The Socioeconomic Effects of Wars

by

Andreas Ferrara

Thesis

Submitted to the University of Warwick

for the degree of

Doctor of Philosophy

Department of Economics

April 2019

Contents

Acknowledgments vii

Declarations viii

Abstract ix

Chapter 1 World War II and African American Socioeconomic Progress 1

1.1 Introduction 1
1.2 Black Economic Progress Pre- and Post-WWII 5
1.3 White War Casualties and the Black Occupational Upgrade 8
1.3.1 Computing a Casualty Rate for Semi-Skilled Whites 8
1.3.2 Evidence from Data on Southern Counties, 1920-70 10
1.3.3 Further Evidence from Individual Census Data 15
1.4 The Relation between World War II and Black Economic Progress in the Post-War Era 17
1.5 How the Black Occupational Upgrade Affected Black-White Social Relations in the South in 1961 19
1.6 Conclusion 23
1.7 Tables 25
1.8 Figures 35
1.9 Appendix 47
A Black occupational upgrade 47
B Commuting Zone Appendix 68
C NPPS Additional Results 73
D Data Appendix 83

Chapter 2 The Long-Term Effects of Losing a Father in the U.S. Civil War 1

2.1 Introduction 1
2.2 Historical Background 4
2.2.1 The Union Army: Recruitment and Organization 5
2.3 Data Sources and Record Linking 6
2.3.1 Linking Census and Military Records 8
2.3.2 Final Data Set 9
2.4 Empirical Strategy 11
2.4.1 Selection into the army and selection into death 11

2.4.2 First identification strategy: OLS . . . 13
2.4.3 Second identification strategy: battle IV . . . 13
2.4.4 The impact of linkage errors . . . 14
2.5 Results . . . 15
2.6 Heterogeneity and discussion of results . . . 17
2.7 Conclusion . . . 19
2.8 Tables . . . 20
2.9 Figures . . . 28
2.10 Appendix . . . 31
A Data Sources . . . 31
B Front Line Service and Socioeconomic Regiment Composition . . . 32
C Union Army pensions . . . 39
D The Bias of OLS and IV Resulting from Linkage Errors . . . 40
E Evidence from a Simulation Exercise . . . 45
F Additional results . . . 47

Chapter 3 WWI Anti-German Sentiment and Economic Growth in U.S. Counties 1
3.1 Introduction . . . 1
3.2 Historical Background and Related Literature . . . 4
3.3 WWI Casualties and German Outmigration . . . 7
3.4 Anti-German Sentiment as Channel of Forced Migration . . . 10
3.4.1 Instrumental Variables Results . . . 14
3.4.2 Sensitivity to Linkage Errors . . . 15
3.4.3 Determinants of Migration . . . 16
3.5 Economic Consequences of German Outflows . . . 16
3.6 Conclusion . . . 19
3.7 Tables . . . 21
3.8 Figures . . . 28
3.9 Appendix . . . 36
A Additional Results . . . 36
B Data Appendix . . . 42

Bibliography 1

List of Tables

1.1 Summary Statistics - WWII Enlistment Records . . . 25
1.2 County Data Summary Statistics, 1920-1970 . . . 26
1.3 County Level Difference-in-Differences Results, 1920-1970 . . . 27
1.4 Difference-in-Differences with Casualty Rates by Ethnicity and Skill-Group . 28
1.5 Micro Census Triple Differences Results, 1920-1970 . . . 29
1.6 Triple Differences Results by Industry, 1920-1970 . . . 30
1.7 WWII Casualties and Blacks' Economic Outcomes . . . 31
1.8 Interview Questions and Outcome Coding Scheme . . . 32
1.9 Summary Statistics - Outcome Variables by Race . . . 32
1.10 Summary Statistics - Individual Characteristics by Race . . . 33
1.11 The Skill Upgrade and Black-White Social Relations - OLS and IV Results . 34
1.A.1 Sensitivity Analysis Using Observable County Characteristics . . . 53
1.A.2 Difference-in-Differences Results with Average Soldier Characteristics . . . 54
1.A.3 Difference-in-Differences Results with Alternative Treatment Denominator . 55
1.A.4 Difference-in-Differences Results with Fixed Outcome Denominator . . . 56
1.A.5 Difference-in-Differences Results with First Differenced Outcome . . . 57
1.A.6 Spatial Independence Test of WWII Casualty Rates . . . 58
1.A.7 Testing for Hot and Cold Spots of WWII Casualty Rates . . . 58
1.A.8 County Level Difference-in-Differences Results with Conley Standard Errors 59
1.B.1 Micro Census Triple Differences Results using the Semi-Skilled Treatment . . 69
1.B.2 Movers vs. Birth-State Stayers, all U.S. . . . 71
1.B.3 Movers vs. Birth-State Stayers, South . . . 72
1.C.1 The Skill Upgrade and Black-White Social Relations - Black Sample . . . 74
1.C.2 The Skill Upgrade and Black-White Social Relations - White Sample . . . 75
1.C.3 The Skill Upgrade and Black-White Social Relations - Weighted Regressions 76
1.C.4 The Skill Upgrade and Black-White Social Relations - 1940 to 1960 Differenced Treatment . . . 77
1.C.5 Causal Mediation Analysis Results . . . 82
1.D.1 County Level Difference-in-Differences Results, 1940-1970 . . . 85

2.1 Military Records Summary Statistics . . . 20
2.2 Union Regiment Statistics . . . 21
2.3 Correlates of Death, Disability, and Injury Probabilities . . . 22
2.4 Information on the 10 Major Land Battles of the U.S. Civil War . . . 23
2.5 First Stage of Fathers' Death Probability and the Top 10 Deadliest Battles . . 24

2.6 Effect of Dead or Disabled Fathers on Children's 1880 Socioeconomic Outcomes 24
2.7 OLS Effect of Dead, Disabled, or Injured Fathers 25
2.8 Effect of Dead Fathers on Children's 1880 Socioeconomic Outcomes: Battle IV Results 25
2.9 Death and Disability Effects by Father's Occupation 26
2.10 Heterogeneity by Father's 1860 Wealth 27
2.11 Effect on the Grandchildren 27
2.A.1 List of Sources for the Union Soldier Data 31
2.B.1 Battle Distance Summary Statistics 34
2.B.2 Determinants of Distance to Nearest Enemy on the Battlefield 36
2.E.1 Summary Statistics for Simulated OLS and IV Estimations with a Mis-Measured Binary Treatment due to Linkage Errors 46
2.F.1 Effect of Losing a Father and One or More Brothers 47
2.F.2 IV Results Controlling for Disabled Fathers 48
2.F.3 Effect of Dead or Disabled Fathers, restriction to uncommon names 48
2.F.4 Heterogeneity by child's age when father enlisted 50

3.1 Average Economic and Social Characteristics by Group in 1910 21
3.2 County-Level Effect of WWI Casualties on German Migration, 1870-1940 . . 21
3.3 Summary Statistics for the Linked Census Sample of Germans and German-Americans 22
3.4 OLS and Reduced Form 23
3.5 Determinants of Anti-German Sentiment 24
3.6 Anti-German Sentiment and WWI Casualties First Stage Regressions 25
3.7 Anti-German Sentiment and Migration Decisions - 2SLS Results 26
3.8 Economic Effects of WWI-Induced Outflows of Germans - OLS Regression . 27
3.9 Economic Effects of WWI-Induced Outflows of Germans - 2SLS Regression 27
3.A.1 Differences in Characteristics of Predicted Movers and Stayers 36
3.A.2 OLS and Reduced Form Robustness to Census Linkage Quality 37
3.A.3 2SLS Regression Robustness to Census Linkage Quality 38
3.A.4 Economic Effects of WWI-Induced Outflows of Germans - OLS Regression with Time-Varying Population and Labor Market Controls 39
3.A.5 Economic Effects of WWI-Induced Outflows of Germans - 2SLS Regression with Time-Varying Population and Labor Market Controls 39

List of Figures

1.1 Share of Semi- and High-Skilled Employment Among Black Men, 1870 to 2010 . . . 35
1.2 Number of Drafted and Fallen Soldiers by Month and Year . . . 36
1.3 Draft and Casualty Records Example . . . 37
1.4 WWII Casualty Rates among Semi-Skilled Whites in the U.S. South . . . 38
1.5 Scatter Plots for WWII Casualty Rates and the Share of Blacks in Semi-Skilled Jobs in Levels and First Differences . . . 39
1.6 Unconditional Share of Blacks in Semi-Skilled Jobs by Casualty Rate Quartile 40
1.7 Difference-in-Differences Coefficient Plot . . . 41
1.8 Spatial Distribution of WWII Casualty Rates among Semi-Skilled Whites . . . 42
1.9 Triple Differences Coefficients Plot . . . 43
1.10 Triple-Differences Coefficient Plots: WWII Casualty Treatment, all U.S. . . . 44
1.11 Triple-Differences Coefficient Plots: WWII Casualty Treatment, South only . 45
1.12 Location of NPPS Respondents . . . 46
1.A.1 Difference-in-Differences Coefficient Plots using Alternative Specifications . 60
1.A.2 Difference-in-Differences Cross-County Migration Test . . . 61
1.A.3 Scatter Plots for WWII Casualty Rates and the Share of Blacks in Semi-Skilled Jobs in Levels and First Differences . . . 62
1.A.4 Difference-in-Differences Coefficient Plot with Alternative Treatment . . . 63
1.A.5 Voluntary Enlistment Rates . . . 64
1.A.6 Leave-One Out DiD Sensitivity Check . . . 65
1.A.7 Observable Determinants of Outcome and Treatment . . . 66
1.A.8 Black Semi-Skilled Employment in Levels - Conditional and Unconditional . 67
1.C.1 Conley et al. (2012) IV Bounds . . . 79
1.C.2 Directed Acyclical Graph for Causal Mediation Effects . . . 80
1.D.1 Data Source for Semi-Skilled Employment of Blacks . . . 84

