.188	.186	.184
25th or 75th percentile	$\geq US\$83,219$		$\leq US\$50,897$
R^2	0.0944	0.0736	0.112
Observations	11293	23544	7740
Panel (V): New Orleans			
Not instant × After	0.0259	-0.00140	0.0206
	(0.0123)	(0.00886)	(0.0256)
Mean[Guest is non-white]	.179	.181	.196
25th or 75th percentile	$\geq US\$64,333$		$\leq US\$32,862$
R^2	0.116	0.0765	0.0816
Observations	5680	24046	10731

Note: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. The cluster standard errors at the listing level are in parentheses. We restrict to listings entered before August 2016 and exited after October 2016. We restrict the period from May 1, 2016, to April 1, 2017. We report the estimated coefficient δ from Equation 3.4.1. We control for three dummy variables: Not Instant, After and Not Instant × After, host's ethnicity, how long the listing is present on Airbnb, the total number of reviews, total listing rating, the room type, superhost status, price per night, availability of listings in the next 30 days and week, month, city, neighborhood, month-neighborhood, listing fixed effects. A variable 25th or 75th percentile shows the threshold level of distribution to include a listing in one of three subgroups: Rich, Medium and Poor. In Column 1, 2 and 3 we restrict on listings located in census tracts above the 75th percentile, above the 25th percentile and below the 75th percentile, and below the 25 percentile of the median income distribution respectively.

3.10 Appendix: Extensions

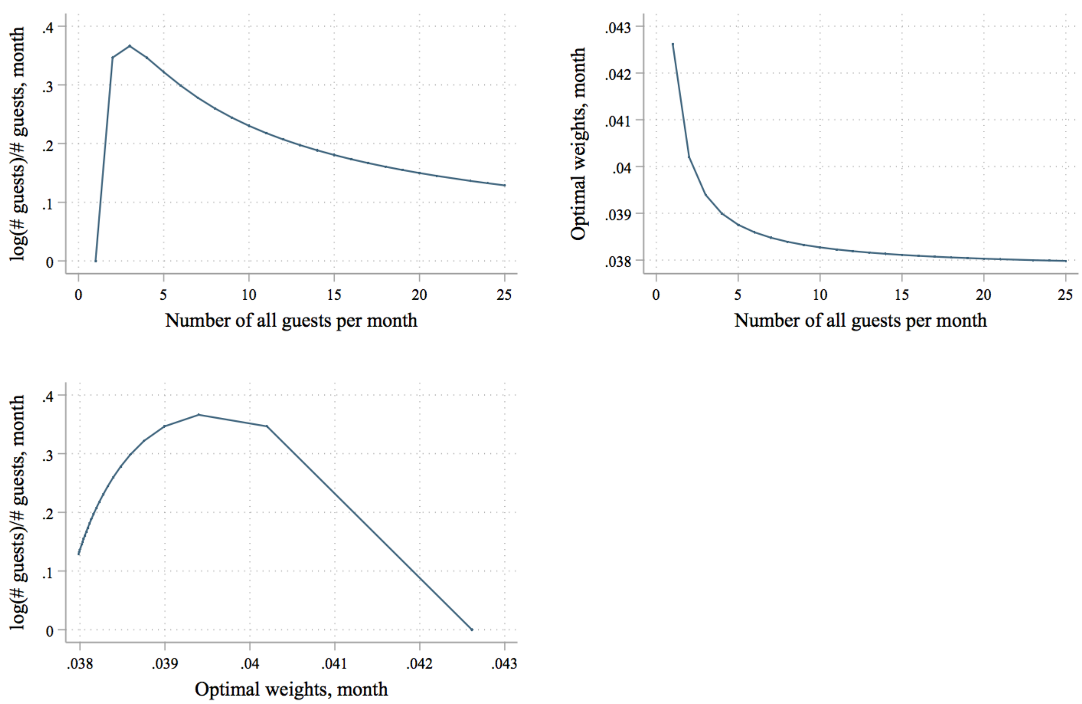
3.10.1 Weighted Analysis

Fig. C.5. Empirical Distribution of the Number of Guests Per Month Per Listing



Note: We pool together all the cities and exploit all the available periods from 2015 to 2019.

Fig. C.6. Distribution of Weights and Optimal Weights Across the Number of Guests Per Month



Note: We pool together all the cities and exploit all the available periods from 2015 to 2019.

3.10.2 Weighted Sample

Table C.2: Difference-in-Differences Set-Up Using the Logarithmic Weights

	All cities (1)	San Francisco (2)	New York (3)	Portland (4)	New Orleans (5)
Dependent variable: Guest is non-white					
Mean[Guest is non-white]	.226	.224	.245	.187	.185
Panel (I): No fixed effects, no controls					
Not instant × After	-0.00185	0.000556	-0.00382	-0.00207	0.000320
	(0.00212)	(0.00486)	(0.00316)	(0.00419)	(0.00475)
R^2	0.000222	0.00117	0.000449	0.000158	0.000252
Observations	473447	84469	261557	70326	57095
Panel (II): Fixed effects, no controls					
Not instant × After	-0.000580	0.00173	-0.00250	-0.00116	0.00213
	(0.00211)	(0.00485)	(0.00314)	(0.00441)	(0.00489)
R^2	0.0282	0.0159	0.0225	0.0224	0.0214
Observations	473447	84469	261557	70326	57095
Panel (III): Fixed effects, controls					
Not instant × After	-0.00138	-0.000974	-0.00266	-0.000203	0.00181
	(0.00210)	(0.00480)	(0.00314)	(0.00440)	(0.00491)
R^2	0.0315	0.0188	0.0267	0.0245	0.0236
Observations	464590	83298	255448	69358	56486

Note: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. The cluster standard errors at the listing level are in parentheses. We restrict on the listings entered before August, 2016 and exited after October, 2016. We restrict the period from May 1, 2016 to April 1, 2017. We use a weighted sample, where each observation gets the following weights: $\frac{\log(N)}{N}$, where N is the total number of guests per month. We report the estimated coefficient δ from Equation 3.4.1. In *Panel (I)*, we control for three dummy variables: Not Instant, After and Not Instant × After. In *Panel (II)*, we add further controls for week, month, city, neighborhood, month-neighborhood fixed effects. In *Panel (III)*, we include observable characteristics: host's ethnicity, how long the listing is present on Airbnb, the total number of reviews, total listing rating, the room type, superhost status, price per night, availability of listings in the next 30 days.

Table C.3: Difference-in-Differences Set-Up Using Optimal Weights

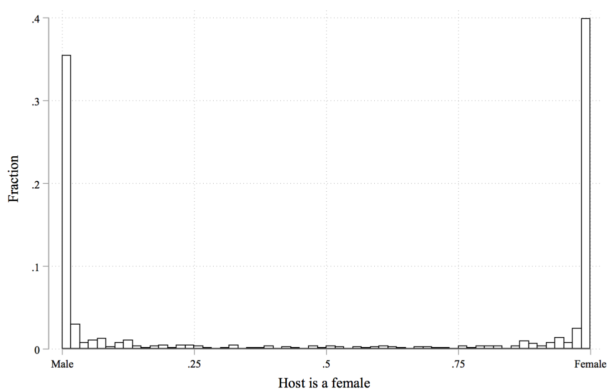
	All cities (1)	San Francisco (2)	New York (3)	Portland (4)	New Orleans (5)
Dependent variable: Guest is non-white					
Mean[Guest is non-white]	.227	.225	.247	.187	.186
Panel (I): No fixed effects, no controls					
Not instant × After	-0.000264	0.000611	-0.00102	-0.00302	0.00128
	(0.00202)	(0.00459)	(0.00301)	(0.00405)	(0.00450)
R^2	0.000130	0.000946	0.000368	0.000180	0.000229
Observations	542190	93960	310138	74511	63581
Panel (II): Fixed effects, no controls					
Not instant × After	0.000175	0.00157	-0.00103	-0.00167	0.00292
	(0.00199)	(0.00454)	(0.00299)	(0.00423)	(0.00464)
R^2	0.0274	0.0144	0.0214	0.0222	0.0213
Observations	542190	93960	310138	74511	63581
Panel (III): Fixed effects, controls					
Not instant × After	-0.000404	-0.00111	-0.000914	-0.00104	0.00300
	(0.00199)	(0.00451)	(0.00299)	(0.00422)	(0.00467)
R^2	0.0307	0.0171	0.0257	0.0245	0.0233
Observations	530560	92417	301941	73413	62789