2.1 Record Linkage Schematic . . . 28
2.2 Geographic Distribution of Soldiers Linked to the 1860 Census . . . 29
2.3 Death Rates by County . . . 29
2.4 22nd MA Volunteer Infantry Regiment Records Example . . . 30
2.B.1 Digitizing Civil War Battle Maps . . . 33
2.B.2 Map of Civil War battles . . . 37
2.B.3 Union Army Organizational Chart . . . 38

2.E.1 Simulated OLS and IV Bias with Mis-Measured Binary Treatment due to Linkage Errors . . . 46
2.F.1 Distribution of child's age when father enlisted . . . 49

3.1 Share of Newspaper Articles on Germans Mentioning the Words *Enemy* or *Huns* . . . 28
3.2 Anti-German Posters and Prints during WWI . . . 29
3.3 Spatial Distribution of German Population Flows and WWI Casualty Rate . . . 30
3.4 Change in the Share of Germans over WWI Casualty Rates . . . 31
3.5 Effect of WWI Casualty Rates on the Share of Germans: Top and Bottom Casualty Rate Quintile . . . 32
3.6 Treatment Effect Heterogeneity by State . . . 33
3.7 Newspaper Articles Examples and Spatial Distribution of Anti-German Reporting . . . 34
3.8 Characteristics of Predicted Movers and Stayers . . . 35
3.A.1 Effect of WWI Casualty Rates on the Share of Germans by Casualty Quintile 40
3.A.2 Absolute Birth Year Difference of Linked Individuals . . . 41
3.B.1 WWI Casualty Lists . . . 43
3.B.2 WWI Draft and Enlistment Data . . . 44

Acknowledgments

I am grateful to my supervisor, Sascha O. Becker, for his continued support and guidance over the past ten years, and for inspiring me to pursue a career in academia. He has been a constant source of inspiration, motivation, and knowledge. Price Fishback and Luigi Pascali have also greatly influenced me in my professional development. Their advice and help has been invaluable. I thank Dan Bernhardt, James Fenske, Claudia Rei, and Natalie Deven for their tremendous support, especially during the job market period. Finally, I am grateful to my parents, Karin and Bernd Ferrara, without whom none of this would have been possible.

Abstract

This thesis consists of three essays at the intersection of the fields of economic history and labor economics. Using the case of the United States during the two World Wars and the Civil War, the thesis shows the unintended consequences of wars on the socioeconomic outcomes of those who stay behind especially through the channel of war deaths which disrupt labor markets, family structures, or social attitudes, among others. Chapter one studies how deaths among semi-skilled whites during World War II opened employment opportunities for African Americans from which they had been barred in the past. These improved opportunities in the labor market not led to better economic outcomes for blacks, such as wages, education, or house values, but also led to better black-white social relations such as friendships or attitudes towards integration. Chapter two uses linked Census data and information on soldiers from the U.S. Civil War to study the effects of losing a father on the long-term effects of children. While the negative results are expected, this is one of the first studies to follow children over such a long period of time and it also provides an identification strategy based on allocation of soldiers to battles that were unexpectedly costly. The final chapter estimates the effect of discrimination against Germans in the U.S. during World War I on economic growth. Counties with higher anti-German sentiment during the war years discriminated away their German-born population at the cost of reduced economic growth. This particularly affected the manufacturing sector, a high-productivity sector with a disproportionally large share of German workers.

Chapter 1

World War II and African American Socioeconomic Progress

1.1 Introduction

The gap in the social and economic outcomes and opportunities between blacks and whites has been a constant in the United States.[1] Differences in wages (Bayer and Charles, 2018) and residential segregation (Boustan, 2010) follow stubbornly persistent historic patterns. Changes over the last century have been episodic. The situation for blacks before 1940 was stagnant (Myrdal, 1944), while Margo (1995) and Maloney (1994) documented sharp improvements from the 1940s to 60s which continued through the Civil Rights era (Donohue and Heckman, 1991; Wright, 2013), followed by the decline in black economic fortunes after the mid-1970s (see Bound and Freeman, 1992).

These episodes are reflected in the skill composition of black men and are shown in figure 1.1. The 1940s and the immediate post-war decades stand out. Between 1940 and 1950, the share of semi-skilled employment among blacks almost doubled. In this one decade alone, blacks made more occupational progress than in the 70 years since the end of the Civil War. Collins (2001) called this period a turning point in African American economic history.

In this paper I study the origins of this turning point, and the effect of the unprecedented occupational upgrade on the economic and social status of blacks in the U.S. My main hypothesis is that higher WWII casualty rates among semi-skilled white workers drove the occupational upgrading of blacks. These deaths and the tight labor market during the war years opened employment opportunities from which blacks had been barred in the past. I argue that the casualty-induced occupational upgrade not only improved economic outcomes, such as wages, house values, or education, but that it also had a positive effect on blacks' social status.

[1] For an overview of recent trends, especially with respect to the social outcomes and interactions between blacks and whites, see Fryer (2007).

African American economic progress during the 1940-60s has been studied with respect to the narrowing of the black-white wage gap (Margo, 1995; Maloney, 1994; Bailey and Collins, 2006), migration and urbanization (Boustan, 2009, 2010, 1961), home ownership (Collins and Margo, 2011; Boustan and Margo, 2013; Logan and Parman, 2017), and education (Smith, 1984; Turner and Bound, 2003). Our knowledge about the root causes of this sudden success is less developed and especially its relation to the occupational upgrade is less well studied (Margo, 1995).

The occupational upgrade at mid-century coincides with several major events, including the Great Migration, the first anti-discrimination policies enforced by the Fair Employment Practice Committee (FEPC), and World War II. This makes it challenging to isolate any single cause. The Great Migration to the North and West, which began during the 1940s, substantially benefited African Americans who migrated (Boustan, 2009, 1961). Panel (b) of figure 1.1 suggests tough that the occupational gains were not solely concentrated in the North. The FEPC was disbanded shortly after the war and did not have a strong impact in the South (Collins, 2001).

Previous work on the labor market and educational effects of the war has primarily focused on women (Goldin, 1991; Acemoglu, Autor and Lyle, 2004; Goldin and Olivetti, 2013; Jaworski, 2014; Shatnawi and Fishback, 2018). Two exceptions are Collins (2000) who studies the role of veteran status in black males' economic mobility during the 1940s, and Turner and Bound (2003) who estimate the educational effects of the G.I. Bill on black veterans. The occupational upgrading, however, was mostly driven by non-veterans and especially by the one million blacks who entered semi-skilled employment during the war years (Wolfbein, 1947). The war therefore provides a potential explanation for this development which goes beyond the gains made by veterans.

This paper makes three contributions to the literature. First, I construct a novel data set of military casualty records and combine them with Southern county-level Census data from 1920 to 1970. Difference-in-differences results provide causal evidence that the occupational upgrade of blacks was driven by higher WWII casualty rates among semi-skilled white workers. Using casualty instead of draft rates is motivated by the fact that they are free from the displacement effects created by soldiers returning after the war.[2] The effect of the draft on female labor supply was temporary as returning soldiers displaced most female workers again (see Acemoglu et al., 2004). Casualties instead have the potential to explain the persistent employment effects seen in figure 1.1.

[2]Given the previous literature of WWII and the draft, I always control for the draft rate as well.

Results show that counties with an average WWII casualty rate among semi-skilled whites increased the share of blacks in semi-skilled jobs by 13 to 16% relative to the pre-war mean. The average casualty rate can explain between 75 to 90% of the overall inflow of blacks into this occupational group between 1940 and 1950. The effect is persistent and lasts until the end of the sample period in 1970. The results are robust to several specifications, and placebo tests provide evidence that they are not driven by casualties among race or skill-groups.

To generalize these results to the entire country, I repeat the previous analysis using individual level Census data from 1920 to 1970 in a triple differences estimation framework with the casualty rate treatment being assigned at the commuting zone level. This is to show that occupational upgrading did occur for blacks (both in the South and outside) but not for whites. This is evidence that the war casualties not merely induced a labor supply shock, but that it removed barriers to entry into these occupations which blacks had faced before the war. The individual level data also have the advantage that they can be used to more meticulously probe for effect heterogeneity. In particular, I provide evidence that the upgrading was not driven by differential cross-state migration or education patterns for blacks, and that the upgrading effect was especially concentrated in manufacturing. There was no effect in placebo sectors that remained segregated throughout and after the war such as retail or telecommunications.

Second, I use the same triple differences estimation framework to show that the outcomes considered by previous studies analyzing black economic progress at mid-century are systematically related with the WWII casualty rate among semi-skilled whites. The outcomes include wages, urbanization, migration, home ownership, house values, and educational attainment for blacks.[3] The relationship between the casualty rates, as driver of the black occupational upgrade, and the economic outcomes is strongest for house values, wages, and education. Effects on home ownership are only short-lived and urbanization does not appear to be affected at all. Blacks living in areas with higher casualty rates had a lower probability for migrating out of their birth state. This is likely because the improvements in local employment opportunities reduced the need to relocate to other states. The results are robust to several specifications and inclusion of different types of time trends, and are not driven by differential changes in mobility or educational attainment across blacks and whites, or mere North-South differences. The majority of the outcomes that have been considered in studies of black economic progress at mid-century can therefore be directly linked to the war as one of their common root causes.

[3]For work on wages see Maloney (1994), Margo (1995), and Bailey and Collins (2006), for migration Boustan (1961), for home ownership Collins and Margo (2011), Boustan and Margo (2013), and Logan and Parman (2017), for education Smith (1984), and Turner and Bound (2003).

Third, I return to the Southern-specific context and estimate the effect of the occupational upgrade on blacks' social standing. For the analysis I use individual-level survey data on 1,068 black and white individuals from 24 Southern counties in 1961. Despite the relatively small sample size, the timing is ideal for studying this question as the data were collected before the major Civil Rights legislation, mainly the Civil Rights Act of 1964, as well as before the outbreak of violence during the Civil Rights protests. I instrument the occupational upgrade with the WWII casualty rates in instrumental variables regressions in order to provide causal estimates. Both black and white respondents who live in areas with a casualty-induced occupational upgrade of African Americans are significantly more likely to have an interracial friendship, to live in mixed-race areas, and to favor integration over segregation. Previous work on the Civil Rights movement has argued that it was the Civil Rights Act of 1964 which has brought about the major break from past trends in the economic and social segregation of blacks (Wright, 2013). I offer a new viewpoint wherein these breaks already occur during and due to WWII.