Note: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. The cluster standard errors at the listing level are in parentheses. We restrict on the listings entered before August, 2016 and exited after October, 2016. We restrict the period from May 1, 2016 to April 1, 2017. We use a weighted sample, where each observation gets the following weights: $\frac{\log(N)}{N}$, where N is the total number of guests per month. We report the estimated coefficient δ from Equation 3.4.1. In *Panel (I)*, we control for three dummy variables: Not Instant, After and Not Instant × After. In *Panel (II)*, we add further controls for week, month, city, neighborhood, month-neighborhood fixed effects. In *Panel (III)*, we include observable characteristics: host's ethnicity, how long the listing is present on Airbnb, the total number of reviews, total listing rating, the room type, superhost status, price per night, availability of listings in the next 30 days.

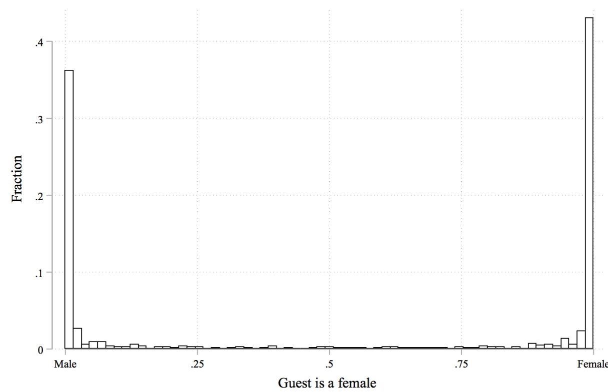
3.10.3 Gender Discrimination

To the best of our knowledge, there has not been a proper discussion on the media about gender discrimination of users on Airbnb (in contrast with what happened regarding ethnic discrimination). If our consideration is correct, by looking at the guest's gender, we should not find any policy effect.