OLS and IV results are similar and estimate an increase in respondents' probability of reporting an interracial friendship, of living in a mixed-race area, and a of favoring integration over segregation. The results are sizable relative to the outcome averages. They are not driven solely by black respondents but are similar across the two groups, and they hold up also for small violations of the exclusion restriction using the test by Conley, Hansen and Rossi (2012).

Studying the relationship between the war and black socioeconomic progress shows how improvements in labor market opportunities for a disadvantaged minority group can positively affect both economic and social outcomes for members of this group. This is a relevant topic for countries with economically and socially segregated minority groups given a literature which shows that such fragmentation is detrimental for societal outcomes (see Alesina, Baqir and Easterly, 1999). It is also related to the debate about the effectiveness of affirmative action policies (Coate and Loury, 1993). Importantly, the casualty-induced shock to blacks' labor market opportunities here is not coming from the potentially endogenous choices of a policy-maker but from a natural experiment. Hence this setting can allow to more cleanly identify the economic and social spillover effects of policies that seek to improve the labor market opportunities for a minority group.

The remainder of the paper is structured as follows. Section 1.2 provides a brief overview of African American economic history in the 20th century to highlight previous directions of research and to put this paper into context. Section 1.3 describes the enlistment and casualty data, features of the draft system, how the data are linked, and how they are used to construct WWII casualty rates by skill group and race. It then outlines the difference-in-differences regression framework used to estimate the effect of casualties among semi-skilled whites on the promotion of blacks into semi-skilled work. This is followed by an extension of the analysis to the whole country using individual level Census data in a triple differences setting. Section 1.4 uses the same individual level Census data and estimation strategy for the South and the entire U.S. to relate the casualty rate measure at the commuting zone level to previously studied economic outcomes regarding African American economic progress. Section 1.5 describes the data and instrumental variables framework to estimate the effect of the occupational upgrade on black-white social relations in a cross-sectional survey in the South in 1961. The final section concludes.

1.2 Black Economic Progress Pre- and Post-WWII

Myrdal (1944) provides an account of the pre-war conditions of blacks in the U.S.: "They own little property; even their household goods are mostly inadequate and dilapidated. Their incomes are not only low but irregular. They thus live from day to day and have a scant security for the future." (p. 205). This is reflected in figure 1.1. Before 1940, 70-90% of black men were employed in low-skilled occupations. In the Southern states, the share of black men in semi-skilled occupations rose by 8 p.p. between 1870 and 1940 but increased by 11.4 p.p. from 1940 to 1950. Blacks made more economic progress in the decade of WWII than in the last seven decades after the end of the Civil War. This exceptional period has attracted the attention of labor economists and economic historians alike. Economic progress for blacks during the 1940s and 1950s has been documented for wages and inequality, education, urbanization and home ownership, among others.

Margo (1995) and Maloney (1994) make two seminal contributions that assess the factors behind black-white wage convergence between 1940-50 in a wage decomposition exercise. Margo (1995) shows that the decrease in black-white wage differentials can be attributed to the Great Compression,[4] but also to the shift of African American workers into better-paying jobs, migration to the North and better education opportunities for blacks. Also Maloney (1994) reaches this conclusion in a similar decomposition exercise. Bailey and Collins (2006) provide a wage decomposition for African-American women in the 1940s. They also document a rapid decrease in the racial wage gap in this period and attribute it to occupational shifts for this group. However, none of these studies examined the causal roots behind the occupational upgrading.

[4]The Great Compression refers to the significant reduction of the dispersion of wages across and within education, experience, and occupation groups (see Goldin and Margo, 1992).

Education for blacks at mid-century developed more steadily. Results by Smith (1984) do not show a particular uptick in educational attainment during the 1940-50 period. The share of illiteracy among blacks declined from 16.3 to 11.5% between 1930-40, but reduced only from 11.5 to 10.2 % between 1940-52 (Smith, 1984). The base for later economic success was founded in improved access and quality of schooling in the earlier part of the century. Aaronson and Mazumder (2011) show that the spread of Rosenwald schools in the South improved educational attainment of blacks with access to such facilities by one year in rural areas for those born between 1910 and 1925. They can explain 40% of the black-white convergence in education for these cohorts. College education for blacks started to increase slowly after WWII (Collins and Margo, 2006), but only increased at a more rapid pace after the 1960s. Turner and Bound (2003) provide evidence that the G.I. Bill significantly increased college education for both black and white men but not for those black veterans who were born in the South.

Outmigration of blacks from the South to Northern cities and its effects on local labor and housing markets has been well documented. Migration from the rural South to the Northern industrial centers during WWII was an opportunity for economic elevation through better employment opportunities (Boustan, 1961). However, while migrants benefited, the additional competition impeded the wage growth of black workers who already lived in the North (Boustan, 2009). The arrival of Southern blacks also produced a response by whites. Boustan (2010) estimates that 2.7 whites departed for each black arrival in a Northern city. White flight might have contributed to increased black home ownership in the city centers, according to Boustan and Margo (2013). Generally, home ownership has increased significantly for African Americans after WWII, though benefits from the G.I. Bill do not appear to drive this result (Logan and Parman, 2017). Moving North was not always related with positive outcomes. For some, this was correlated with higher levels of child mortality or incarceration instead (Eriksson and Niemesh, 2016; Eriksson, 2018).

While there are good explanations for the evolution of black education and the migration patterns at mid-century, there is still little insight into the unprecedented occupational upgrade of African Americans. It cannot be explained by education because black education expanded more gradually and long before the war. Migration alone is not a sufficient explanation as occupational upgrading not only occurred in the North: panel (b) of figure 1.1 documents a very similar pattern for the South. Institutional factors played a role in helping blacks gain better employment or to reduce inequality, but these factors do not appear to play a major role in the South. The Fair Employment Practice Committee (FEPC) generated substantial employment and wage gains for blacks but was ineffective in the South (Collins, 2001).

The FEPC was disbanded shortly after the war and nationwide affirmative action policies were only implemented with or after the Civil Rights Act.

Another strand of the literature mainly attributes post-war black economic and social progress to the Civil Rights movement (see Wright, 2013). Several Supreme Court decisions and laws, most notably the Civil Rights Act of 1964, sought to improve the economic and social equality of African Americans. This includes enforcement of voting rights and interracial marriage after the 1965 Voting Rights Act and the 1967 Supreme Court ruling in Loving versus Virginia, respectively. The affirmative action policies of the 1960s played an important role in desegregating firms (Miller, 2017). Wright (2013) argues that the Civil Rights movement was the main breaking point from past trends and that it set in motion the process of economic and social integration of blacks. Despite the importance of the Civil Rights Act for the social and economic progress made by blacks, figure 1.1 suggests that the break in occupational segregation had already occurred during the 1940s.

If migration, improved education, and other regulatory and institutional factors do not explain the sudden and large occupational shift from low- to semi-skilled jobs for African Americans, the question then is what other factor could have been at the root of this phenomenon. A natural starting point is World War II. Using data from the Civil War, Larsen (2015) provides evidence for how war related labor shortages reduced lynchings of blacks and increased political participation. The labor market effects of World War II, and in particular of the draft, have been extensively studied for women (Goldin, 1991; Acemoglu et al., 2004; Goldin and Olivetti, 2013; Jaworski, 2014; Shatnawi and Fishback, 2018). The effect of the war on African Americans' economic progress has received comparatively little attention.

Labor economists at the time, such as Wolfbein (1947), observed that a, "significant shift occurred from the farm to the factory as well as considerable upgrading of Negro workers, many of whom received their first opportunity to perform basic factory operations in a semiskilled or skilled capacity" (p. 663). He attributed this to the labor shortages during the war. Likewise, Weaver (1945) describes how labor shortages in the aircraft industry opened job opportunities for blacks beyond low-skilled work. If the labor shortages during the war were the only reason, why did the blacks maintain their labor market gains in the post-war period unlike women?

From the historic accounts it appears that the war played a significant role in the skill-upgrade of blacks which translated into other economic gains such as higher wages (Maloney, 1994; Margo, 1995; Collins, 2000), but the precise channel of this lasting effect is not well known. This has been an understudied part of black economic history: "The story of black occupational upgrading is somewhat less well known than the story of black migration" (Margo, 1995, p. 472).

1.3 White War Casualties and the Black Occupational Upgrade

1.3.1 Computing a Casualty Rate for Semi-Skilled Whites

To compute county-specific casualty rates among semi-skilled whites, I match two data sources, the WWII Enlistment Records and the WWII Honor List of Dead and Missing, for the Army and Army Air Force.[5] The Army kept meticulous records of their drafted and enlisted soldiers during the war. Upon entry, an IBM punch card would store a soldier's name, unique Army serial number, age, education, race, marital status, residence, date and place of entry, and their pre-war occupation codified in three-digit groups using the Dictionary of Occupational Titles of 1939. The National Archives and Records Administration digitized these enlistment records.

The data do not contain soldiers in other service branches such as the Navy, Marines, or Coast Guard. However, the 8.3 million individuals in the Army comprise the majority of the 10 million drafted men during World War II. Due to the high manpower demands by the armed forces there was almost no scope for drafted soldiers to choose a service branch (Flynn, 1993). Volunteering provided more choice regarding the branch of service but was forbidden in 1942 to give the military more control over who entered into service (Flynn, 1993). The removal of volunteering came before the largest battles and casualties were sustained but after the majority of the drafting was completed (see figure 1.2). It therefore would have been difficult to form a prior as to which service branch was the least dangerous in order to enlist strategically.

Deferments were only obtained by fathers with dependents, workers in war-related industries and farmers, or conscientious objectors. Out of 40 million men who had been assessed by their local draft boards only 11,896 men registered as conscientious objectors based on religious reasons (Flynn, 1993). Given that the draft was enacted during peacetime, it had to be significantly more just and equal than the prior drafts to pass the substantial resistance by politicians and the public. Going to college or buying out was not possible. Kriner and Shen (2010) show that there was no significant difference in casualty rates across socioeconomic groups during WWII. Only from the Korean War onwards such a gap emerged.