Fig. C.7. Gender of Users on Airbnb



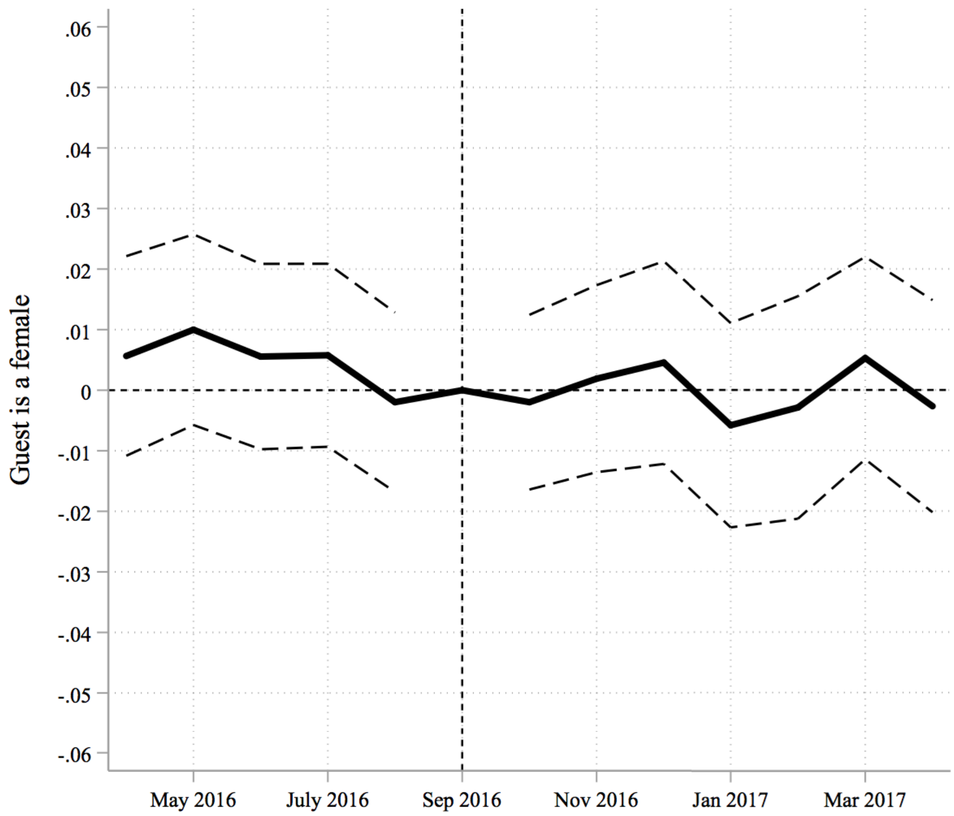
(a) Host



(b) Guest

To estimate the policy impact on having a female guest, we use a new dependent variable. In the main identification strategy, we replace the probability of being non-white with having a guest with a female sounded name. Also, we replace the host's ethnicity with the host's gender in our covariates. Fig. C.8 shows the results of event study. There is no impact of the policy on guest's gender as we expected.

Fig. C.8. Impact of Nondiscrimination Policy on Female Guests



Note: The graph shows the estimated coefficients γ_k from Equation 3.4.2 using the probability for a guest to be a female as the dependent variable. We use all cities and restrict to listings entered before August 2016 and exited after October 2016. We restrict the period from February 1, 2016, to July 1, 2017. We do not label the endpoints since they are binned. In the regression, we control for observable characteristics: host's ethnicity, how long the listing is present on Airbnb, the total number of reviews, total listing rating, the room type, superhost status, price per night, availability of listings in the next 30 days, and week, month, city, neighborhood, month-neighborhood, listing fixed effects. The dashed lines correspond with 95 percent standard intervals and are based on cluster standard errors at the listing level.

3.10.4 Black Lives Matter and Trump Popularity

The other concern for our specification regards confounding effects related to events that nudged or discouraged inclusive behavior of ethnic minorities in society. As examples of these events, we considered the Black Lives Matter movement and Trump's popularity. Both events do not specifically target Airbnb users; still, hosts can accept more non-white due to a general change in the attitude to non-whites.

Using Google Trends for two queries: "Black Lives Matter" and "Trump popularity", we plot their weekly popularity over 2016 in the USA in Fig. C.9, the dashed line shows the moment of policy. There are two peaks of requests around August 2016 and November 2016, the sharp increase in popularity of "Black Lives Matter" can explain the change in the probability to have a non-white guest. To see if it is the case, we use the third specification of Equation 3.4.1 for all the cities and check how robust δ to further controls for "Black Lives Matter" and "Trump popularity" frequency of requests.

Table C.4 reports that the estimate is not stable once we include the new controls. Thus, we cannot rule out that the policy's timing is precise, and in the absence of the other movements that happened in 2016, the policy could lead to different results. Especially, the variable that matters among all is the popularity of "Trump popularity" requests. Further analysis is required to explain the channel, but it is striking that guests get less accepted in not instant booking lodgings once the Trump popularity increases.