Generally, the willingness to join the war effort was high. Out of 16 million WWII soldiers some 50,000 deserted compared to the 200,000 out of 2.5 million Civil War soldiers (Glass, 2013). There is little historic evidence that draft evasion and avoidance were a major issue during WWII, especially after Pearl Harbor.[6]

[5] The Air Force only became an independent service branch after the war in 1947.

[6] Appendix A shows that results here are not driven by differential volunteering or other soldier characteristics

To supplement the enlistment data with information about a soldier's survival, I digitized 310,000 entries from the WWII Honor List of Dead and Missing. The casualty records include the name, state and county of residence, cause of death, and the Army serial number. The unique serial number is what identifies soldiers across the two data sources. This limits the need to rely on fuzzy name-matching techniques. Figure 1.3 shows examples of the enlistment and casualty records. More details on merging the enlistment and casualty records is provided in the data appendix. Summary statistics for the matched data for different sample splits comparing blacks and whites, enlisted and drafted, and Northern with Southern soldiers are reported in table 1.1. The unconditional death probability is the same across all splits except for the comparison of black and white soldiers. Blacks were mainly employed in comparatively safer support and supply activities due to racist attitudes that saw them unfit for fighting (Lee, 1965).[7] Due to racism in the military, blacks were both drafted and killed at a lower rate and only towards the end of the war did black draft rates approach their population share.

Using the information on residence, race, pre-war occupation and casualty status, the casualty rate among semi-skilled whites in county c can be computed as,

$$\text{Casualty rate}_c = \frac{\text{white semi-skilled casualties}_c \times 100}{\text{white semi-skilled soldiers}_c} \tag{1.1}$$

which is the percentage of those who went to war and who needed a replacement at their pre-war workplace, but did not return. The denominator was chosen to be the number of serving semi-skilled whites rather than the total number of semi-skilled whites in a county. Using the latter is potentially problematic because workers in war related industries had a higher chance of receiving deferments. Without exact knowledge about the number of deferred men it is not possible to compute an accurate measure of wartime demand for alternative labor such as women or black workers.[8]

The spatial distribution of this casualty rate measure for counties in Southern states is plotted in figure 1.4. The casualty rate measure can be constructed for the whole of the U.S. but the outcome variable of interest, i.e. the share of blacks in semi-skilled jobs, can only be computed at the county-level for the mapped Southern states. These states are the only ones to provide occupational counts by race in their county level Census files.

across counties.

[7] Few black fighting units existed, such as the Tuskegee Airmen, but among the almost 1 million black servicemen these made up a small fraction.

[8] For robustness checks, I later also use the casualty measure with the denominator being all semi-skilled whites in 1940 (see appendix A).

1.3.2 Evidence from Data on Southern Counties, 1920-70

The outcome of interest is the percentage share of blacks in semi-skilled employment in county c and decade t. Following the U.S. Census Bureau's occupational classification of 1950, semi-skilled jobs are those classified in the craftsmen and operatives categories. Data refer to male workers only. Aggregate data on the number of employed workers by skill group at the county level is available for the U.S. Census files between 1920 and 1970. After 1970 the county level statistics of the Census underwent significant definitional changes for reported occupations, preventing consistent construction of the outcome after 1970.

An additional restriction is that only Southern states tabulated occupational counts by race.[9] For the 16 states plus D.C. there is a total of 1,388 counties which are kept fixed at their 1940 borders. The definition of county borders is not crucial given that over this period there are almost no creations or removals of counties, nor were there substantial boundary changes (see Forstall, 1996).

The raw correlation between casualty rates and the share of blacks in semi-skilled employment in the cross section of counties and across time is shown in figure 1.5.[10] The plots show a strong linear relationship. The time evolution of the unconditional outcome over quartiles of the casualty rate is plotted in figure 1.6. The outcome trends across casualty quartiles are parallel before the war. After the war in 1950, the share of blacks in semi-skilled jobs is increasing with the casualty rate quartile, with the exception of the lowest quartile which also experiences a short-lived uptick in the outcome in 1960.

The difference-in-differences specification is,

$$\%\ \text{semi-skilled blacks}_{ct} = \alpha_c + \lambda_t + \beta\ \text{Casualty rate}_c \times \text{Post-war}_t + X'_{ct}\phi + \eta_{ct} \qquad (1.2)$$

which allows for variable treatment intensities. Under the usual parallel trends assumption and in the absence of time-varying confounding factors, the coefficient β captures the causal effect of a one percentage point increase in the WWII casualty rate among semi-skilled whites on the share of blacks in semi-skilled occupations after the war.

Time-invariant determinants of the share of blacks in semi-skilled occupations across counties are absorbed by county fixed effects α_c. Time-varying shocks common to all counties are controlled for by time fixed effects λ_t. Alternative specifications include state-specific

[9] These are Alabama, Arkansas, Delaware, Florida, Georgia, Kentucky, Louisiana, Maryland, Mississippi, North Carolina, South Carolina, Oklahoma, Tennessee, Texas, Virginia, and West Virginia, and Washington D.C. Note that even though I refer to mentioned states as "South", this deviates from the typical definition of the South as the former Confederacy, unless stated otherwise.

[10] Conditional scatter plots that partial out county characteristics in 1940 such as population, share of black males, and the share of agricultural and manufacturing employment are shown in appendix A, figure 1.A.3.

flexible time trends ρ_{st} or county-specific linear time trends $\alpha_c t$ to probe for robustness of the results with respect to treatment of the time dimension. This allows for partialling out state- or county-specific secular changes in the outcome that would have occurred in the absence of the casualty shock. This includes the introduction of state-specific legislation, or differences in the underlying economic trends across counties that are not captured by the controls.

The vector X_{ct} contains controls that seek to capture other potential changes in observables that might determine the share of blacks in semi-skilled jobs and which correlate with the casualty rate among semi-skilled whites. The draft rate accounts for the remaining workforce during the war as well as for the share of the male population under threat of being killed in the war. It also provides an estimate of the male population eligible for benefits under the G.I. Bill after the war (Turner and Bound, 2003). To account for spillover effects, I include the average casualty rate in the adjacent counties of a given county c. The log of WWII related spending per capita captures governmental spending as potential stimulus to the local economies (see Fishback and Cullen, 2013). Data for WWII expenditure comes from the County and City Data Book 1947 published by the United States Department of Commerce (2012).

Demographic and political controls include the share of rural population and the share of black men from the Census, and the Republican vote share from data by Clubb, Flanigan and Zingale (2006). To control for factors specific to blacks in the South, the number of lynchings between 1900 and 1930 per 1,000 blacks, and the number of slaves in 1860 (both interacted with decade fixed effects) are included. Lynchings had a significant effect on economic growth generated by black inventors (Cook, 2014). I also include the number of Rosenwald schools per 1,000 blacks, which are significant determinants of black education (Aaronson and Mazumder, 2011) and the share of acres flooded by the Mississippi in 1928 interacted with time as a major shock to internal migration of blacks (Hornbeck and Naidu, 2014).

Given that the manufacturing sector at the time was the main employer of operatives and craftsmen, I also include the number of manufacturing establishments per capita, the average firm size measured as the average number of employees per establishment, the log value added per manufacturing worker as measure for productivity, and the share of employment in manufacturing in a given county.

Agriculture was a major employer for black workers before the war, hence I include variables to rule out that shocks related to agricultural productivity or capital accumulation were driving the shift of blacks to semi-skilled employment. These include the share of land used for agricultural production, the share of acres in cotton, the share of cash tenants as measure for skill available in the agricultural sector that might have been portable to semi-skilled employment, and the average value of machinery per farm. The latter seeks to control for technological changes in the agricultural sector. In particular, the use and quality of tractors expanded at the time, especially in the South and released labor from the farms (see Olmstead and Rhode, Journal of Economic History).

Finally, to account for the major economic changes brought by the Great Depression in the decade just prior to the war, I include measures of New Deal spending per capita from Fishback, Horrace and Kantor (2006). These were distributed as stimulus packages between 1933 and 1935. This includes government loans, money for public works, funds from the Agricultural Adjustment Act (AAA), and by the Federal Housing Administration (FHA), as well as the unemployment rate in 1937. All of these variables are interacted with decade fixed effects. All monetary values are deflated to 2010 U.S. dollars using the CPI provided by the Bureau of Labor Statistics.

An overview of all data sources used to compile the final estimation sample is given in the data appendix. Summary statistics are reported in table 1.2. All remaining variation in the outcome which is not captured by the previously mentioned right-hand side variables is absorbed in the error term η_{ct}. Standard errors are clustered at the county level to account for heteroscedasticity and autocorrelation.

The main results from the estimation of eq. (1.2) are reported in table 1.3 under different model specifications. The effect of a one percentage point increase in the WWII casualty rate among semi-skilled whites on the county share of blacks in semi-skilled occupations is between 0.51 and 0.64 p.p. This effect is significant at the one percent level across all specifications. For an average casualty rate of 3.13% the average effect size thus ranges between 1.6 to 2 p.p. Given the average share of blacks in this skill group in 1940, a $\beta \times 3.13$ p.p. addition corresponds to an increase of 12.9 to 16.1% relative to the pre-war mean. A recent study by Miller (2017) assesses the affirmative action policies under President Johnson in 1965. Affected firms increased their share of black employees by 0.8 p.p. five years after. While the magnitudes are not directly comparable due to differences in sample composition and measurement of variables, it gives context to the effect sizes estimated here.

There was a similar order by President Roosevelt during the war which established the Fair Employment Practice Committee (FEPC). Collins (2001) analyzed its role in the employment of blacks in war related industries. Even though he finds significant effects in the North, he also notes that the FEPC was ineffective in the South due to a lack of cooperation by local authorities. While I do not have measures of the FEPC's effectiveness, the results here are unlikely to be driven by the affirmative action policies under Roosevelt. The FEPC disbanded shortly after the war and new employment policies of this type did not come into effect until the Civil Rights Act of 1964.

Inclusion of the controls does not alter the results in column (2). A potential concern is that some of these controls could themselves be outcomes of the casualty rate, such as the share of manufacturing employment or the share of blacks in a county. To alleviate these concerns, I fix all controls at their pre-war levels in 1940 and interact them with decade fixed effects in column (3). Again the results remain unchanged. Columns (4) and (5) present specifications with flexible state-specific time trends and county-specific linear time trends, respectively, to absorb secular trends in the outcome over time that might otherwise be picked up by the casualty rate.

The final column reports estimates using the doubly-robust selection procedure by Belloni, Chernozhukov and Hansen (2014). Their machine learning covariate selection algorithm tests for the stability of treatment effects and potentially improves inference on such parameters. Suppose that a large set of observed controls includes the most relevant covariates to explain the relation of interest but that these variables are unknown to the econometrician.[11] In a first step, the outcome is regressed on the controls, their squares, and all cross-term interactions, after which the most significant predictors are selected either via LASSO or a simple t-test from a multiple regression if the sample size permits. Here a t-test sufficed. The same is repeated for the treatment, i.e. the casualty rate in this case. In a final step, eq. (1.2) is re-estimated using the union of controls selected in either of the previous two steps. The idea is that the regression learns the most important predictors of outcome and treatment which would be problematic omitted variables.

To probe for the sensitivity of the previous results with respect to the unobservable components, table 1.3 reports the coefficient sensitivity test by Oster (2017) for all specifications. She considers a standard linear regression model $Y = \beta X + W_1 + W_2 + \epsilon$, where $W_1 = \Psi w^o$ is a vector of observable controls and W_2 is an index of unobservables. The treatment variable X here is the casualty rate. She then defines the selection relationship as $\delta\frac{Cov(W_1,X)}{Var(W_1)} = \frac{Cov(W_2,X)}{Var(W_2)}$ and solves for δ (the degree to which selection on unobservables is less than or larger than selection on observables) which would be required to produce $\beta = 0$. This uses the coefficient and R^2 movement from the controlled and uncontrolled regressions results in a bounding argument.

Assuming that W_1 and W_2 can fully explain variation in the casualty rate, i.e. $R_{max} = 1$ in a regression of the casualty rate on W_1 and W_2, a reasonable threshold for the previous

[11]These most influential explanatory variables potentially include interactions and squared terms.

results in table 1.3 to be considered robust is $\delta \geq 1$. This implies that the selection on unobservables would need to be at least as important as selection on observables in order to yield a coefficient of zero for the casualty rate. With the exception of column (5) all specifications pass this threshold.

The main assumption underlying eq. (1.2) is the parallel trends assumption. With a continuous treatment, a typical approach is to generate placebo treatments in order to test whether the casualty rate had an effect on the outcome before there were any casualties. Such differences across high- and low-casualty rate counties would hint towards pre-existing trends in the outcome which would bias the coefficient β. The placebo tests are implemented by estimating,

$$\text{\% semi-skilled blacks}_{ct} = \alpha_c + \lambda_t + \sum_{k \neq 1940} \beta_k \text{ Casualty rate}_c \times \text{Year}_k + X'_{ct}\phi + \eta_{ct} \quad (1.3)$$

for which results are plotted in figure 1.7. The specification includes controls and the state-specific flexible time trends. The coefficients plot shows that up until the war the average conditional evolution of the outcome over time was parallel across counties with differing casualty rates. The coefficients from the interaction of the casualty rate with the post-war decades in $k > 1940$ are similar to the effect estimated in table 1.7. The effect remains stable and persists in the three decades after the war. Miller (2017) also finds a persistent effect of the 1960s affirmative action policies which remains even after their removal.

Another way to attempt to falsify the previous results is to consider the effect of casualty rates in other skill groups for both blacks and whites. If the claim here is correct that it was the death of semi-skilled whites that led to the occupational upgrade of African Americans, then we should not see any effect coming from casualty rates in other skill-race groups. The results are reported in table 1.4 which includes casualty rates by race and skill group in the regression. The estimated coefficients for the semi-skilled white casualty rate are not significantly different from what was estimated in the baseline specification. There is no detectable effect for the casualty rates among low- and high-skilled whites.

Likewise, casualty rates for semi- and high-skilled blacks do not have a significant impact on the outcome. However, there is a smaller but significant negative effect coming from the group of low-skilled blacks. A percentage point increase in the casualty rate for this group decreases the share of semi-skilled blacks by 0.09 to 0.15 p.p. This result is intuitive given that these are the workers who, had they survived, would have replaced the deceased semi-skilled whites after the war.[12]

[12] All further robustness and sensitivity analyses are reported in appendix A, including further specification tests

1.3.3 Further Evidence from Individual Census Data

The previous results show that the occupational upgrading of blacks also occurred in the South and was not merely a phenomenon driven by the Great Migration. Yet it is also insightful to generalize the result to the entire country. Doing so requires to assign casualty rates at the commuting zone level instead of the county level. Commuting zones are clusters of counties that share a common labor market. There are 722 commuting zones which can be consistently constructed using the spatial information available in the individual level data of the 1920 to 1970 U.S. Census files by Ruggles, Flood, Goeken, Grover, Meyer, Pacas and Sobek (2018).[13] Figure 1.8 plots the WWII casualty rate among semi-skilled whites at the commuting zone level.

I use the 1% micro Census files from 1920 to 1950, the 5% file of 1960, and the 1% form metro sample of 1970. The estimation sample includes the non-institutionalized working age (16-65) male population who were participating in the labor force at the enumeration date, who were not enrolled in school or classified as unpaid family workers, and whose ethnicity was classified as black or white. The micro level data provide the advantage of using whites an additional control group. If casualties resulted in a labor supply shock only, then one would expect occupational upgrading to occur for both blacks and whites. However, if semi-skilled professions had higher barriers to entry for blacks that were removed due to the labor shortages induced by the casualties, then only blacks should see an effect on their probability to be employed in such jobs.

In the following triple difference (DDD) regression I compare the probability of semi-skilled employment between blacks and whites, before and after the war, and across commuting zones with differing casualty rates:

$$\begin{aligned}\Pr\left(\text{semi-skilled}=1\right)_{izt} &= \beta_1\left(\text{casualty rate}_z \times \text{post-WWII}_t\right)\\ &\quad + \beta_2\left(\text{casualty rate}_z \times \text{black}_{izt} \times \text{post-WWII}_t\right)\\ &\quad + \alpha_z + \lambda_t + \delta\text{black}_{izt} + X'_{izt}\gamma + \epsilon_{izt}\end{aligned} \tag{1.4}$$

where i, z, and t index individuals, commuting zones, and Census years, respectively. The outcome is an indicator for whether an individual is a semi-skilled worker (craftsman or operative). The coefficients of interest are β_1 for whites and the triple interaction coefficient β_2 for

of the parallel trends assumption, selective migration of blacks, selection on observables, selection of soldiers into the military and into death, alternative treatment and outcome denominators, sensitivity of the results by state, and spatial clustering of the casualty rates.

[13]The crosswalks for 1950 and 1970 are available on David Dorn's website (http://www.ddorn.net/data.htm), and the crosswalk files for the other years were kindly shared by Felix König.

blacks. Controls include age, marital status, year of birth, a self-employment indicator, farm status, and industry fixed effects, and α_z and λ_t are commuting zone and time fixed effects. Standard errors are clustered at the commuting zone level.

The triple differences regression seeks to eliminate potentially confounding trends in the employment probability of blacks in semi-skilled jobs across commuting zones that are unrelated to the war casualties. It also accounts for changes in the employment probability of all workers in high-casualty commuting zones which might have happened due to other shocks that occurred at the same time. Compared to the county level regressions, this framework also allows to estimate the casualty rate effect on i) whites, and ii) on blacks and whites in different industries for the entire U.S.

To visualize the relationship, I interact the casualty rate$_z$ and casualty rate$_z$ $\times$ black$_{izt}$ variables with Census year fixed effects in eq. (1.4), leaving out 1940 as baseline. The resulting coefficients for blacks and whites are plotted in figure 1.9. There is no significant casualty rate effect before the war for either group and remains insignificant for whites also in the post-war period. This means that there are no differential pre-trends for blacks or whites across high- and low-casualty rate commuting zones. For blacks there is a positive post-war effect starting from 1950 which increases over time and peaks in 1970 with a 5 p.p. rise in the semi-skilled employment probability for every one percentage point increase in the commuting zone WWII casualty rate among semi-skilled whites.

Table 1.5 reports results from estimating eq. (1.4) for different model specifications. The triple difference coefficient for black workers is positive and significant in all specifications and ranges between 1.9 to 4.7 p.p. for the whole country and between 1.1 and 3 p.p. for workers in the South. There is no effect on whites with the exception of column (6) where the regression with commuting zone specific time trends shows a small but negative and significant effect for white workers. The null effect on whites is coherent with the historic account by Wolfbein (1947): "the movement of [black] men and women to factories, primarily as semiskilled operatives, was even more pronounced than that of white persons" (p. 665).

The results show that the employment gains for blacks not only occurred in the North or West of the country but that also Southern blacks gained significantly in terms of the occupational upgrading. Another advantage of the micro data is that I can further deal with potential migration responses. I therefore interact an indicator for whether an individual lives outside their state of birth with time fixed effects and the black indicator in column (4). The same interactions are applied to the education variable. The results show that even though the coefficients are smaller, they are still positive and significant. It should be noted that migration and education are potential outcomes of the treatment, hence results from this specification are to be taken with caution. Yet it sheds light on whether the occupational upgrading effect can be explained away by differential migration or educational attainment across black and white workers over time.

Next, I analyze whether the occupational upgrading of blacks is concentrated in particular sectors. Table 1.6 repeats the analysis for the manufacturing sector as a whole, and for the durable and non-durable manufacturing sub-sectors, as well as for telecommunications, retail, and public administration as placebo groups. Unlike the manufacturing sectors, the jobs in the placebo sectors often involved direct customer contact and therefore employers sought to avoid employment of blacks in such positions (Anderson, 1982). Given that these sectors remained segregated throughout and after the war, they should not show any occupational gains made by blacks. The results provide evidence that black occupational upgrading was particularly pronounced in all manufacturing sectors with a 9 to 11 p.p. increase in the probability of semi-skilled employment for blacks for a one percentage points increase in the WWII casualty rate among semi-skilled whites. Except for a slight negative effect in retail, there is no effect on blacks in the high-skilled sectors and for whites the effect is never significant in any sector.