Fig. C.9. "Black Lives Matter" and "Trump" Weekly Popularity in the USA in 2016

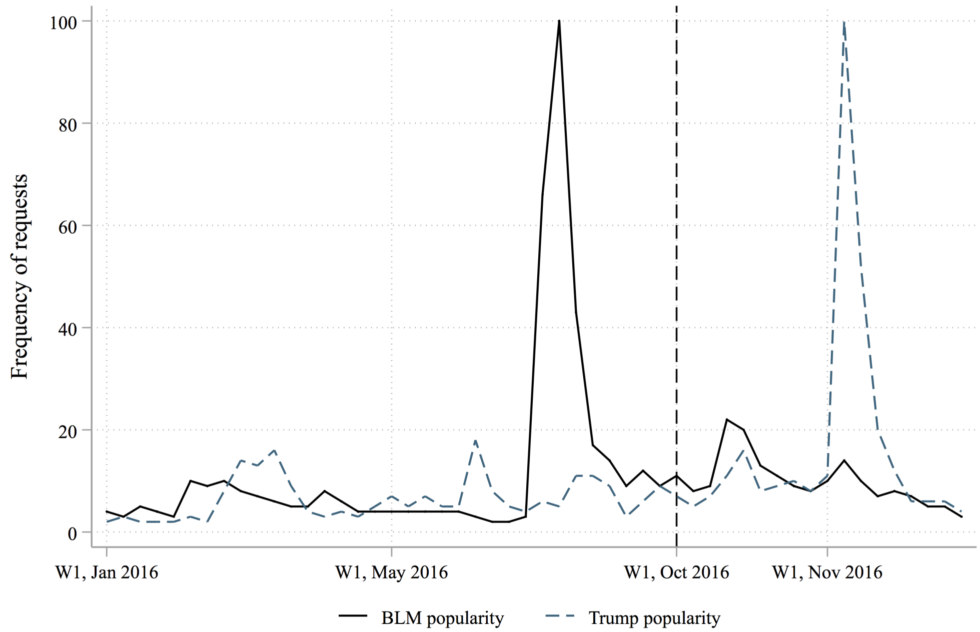


Table C.4: Difference-in-Differences Set-Up Controlling for Black Lives Matter and Trump Popularity

	(1)	(2)	(3)	(4)
Not instant × After	-0.000413	-0.00232	-0.00232	-0.00232
	(0.00199)	(0.00241)	(0.00241)	(0.00241)
BLM popularity		-0.000187		-0.000167
		(0.000304)		(0.000302)
Trump popularity			-0.00588	-0.000619***
			(0.00956)	(0.000111)
R^2	0.0306	0.0287	0.0287	0.0287
Observations	530560	414061	414061	414061

Note: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. The cluster standard errors at the listing level are in parentheses. We restrict on listings entered before August 2016 and exited after October 2016. We restrict the period from May 1, 2016 to April 1, 2017. We report the estimated coefficient δ from Equation 3.4.1. We control for Not Instant, After and Not Instant × After, host's ethnicity, how long the listing is present on Airbnb, the total number of reviews, total listing rating, the room type, superhost status, price per night, availability of listings in the next 30 days and week, month, city, neighborhood, month-neighborhood fixed effects. Column 1 coincides with Column 1, *Panel (III)* in Table 3.2.

Bibliography

Adsera, A., Dalla Pozza, F., Guriev, S., Kleine-Rueschkamp, L., Nikolova, E., oct 2019. Transition, Height and Well-Being. SSRN Electronic Journal.

Alesina, A., Fuchs-Schündeln, N., 2007. Good-bye Lenin (or not?): The effect of communism on people's preferences. American Economic Review 97 (4), 1507–1528.

Atkinson, D., Dallin, A., Lapidus, G. W., 1977. Women in Russia. Stanford University Press.

Backhaus, A., 2019. Fading Legacies: Human Capital in the Aftermath of the Partitions of Poland.

Balia, S., Brau, R., 2014. A country for old men? Long-term home care utilization in Europe. Health Economics 23 (10), 1185–1212.

Barczyk, D., Kredler, M., 2018. Evaluating long-term-care policy options, taking the family seriously. Review of Economic Studies 85 (2), 766–809.

Barczyk, D., Kredler, M., 2019. Long-Term Care Across Europe and the U.S.: The Role of Informal and Formal Care. Fiscal Studies 40 (3), 329–373.

Beblo, M., Görges, L., 2018. On the nature of nurture. The malleability of gender differences in work preferences. Journal of Economic Behavior and Organization 151, 19–41.

Becker, S. O., Grosfeld, I., Grosjean, P., Voigtländer, N., Zhuravskaya, E., 2020a. Forced migration and human capital: Evidence from post-WWII population transfers. American Economic Review 110 (5), 1430–1463.