1.4 The Relation between World War II and Black Economic Progress in the Post-War Era

Several scholars have studied black economic progress at mid-century with respect to wages (Margo, 1995; Maloney, 1994), cross-state migration (Boustan, 1961) and urbanization (Boustan, 2010), home ownership (Collins and Margo, 2011; Boustan and Margo, 2013; Logan and Parman, 2017), or education (Smith, 1984). If African Americans made progress on all these dimensions and at the same time, then it is likely that there exists at least one underlying common factor. Both Maloney (1994) and Margo (1995) discussed the labor shortages during the war as potential reason for the wage gains made by black workers. According to Margo (1995), "the most important example of occupational upgrading was the increase of blacks in semi-skilled operative positions. Such jobs paid far better than farm labor [...] that blacks were accustomed to" (p. 472).

I next study the war, and in particular the role of semi-skilled white casualty rates as driver of the black occupational upgrade, as common denominator for the post-war progress made by blacks on other economic dimensions analyzed in prior work.[14] I again use the individual level data from the Census between 1920 and 1970 from the previous section. To

[14] Appendix B performs this analysis using semi-skilled employment as treatment for comparison purposes. The casualty rate is the more exogenous variable and hence was preferred for the main specification.

test the hypothesis that other economic improvements for blacks are related to the war, I re-run eq. (1.4),

$$
\begin{aligned}
y_{izt} = {} & \beta_1 \left(\text{casualty rate}_z \times \text{post-WWII}_t\right) \\
& + \beta_2 \left(\text{casualty rate}_z \times \text{black}_{izt} \times \text{post-WWII}_t\right) \\
& + \alpha_z + \lambda_t + \delta \text{black}_{izt} + X'_{izt}\gamma + \epsilon_{izt}
\end{aligned} \tag{1.5}
$$

with different economic outcomes y_{izt} which are the log of an individual's real annual wage, years of completed education, an indicator for whether they own their home, the log house value, and an indicator for whether a person's state of residence is not their state of birth. Results for the full sample and for the Southern sub-sample are reported in panels A and B in table 1.7, respectively. The corresponding dynamic coefficient plots are shown in figure 1.10 for the full sample and in figure 1.11 for the Southern sample. A downside of the Census data is that not all outcomes were recorded before 1940, such as wages, education, or house values, which were only collected for the first time with the 1940 Census.

The results in table 1.7 show that almost all outcomes for black economic progress in the post-war period considered by prior work are significantly related to the WWII casualty rate among semi-skilled whites. Blacks living in a commuting zone with a 1 p.p. higher casualty rate tend to have 3 to 4 p.p. higher annual wages, a quarter to a third of a year more of completed education, 7 to 9 p.p. higher house values, and they are 1 to 2 p.p. less likely to be living outside their state of birth. With these casualties leading to better employment opportunities for blacks, this decreased the pressure on black workers to leave their state of birth to find better employment elsewhere. The effect of home ownership follows a more complex dynamic response. This is seen in the coefficient plots in figures 1.10 and 1.11 panel (c). The plots show a strong positive initial increase in the home ownership probability in 1950 which then drops in the subsequent decades and becomes negative.

The results on house values, wages, and employment are positive and significant for blacks, irrespective of whether the full sample or the South-only sub-sample is considered. While the wage gains associated with higher casualty rates are higher in the full sample, house values and educational attainment have improved more in the South although the difference to the full sample coefficients are not significantly different. The educational results can potentially be explained in parts with the G.I. Bill which provided subsidies for further education of veterans. However, it would not explain the rise in education levels among Southern blacks who did not benefit from the bill (Turner and Bound, 2003).

Turning to the coefficient plots in in figures 1.10 and 1.11, these show an increase in house values for blacks and a penalty for whites. In terms of house value, blacks gain more in the South, whereas the wage response is slightly larger in the full sample. This might be driven by migration to the North where wages were generally higher and especially high for those who migrate there (Boustan, 2009). The effect on education does not produce a negative or only a weakly negative effect for whites but a strong positive effect on blacks. The initial spike could be explained by the G.I. Bill, whereas the later results, which are weaker but with an increasing trend, can be rationalized by younger cohorts of African Americans. The wartime cohort basically showed that semi-skilled employment is now within reach for blacks, meaning that the benefits of acquiring more education before entering the labor market were more tangible to the newer cohorts. The coefficient plots in figures 1.10 and 1.11 reveal that any negative effect on whites is short-lived and zero otherwise. The wage coefficients display a strong upward trend for blacks, especially in 1970 when the Civil Rights Act of 1964 likely reinforced the wage effect.

1.5 How the Black Occupational Upgrade Affected Black-White Social Relations in the South in 1961

The war elevated African American's economic position by providing them with access to better-paid semi-skilled jobs, especially in the manufacturing sector. During the war, this was not always embraced by white workers. In 1944, the Philadelphia Transportation Company began to alleviate labor shortages by allowing blacks to enter semi-skilled occupations. White workers initiated a strike which was broken when the Army threatened to re-evaluate the draft deferments of striking workers (Collins, 2001). As with the Civil Rights movement, it took some time for whites to adapt to the new workplace realities (see Wright, 2013). What was the longer-term effect of the casualty-induced economic upgrading of blacks on their social status and their relationship with whites?

The answer to this question is not obvious a priori. A well-established concept in the study of network formation is homophily whereby individuals prefer contact with other agents who are more like themselves in terms of age, race, income, and other characteristics (see Currarini, Jackson and Pin, 2009). As the economic position of African Americans improved during and after the war, they became more similar to whites in economic characteristics and therefore their relations may have improved. However, if whites perceived blacks as economic rivals, such as in the case of the Philadelphia Transport Company, the exact opposite could have happened.

To study the above question, I use the "Negro Political Participation Study" (NPPS) of 1961 by Matthews and Prothro (2006). The study was conducted in states of the former Confederacy for a random sample of 540 black and 528 white adults in 1961. For the analysis I coded responses to questions regarding the social integration and status of blacks into binary variables.[15] The outcomes are interracial friendships, living in mixed-race neighborhoods, and attitudes towards integration of respondents and their church ministers. A complete list of the specific questions and the coding scheme for the outcome variables is provided in table 1.8. The summary statistics are reported in table 1.9.

Despite the relatively small sample size, this data set provides a unique opportunity to study the social standing of African Americans in the South before the riots and violence between 1963 and 1970, and before the major legislative and legal reforms against segregation were passed and implemented. Major desegregation laws, such as the Civil Rights Act of 1964, the Voting Rights Act of 1965, the Fair Housing Act of 1968, or Supreme Court rulings such as Loving vs. Virginia 1967, which invalidated anti-miscegenation laws, were only enacted later. The only exception is the Supreme Court case of Brown vs. Board of Education of Topeka in 1954 wherein segregation at public schools was declared unconstitutional. However, it took more than a decade to be fully implemented (Wright, 2013).

Regressing outcomes related to black-white social interaction and attitudes on the share of blacks in semi-skilled occupations as in,

$$\text{social outcome}_{ic} = \beta\Delta\text{share of blacks}_c + \alpha \text{ share of blacks}_{c,1940} + X'_{ic}\delta + \epsilon_{ic} \tag{1.6}$$

where i and c index individuals and counties, respectively, and where social outcomes are the ones described in table 1.8, may not provide unbiased and consistent estimates. A potential issue is reverse causality. The regression in eq. (1.6) assumes that an individual's economic status affects her social status. The opposite might be true when better job opportunities arise from an increase in social contacts. To address this type of endogeneity problem, I instrument the change in the share of blacks in semi-skilled jobs from 1940 to 1950 (Δshare of blacks$_c$) with the WWII casualty rate among semi-skilled whites:

$$\Delta\text{share of blacks}_c = \phi\text{casualty rate}_c + \pi \text{ share of blacks}_{c,1940} + X'_{ic}\gamma + \rho_c \tag{1.7}$$

The casualty rate is defined as before, ρ_c and ϵ_{ic} are stochastic error terms, and X'_{ic} is a vec-

[15] Social integration here refers to any question concerning non-market interactions between blacks and whites, or attitudes towards people from the opposite race.

tor of individual and county level controls as well as state fixed effects. Controlling for the pre-war level of the share of blacks in semi-skilled jobs accounts for cross-county level differences in market-based discrimination before. For a given level of blacks in this skill group, Δshare of blacks$_c$ then provides the additional inflow of blacks into this skill group during the war years. The effect of this inflow might have a different impact when starting from a low or high pre-war level. This simply is a way to leverage the time information on the treatment in cross sectional survey data.

The main assumptions required for identification are that the casualty rate is a sufficiently relevant predictor of Δshare of blacks$_c$ and that it does not correlate with the error term of a given social outcome. A threat to identification would be joint service of blacks and whites in the war. Draft and casualty rates correlate positively. Serving together in battle could have created bonds between black and white soldiers. If those translated to better social relations in the workplace because of their common war experience, this would violate the exclusion restriction. To alleviate such concerns, all regressions control for a respondent's veteran status and the county draft rate.

Further controls for interracial social relations and that might correlate with semi-skilled employment include gender, age, race, the county an individual grew up in, the number of years an individual has spent in their current county of residence, and place size. Additional county level controls include the percentage of blacks, the share of people born in other counties, the WWII draft rate, the number of lynchings between 1900 and 1930, and the number of Rosenwald schools per 1,000 blacks, as well as the number of slaves in 1860.

Another important control is the location of a respondent's dwelling (rural, rural non-farm, suburban, and urban). Boustan (2010, 1961) shows that in-migration of blacks to the centers of Northern cities led whites to move to the periphery. This phenomenon is known in the literature as white flight. If unaccounted for, blacks would find semi-skilled occupations in the city centers and make friends with whites though not because of their improved economic position but because all the whites who had a distaste for interactions with blacks moved to the suburbs. Summary statistics for the individual level controls by race are reported in table 1.10.

A significant shortcoming of this data set is that these individuals cluster in only 24 different counties. This is mainly an inference problem due to the sampling scheme employed. First, primary sampling units (counties or collections of counties) were drawn at random within each Southern state, then individuals were sampled from within a chosen area. The data are therefore representative of the Southern population as argued by Matthews and Prothro (2006). The sample counties are mapped in figure 1.12.

Nevertheless, 24 clusters are not enough for the conventionally used cluster-robust variance-covariance estimator to be consistent as it relies on large sample asymptotics. Cluster-robust standard errors are reported in parentheses for purposes of comparison. The standard errors in squared brackets are estimated via the wild cluster bootstrap t-percentile procedure by Cameron, Gelbach and Miller (2008) for the OLS models, and by combining the same OLS approach with the wild restricted efficient residual bootstrap for IV models by Davidson and MacKinnon (2010). These correct inference for the smaller number of clusters.

OLS and IV results for the regression equation in eq. (1.6) are reported in table 1.11. The sample size is kept constant for all regressions using information from the 540 black and 528 white respondents. The first stage F-statistic on the instrument is sufficiently large with a value of 43.8. I also report the efficient F-statistic by Olea and Pflueger (2013), which is robust to heteroscedasticity and clustering, with a value of 45.8. Most of the IV results are similar to the OLS estimates and show a significant and positive effect of the black skill-upgrade on social relations between blacks and whites. Issues related to omitted variables or selection appear to be less relevant in the context of these outcomes.

A casualty-induced one percentage point increase in Δshare of blacks$_c$ is associated with an 1.8 p.p. increase in a respondent's probability of reporting an interracial friendship. The OLS and IV estimates are virtually the same. An increase in the share of blacks in semi-skilled jobs at the average casualty rate thus increases this probability by 2.9 p.p.[16] Camargo, Stinebrickner and Stinebrickner (2010) show that white students who were randomly assigned a black roommate in their first year of college had a 10.5 p.p. higher probability of having an interracial friendship in the second year. Compared to their estimates, the friendship effect at the average casualty rate is abut 28% of the exposure treatment for college students in the early 2000s. This seems reasonable and puts the magnitude of the estimated coefficients into perspective.

Respondents in treated counties stated with a 1.2 p.p. higher probability that they lived in mixed-race areas. Relative to the outcome mean of 12.4% this is a sizable effect. Given that the share of blacks in the county and dwelling location are controlled for, this is not a mere population composition effect but must have been an active choice by respondents. The black occupational upgrade also had significant effects on attitudes towards integration. Each percentage point increase in Δshare of blacks$_c$ is associated with a 1 p.p. higher probability of respondents favoring integrating in the OLS and 2 p.p. higher in the IV estimation.

[16] Section 1.3.2 estimated an increase in the share of blacks in semiskilled jobs of 0.515 for a 1 p.p. increase in the casualty rate. Since the regression includes fixed effects, this will be similar to a regression in first differences using Δshare of blacks$_c$ as outcome. Hence the friendship effect at an average casualty rate is $3.1 \times 0.515 \times 1.8 = 2.87$.

Breaking this down further, support for integration at school increased by 1 p.p. and by 0.3 (OLS) and 0.8 (IV) p.p. for integration at church. Favoring interracial exposure of their children or in their churches provides significant evidence for the extent of the effects of the improved economic position of blacks on black-white social relations. The results relating to integration at church indicate a willingness to accept the other racial group into the most intimate spheres of social life. Even nowadays there is a strong racial divide in church memberships and service, and Martin Luther King stated in several speeches that 11 o'clock on Sunday is the most segregated hour in American life (see Fryer, 2007). There also appears to be an institutional component since respondents in treated counties were 0.5 to 1.5 p.p. less likely to report their ministers preaching in favor of segregation. However, given the data it is not possible to say whether this was a demand or supply effect. Individuals with higher interracial exposure or contacts might have demanded less segregationist priests, while another possibility is that such priests were predominantly assigned to areas were racial tensions were lower. Overall, the results suggest that the casualty-induced skill-upgrade of African Americans not only came with a rise in economic but also in social status.[17]

1.6 Conclusion

Much has changed since the negative assessment of the economic and social fortunes of African Americans by Myrdal (1944). This is particularly true for the middle of the last century. While writing his book, Myrdal had recognized the importance of the war for the employment of blacks: "The present War is of tremendous importance to the Negro in all respects. He has seen his strategic position strengthened not only because of the desperate scarcity of labor but also because of a revitalization of the American Creed." (1944, p. 409). This paper shows that this scarcity was particularly pronounced in areas with higher WWII casualty rates among semi-skilled whites. These losses opened up new employment opportunities for blacks and contributed to the largest occupational upgrading of African Americans since the end of the Civil War.

Understanding the roots of this unprecedented occupational gain helps to understand African American progress at mid-century. While some path breaking work has assessed black economic progress at mid-century with respect to wages (Margo, 1995; Maloney, 1994; Bailey and Collins, 2006), migration and urbanization (Boustan, 2009, 2010, 1961), home ownership

[17] Appendix C provides further heterogeneity analyses by repeating the estimation for the black and white subsamples, as well as robustness checks with respect to weighting blacks by their population share in the county, changing the definition of the treatment variable, and to assess sensitivity of the IV estimates with respect to mild violations of the exclusion restriction. It also provides a causal mediation analysis to see whether higher incomes for blacks are a mechanism that mediates the effects found in the main analysis.

(Collins and Margo, 2011; Boustan and Margo, 2013; Logan and Parman, 2017), or education (Smith, 1984; Turner and Bound, 2003), our knowledge of the origins of the sudden and strong improvements during and after the war has been limited. The analysis here provides evidence that several of the economic outcomes considered by previous work can be directly related to the war. In particular, they relate to the casualty rate among semi-skilled whites as driver of the black occupational upgrade. I rule out alternative explanations for this pattern based on migration or increased educational attainment by blacks.

The improvements in the position of blacks go beyond the economic gains. The survey data results provide some insights which indicate that areas with a larger wartime upgrading of blacks into semi-skilled employment also saw a rise in their social status. This ranges from increased interracial friendships to higher acceptance of the other group at school or church. The economic upgrading of a minority group thus has the potential to even affect strongly embedded social values in a conservative setting such as the Bible Belt in the early 1960s.

Even though this paper has quantified the relationships between the war casualties and the occupational upgrade, as well as the economic and social outcomes of blacks, it remained mostly silent on the specific mechanisms behind these relationships. The difficulty is to determine which variables are outcomes, treatments, or mediators. Several channels of causation may exist at the same time. The occupational upgrade not only came with better-paying jobs but also with the opportunity to interact more with white workers in the workplace. Is the improvement in social relations driven by inter-group contact at work or by the relaxation of black households' budget constraints that allow for social activities or for moving to better neighborhoods? Exploring these questions might offer a promising avenue for future research.

1.7 Tables

Table 1.1: Summary Statistics - WWII Enlistment Records

	Panel A							
	Black (n = 807,116)				White (n = 7,228,570)			
	mean	st. dev.	min.	max.	mean	st. dev.	min.	max.
Age	25.03	5.80	18	49	24.59	5.69	18	49
Education	9.29	1.86	8	18	10.68	2.24	8	18
AGCT	70.19	19.54	40	187	100.46	22.17	40	199
Married	0.23	0.42	0	1	0.23	0.42	0	1
Height (in.)	68.21	3.51	59	82	68.49	3.25	59	82
Weight (lbs.)	148.42	17.90	94	249	149.59	19.97	88	257
Died	0.019	0.139	0	1	0.029	0.169	0	1
	Panel B							
	Enlisted (n = 1,670,352)				Drafted (n = 6,622,454)			
	mean	st. dev.	min.	max.	mean	st. dev.	min.	max.
Age	22.859	5.155	18	48	25.156	5.809	18	49
Education	11.456	2.148	8	20	10.306	2.244	8	20
AGCT	133.181	27.585	1	199	95.777	22.773	1	199
Married	0.121	0.326	0	1	0.256	0.436	0	1
Height (in.)	68.821	2.839	59	82	68.328	3.414	59	82
Weight (lbs.)	149.056	19.256	90	257	149.311	20.066	88	257
Died	0.027	0.162	0	1	0.029	0.167	0	1
	Panel C							
	South (n = 2,249,203)				Non-South (n = 6,043,984)			
	mean	st. dev.	min.	max.	mean	st. dev.	min.	max.
Age	22.288	5.570	18	46	24.844	5.819	18	49
Education	10.157	2.207	8	20	10.680	2.280	8	20
AGCT	90.722	25.958	1	199	99.825	22.727	1	199
Married	0.252	0.434	0	1	0.220	0.414	0	1
Height (in.)	68.658	2.308	59	82	68.364	3.293	59	82
Weight (lbs.)	148.076	19.501	90	256	149.657	19.989	88	257
Died	0.028	0.166	0	1	0.028	0.166	0	1

Note: Summary statistics for data from drafted soldiers in the Army or Army Air Force between 1940 and 1946. AGCT is the Army General Classification Test, an ability test administered during the draft examinations. This measure is only available for a subset of men drafted in 1943. The similarities in the minimum values for the AGCT, education levels, and height across groups are due to the minimum requirements imposed by the Army on the draft. The indicator for a soldier's death equals one for those who were killed in combat or who died due to all other reasons such as battle and non-battle injuries, accidents, self-inflicted wounds or diseases.