Becker, S. O., Mergele, L., Woessmann, L., 2020b. The separation and reunification of Germany: Rethinking a natural experiment interpretation of the enduring effects of communism. Journal of Economic Perspectives 34 (2), 143–171.

Becker, S. O., Woessmann, L., 2008. Luther and the girls: Religious denomination and the female education gap in nineteenth-century Prussia. Scandinavian Journal of Economics 110 (4), 777–805.

Bieliauskienė, R., 2014. Education System in Lithuania Historical Aspect. International Interdisciplinary Journal of Scientific Research 1 (2), 43–66.

Boelmann, B., Raute, A., Schönberg, U., 2020. Wind of Change? Cultural Determinants of Maternal Labor Supply.

Bolin, K., Lindgren, B., Lundborg, P., 2008a. Informal and formal care among single-living elderly in Europe. Health Economics 17 (3), 393–409.

Bolin, K., Lindgren, B., Lundborg, P., 2008b. Your next of kin or your own career?. Caring and working among the 50+ of Europe. Journal of Health Economics 27 (3), 718–738.

Bonsang, E., 2009. Does informal care from children to their elderly parents substitute for formal care in Europe? Journal of Health Economics 28 (1), 143–154.

Bonsang, E., van Soest, A., 2020. Time devoted to home production and retirement in couples: A panel data analysis. Labour Economics 65 (2017), 101810.

Brainerd, E., 2017. The lasting effect of sex ratio imbalance on marriage and family: Evidence from world war II in Russia. Review of Economics and Statistics 99 (2), 229–242.

Brugiavini, A., Orso, C. E., Genie, M. G., Naci, R., Pasini, G., 2019. Combining the retrospective interviews of wave 3 and wave 7: the third release of the SHARE Job Episodes Panel.

Brunello, G., Fabbri, D., Fort, M., 2013. The causal effect of education on body mass: Evidence from Europe. Journal of Labor Economics 31 (1), 195–223.

Brzezinski, Z., 1967. The Soviet Bloc: Unity and Conflict. Harvard University Press.

Bukowski, P., 2019. How history matters for student performance. lessons from the Partitions of Poland. Journal of Comparative Economics 47 (1), 136–175.

Bukowski, P., Novokmet, F., 2017. Top Incomes during Wars, Communism and Capitalism: Poland 1892-2015.

Campa, P., Serafinelli, M., 2019. Politico-economic regimes and attitudes: Female workers under state socialism. Review of Economics and Statistics 101 (2), 233–248.

Casado-Marín, D., García-Gómez, P., López-Nicolás, Á., 2011. Informal care and labour force participation among middle-aged women in Spain. SERIEs 2 (1), 1–29.

Chiappori, P.-A., 2015. Gary Becker's Contribution to the Economics of Matching and Marriage. Journal of Demographic Economics 81 (1), 7–11.

Ciccarelli, N., van Soest, A., 2018. Informal Caregiving, Employment Status and Work Hours of the 50+ Population in Europe. Economist (Netherlands) 166 (3), 363–396.

Costa-Font, J., Karlsson, M., Øien, H., 2016. Careful in the Crisis? Determinants of Older People's Informal Care Receipt in Crisis-Struck European Countries. Health Economics (25), 25–42.

Crampton, R. J., apr 2002. Eastern Europe in the Twentieth Century – And After. Routledge.

Crespo, L., López-Noval, B., Mira, P., 2014. Compulsory schooling, education, depression and memory: New evidence from SHARELIFE. Economics of Education Review 43, 36–46.

Crespo, L., Mira, P., 2014. Caregiving to elderly parents and employment status of European mature women. Review of Economics and Statistics 96 (4), 693–709.

Djankov, S., Nikolova, E., Zilinsky, J., 2016. The happiness gap in Eastern Europe. Journal of Comparative Economics 44 (1), 108–124.

Dolado, J. J., García-Penalosa, C., Tarasonis, L., 2020. The Changing Nature of Gender Selection into Employment over the Great Recession. Economic Policy 11.

Duflo, E., 2012. Women empowerment and economic development. Journal of Economic Literature 50 (4), 1051–1079.

Eberhardt, P., 2003. Ethnic groups and population changes in twentieth century Eastern Europe: history, data and analysis: history, data and analysis.

Edelman, B., Luca, M., Svirsky, D., 2017. Racial discrimination in the sharing economy: Evidence from a field experiment. American Economic Journal: Applied Economics 9 (2), 1–22.

Edelman, B. G., Luca, M., 2014. Digital discrimination: The case of Airbnb.com. SSRN Electronic Journal.

Eidintas, A., Bumblauskas, A., Kulakauskas, A., Tamovaitis, M., 2016. History of Lithuania. Eugrimas Pub. House.

European Commission, 2014. 25 years after the fall of the Iron Curtain. The state of integration of East and West in the European Union.

Eurostat - European Commission, 2020. Ageing Europe - Looking at the lives of older people in the EU.

Ewens, M., Tomlin, B., Wang, L. C., 2014. Statistical discrimination or prejudice? A large sample field experiment. Review of Economics and Statistics 96 (1), 119–134.

Fort, M., Schneeweis, N., Winter-Ebmer, R., 2016. Is Education Always Reducing Fertility? Evidence from Compulsory Schooling Reforms. Economic Journal 126 (595), 1823–1855.

Fuchs-Schündeln, N., Masella, P., 2016. Long-lasting effects of socialist education. Review of Economics and Statistics 98 (3), 428–441.

Fuchs-Schündeln, N., Schündeln, M., 2020. The long-term effects of communism in Eastern Europe. Journal of Economic Perspectives 34 (2), 172–191.

Girnius, K. K., 1988. The collectivisation of lithuanian agriculture, 1944-1950. Soviet Studies 40 (3), 460–478.

Goldstein, J. R., Kreyenfeld, M., 2011. Has East Germany overtaken West Germany? Recent trends in order-specific fertility. Population and Development Review 37 (3), 453–472.

Gregg, A. G., 2020. Factory productivity and the concession system of incorporation in late imperial Russia, 1894-1908. American Economic Review 110 (2), 401–427.

Grosfeld, I., Rodnyansky, A., Zhuravskaya, E., 2013. Persistent antimarket culture: A legacy of the pale of settlement after the holocaust. American Economic Journal: Economic Policy 5 (3), 189–226.

Grosfeld, I., Zhuravskaya, E., 2015. Cultural vs. economic legacies of empires: Evidence from the partition of Poland. Journal of Comparative Economics 43 (1), 55–75.

Guriev, S., Melnikov, N., 2018. Happiness convergence in transition countries. Journal of Comparative Economics 46 (3), 683–707.

Guriev, S., Zhuravskaya, E., 2009. (Un)Happiness in Transition. Journal of Economic Perspectives 23 (2), 143–168.

Havari, E., Mazzonna, F., 2015. Can We Trust Older People's Statements on Their Childhood Circumstances? Evidence from SHARELIFE. European Journal of Population 31 (3), 233–257.

Heitmueller, A., 2007. The chicken or the egg?. Endogeneity in labour market participation of informal carers in England. Journal of Health Economics 26 (3), 536–559.

Ichino, A., Winter-Ebmer, R., 2004. The Long-Run Educational Cost of World War II. Journal of Labor Economics 22 (1), 57–86.

Jiménez-Martín, S., Vilaplana Prieto, C., 2015. Informal care motivations and intergenerational transfers in European countries. Health Economics 24 (S1), 89–103.

Kakar, V., Voelz, J., Wu, J., Franco, J., 2018. The visible host: Does race guide Airbnb rental rates in San Francisco? Journal of Housing Economics (40), 25–40.

Kessler, G., Markevich, A., 2019. Electronic Repository of Russian Historical Statistics, 18th - 21st centuries. https://ristat.org/.

Kesternich, I., Siflinger, B., Smith, J. P., Winter, J. K., 2014. The effects of world war ii on economic and health outcomes across Europe. Review of Economics and Statistics 96 (1), 103–118.

Klüsener, S., Goldstein, J. R., 2014. A Long-Standing Demographic East – West. Population, Space and Place 22 (1), 5–22.

Lapidus, G. W., 1978. Women in Soviet Society: Equality, Development, and Social Change. Univ of California Press.

Leong, N., Belzer, A., 2017. The new public accommodations: Race discrimination in the platform economy. Georgetown Law Journal 105 (5), 1271–1322.

Light, M. A., 2012. What Does It Mean to Control Migration? Soviet Mobility Policies in Comparative Perspective. Law and Social Inquiry 37 (2), 395–429.

Lippmann, Q., Georgieff, A., Senik, C., 2020. Undoing Gender with Institutions: Lessons from the German Division and Reunification. The Economic Journal 130 (629), 1445–1470.

Lippmann, Q., Senik, C., 2018. Math, girls and socialism. Journal of Comparative Economics 46 (3), 874–888.

Markevich, A., Zhuravskaya, E., 2018. The economic effects of the abolition of serfdom: Evidence from the Russian Empire. American Economic Review 108 (4-5), 1074–1117.

Michaud, P. C., Heitmueller, A., Nazarov, Z., 2010. A dynamic analysis of informal care and employment in England. Labour Economics 17 (3), 455–465.

Mommaerts, C., Truskinovsky, Y., 2020. The cyclicality of informal care. Journal of Health Economics 71, 102306.

Polugodina, M., Grigoriadis, T., 2020. East Prussia 2.0: Persistent regions, rising nations.

Sablin, I., Kuchinskiy, A., Korobeinikov, A., Mikhaylov, S., Kudinov, O., Kitaeva, Y., Aleksandrov, P., Zimina, M., Zhidkov, G., 2015. Transcultural Empire: Geographic Information System of the 1897 and 1926 General Censuses in the Russian Empire and Soviet Union. HeiDATA: Heidelberg Research Data Repository 2.

San Francisco Planning Department, 2012. San Francisco neighborhoods socio-economic profiles.

Schmidheiny, K., Siegloch, S., 2019. On event study designs and distributed-lag models: Equivalence, generalization and practical implications. CEPR Discussion Paper No. DP13477.

Schweitzer, M. M., 1980. World War II and Female Labor Force Participation Rates. Journal of Economic History 40 (1), 89–95.

Snyder, T., 2004. The reconstruction of nations: Poland, Ukraine, Lithuania, Belarus, 1569-1999.

Solon, G., Haider, S. J., Wooldridge, J. M., 2015. What are we weighting for? Journal of Human Resources 50 (2), 301–316.

Todisco, M., 2015. Share and share alike: Considering racial discrimination in the nascent room-sharing economy. Stan. L. Rev. Online 67, 121.

Valencia Caicedo, F., 2019. The Mission: Human Capital Transmission, Economic Persistence, and Culture in South America. The Quarterly Journal of Economics 134 (1), 507–556.

Van Houtven, C. H., Coe, N. B., Skira, M. M., 2013. The effect of informal care on work and wages. Journal of Health Economics 32 (1), 240–252.

Van Houtven, C. H., Norton, E. C., 2004. Informal care and health care use of older adults. Journal of Health Economics 23 (6), 1159–1180.

Vitola, A., Grigoriadis, T., 2018. Diversity & Empire: Baltic Germans & Comparative Development.

White, S., Batt, J., Lewis, P. G., 1993. Developments in East European Politics. Macmillan International Higher Education.

World Economic Forum, 2020. Global Gender Gap Report 2020: Insight Report.

Wyrwich, M., 2019. Historical and current spatial differences in female labour force participation: Evidence from Germany. Papers in Regional Science 98 (1), 211–239.

Ye, J., Han, S., Hu, Y., Coskun, B., Liu, M., Qin, H., Skiena, S., 2017. Nationality classification using name embeddings. In: Proceedings of the 2017 ACM on Conference on Information and Knowledge Management. pp. 1897–1906.

Ye, J., Skiena, S., 2019. The Secret Lives of Names? Name Embeddings from Social Media. In: Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. pp. 3000–3008.