Table 1.2: County Data Summary Statistics, 1920-1970

	obs.	mean	st. dev.	min	max
Main Outcome					
% blacks in semi-skilled jobs	7,737	14.611	14.228	0.000	87.550
% blacks in semi-skilled jobs in 1940	1,386	12.433	12.567	0.000	67.619
Military					
WWII casualty rate of semi-skilled whites	8,303	3.129	2.211	0.000	22.222
Av. casualty rate in neighboring counties	8,286	1.571	1.764	0.000	11.528
Draft rate	8,303	13.143	13.890	0.000	61.592
Log WWII spending per capita	8,303	0.346	1.209	0.000	9.130
Demographics					
Log median family income	5,515	9.780	0.682	7.756	11.469
% with high school degree	5,543	24.440	11.621	3.700	79.500
% rural population	8,299	78.734	24.475	0.000	100.000
% Republican vote share	7,652	14.452	22.562	0.000	100.000
% black population	7,954	22.421	20.706	0.000	90.772
% black male population	8,299	21.341	20.436	0.000	89.893
Lynchings per 1,000 blacks, 1900-30	7,826	0.450	8.607	0.000	500.000
Rosenwald schools per 1,000 blacks	7,826	0.719	1.655	0.000	71.429
% acres flooded by Mississippi, 1928	8,303	0.420	5.015	0.000	100.000
Number of slaves (000s), 1860	8,303	1.377	2.115	0.000	17.957
Agriculture					
% of land in agriculture	8,299	62.198	24.098	0.000	100.000
% acreage in cotton production	8,289	6.050	9.483	0.000	74.414
Share of cash tenants	8,291	7.261	7.915	0.000	78.284
Av. value of machinery per farm (000s)	8,289	2.466	4.758	0.000	219.461
Manufacturing					
Manufact. establishments per 1,000 pop.	7,887	1.240	0.942	0.000	29.728
Av. manufact. firm size	7,461	41.334	39.119	0.000	629.000
Log manufact. value per worker	6,756	12.411	0.956	0.000	14.793
Share of manufact. employment	7,461	5.014	5.329	0.000	100.000
New Deal controls					
New deal loans per capita, 1933-35	8,280	4.562	17.789	0.000	573.874
Relief per capita, 1933-39	8,280	7.613	23.471	0.000	949.111
Public works per capita, 1933-39	8,280	4.868	21.361	0.000	844.372
AAA spending per capita, 1933-39	8,280	5.316	25.560	0.000	852.113
FHA loans insured per capita, 1934-39	8,280	1.124	5.803	0.000	195.790
Unemployment rate, 1937	8,297	10.981	5.831	0.258	42.288

Note: Summary statistics for 1,388 counties in Southern states between 1920 and 1970. Monetary values are deflated to 2010 dollars.

Table 1.3: County Level Difference-in-Differences Results, 1920-1970

	Outcome: % blacks in semi-skilled jobs (pre-war mean = 12.433)					
	(1)	(2)	(3)	(4)	(5)	(6)
Casualty rate$_c \times$ Post-war$_t$	0.515***	0.546***	0.508***	0.548***	0.587***	0.636***
	(0.119)	(0.141)	(0.144)	(0.148)	(0.214)	(0.122)
Controls		Yes		Yes	Yes	Yes
1940 controls $\times$ time			Yes			
Flexible state time trends				Yes		
Linear county time trends					Yes	
Doubly-robust selection						Yes
Observations	7,737	5,713	5,692	5,713	5,713	6,429
Counties	1,388	1,320	994	1,320	1,320	1,375
Adj. R^2	0.855	0.877	0.873	0.883	0.915	0.869
Oster's δ	1.273	1.291	1.112	1.486	0.614	1.494

Note: Difference-in-differences regressions of the county-level share of blacks in semi-skilled occupations on the WWII county casualty rate among semi-skilled whites interacted with a post-war indicator. The estimation sample uses decennial U.S. Census data on counties in Southern states from 1920 to 1970. Controls include county and decade fixed effects, the county draft rate, average casualty rate in the neighboring counties, log WWII spending per capita, share of black men, share of rural population, no. of manufacturing establishments per capita, average manufacturing firm size, log manufacturing value added per worker, share of employment in manufacturing, share of land in agricultural production, share of acres in cotton production, share of cash tenants, average value of machinery per farm, lynchings per 1,000 blacks between 1900 and 1930, no. of Rosenwald schools per 1,000 blacks, share of acres flooded by the Mississippi in 1928, no. of slaves in 1860, Republican vote share, New Deal spending per capita 1933-35 (loans, public works, AAA, FHA loans), and the unemployment rate in 1937. Time-invariant controls are interacted with decade fixed effects. Monetary values are deflated to 2010 U.S. dollars. The doubly-robust selection method implements the Belloni et al. (2014) machine learning covariate selection algorithm for testing the stability of treatment effects with respect to the observables. Oster's (2017) test for selection on unobservables is reported in the final row by computing the coefficient of proportionality δ for which the coefficient on the semi-skilled casualty rate among whites would equal zero. Standard errors clustered at the county level. Significance levels are denoted by * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 1.4: Difference-in-Differences with Casualty Rates by Ethnicity and Skill-Group

	Outcome: % blacks in semi-skilled jobs (pre-war mean = 12.433)					
	(1)	(2)	(3)	(4)	(5)	(6)
White Casualty Rates$_c$ $\times$ Post-war$_t$						
Low-skilled	-0.029	-0.053	0.071	-0.042	-0.196	-0.052
	(0.134)	(0.203)	(0.154)	(0.195)	(0.301)	(0.173)
Semi-skilled	0.557***	0.619***	0.452***	0.585***	0.646***	0.612***
	(0.134)	(0.161)	(0.161)	(0.167)	(0.237)	(0.148)
High-skilled	-0.093	-0.138	0.027	-0.161	-0.220	-0.090
	(0.169)	(0.193)	(0.190)	(0.194)	(0.341)	(0.187)
Black Casualty Rates$_c$ $\times$ Post-war$_t$						
Low-skilled	-0.085**	-0.140**	-0.086*	-0.115*	-0.132	-0.154***
	(0.041)	(0.056)	(0.048)	(0.060)	(0.083)	(0.058)
Semi-skilled	0.057	0.003	0.055	0.014	0.093	-0.011
	(0.054)	(0.057)	(0.054)	(0.047)	(0.093)	(0.055)
High-skilled	-0.051	-0.066	0.008	-0.046	0.008	-0.074
	(0.045)	(0.067)	(0.068)	(0.067)	(0.116)	(0.069)
Controls		Yes		Yes	Yes	Yes
1940 controls $\times$ time			Yes			
Flexible state time trends				Yes		
Linear county time trends					Yes	
Doubly-robust selection						Yes
Observations	7,737	5,713	5,692	5,713	5,713	5,634
Counties	1,388	1,320	994	1,320	1,320	1,299
Adj. R^2	0.855	0.879	0.883	0.884	0.915	0.878
Oster's δ	1.119	1.182	0.833	1.251	0.299	1.152

Note: Difference-in-differences regressions of the county-level share of blacks in semi-skilled occupations on the WWII county casualty rate by race and skill group interacted with a post-war indicator. The estimation sample uses decennial U.S. Census data on counties in Southern states from 1920 to 1970. Controls include county and decade fixed effects, the county draft rate, draft share of each race and skill group, average casualty rate in the neighboring counties, log WWII spending per capita, share of black men, share of rural population, no. of manufacturing establishments per capita, average manufacturing firm size, log manufacturing value added per worker, share of employment in manufacturing, share of land in agricultural production, share of acres in cotton production, share of cash tenants, average value of machinery per farm, lynchings per 1,000 blacks between 1900 and 1930, no. of Rosenwald schools per 1,000 blacks, share of acres flooded by the Mississippi in 1928, no. of slaves in 1860, Republican vote share, New Deal spending per capita 1933-35 (loans, public works, AAA, FHA loans), and the unemployment rate in 1937. Time-invariant controls are interacted with decade fixed effects. Monetary values are deflated to 2010 U.S. dollars. The doubly-robust selection method implements the Belloni et al. (2014) machine learning covariate selection algorithm for testing the stability of treatment effects with respect to the observables. Oster's (2017) test for selection on unobservables is reported in the final row by computing the coefficient of proportionality δ for which the coefficient on the semi-skilled casualty rate among whites would equal zero. Standard errors clustered at the county level. Significance levels are denoted by * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 1.5: Micro Census Triple Differences Results, 1920-1970

	Outcome: $\Pr(\text{semi-skilled}_{izt}) = 1$					
	(1)	(2)	(3)	(4)	(5)	(6)
	Panel A: All U.S.					
Casualty rate$_z \times$ Post-war$_t$	-0.002	-0.002	-0.002	-0.004	-0.003	-0.009**
	(0.007)	(0.007)	(0.004)	(0.004)	(0.005)	(0.004)
Casualty rate$_z \times$ Black$_{izt} \times$ Post-war$_t$	0.047***	0.047***	0.043***	0.019***	0.043***	0.043***
	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)
Observations	4,348,026	4,348,026	4,335,873	3,119,300	4,335,873	4,335,873
Adj. R^2	0.031	0.042	0.044	0.135	0.046	0.047
	Panel B: South only					
Casualty rate$_z \times$ Post-war$_t$	-0.012	-0.013	0.005	-0.009	-0.011	-0.011*
	(0.010)	(0.009)	(0.007)	(0.006)	(0.008)	(0.006)
Casualty rate$_z \times$ Black$_{izt} \times$ Post-war$_t$	0.029***	0.030***	0.028***	0.011***	0.028***	0.028***
	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)
Observations	1,272,016	1,272,016	1,269,553	911,418	1,269,553	1,269,553
Adj. R^2	0.061	0.073	0.075	0.140	0.077	0.080
Individual controls		Yes	Yes	Yes	Yes	Yes
Commuting Zone controls			Yes	Yes	Yes	Yes
Migration and education				Yes		
State time trends					Yes	
Commuting zone time trends						Yes

Note: Difference-in-differenece-in-differences regression of a semi-skilled indicator on the commuting zone WWII casualty rate among semi-skilled whites interacted with a post-WWII dummy, and with a black indicator for individuals living in 722 commuting zones in the whole U.S. and 300 commuting zones in the South. The estimation sample contains data from the decennial U.S. micro Census from 1920-70 on non-institutionalized, working black and white males aged 15-65 who are not currently attending school. All regressions include commuting zone and Census year fixed effects. Individual level controls include age, marital status, age and place of birth dummies. Column (4) adds cross-state migration and education controls interacted with race and time fixed effects. Commuting zone level controls are the WWII draft rate, log WWII spending per capita, share of black men, share of rural population, no. of manufacturing establishments per capita, average manufacturing firm size, log manufacturing value added per worker, share of employment in manufacturing, share of land in agricultural production, share of acres in cotton production, share of cash tenants, average value of machinery per farm, lynchings per 1,000 blacks between 1900 and 1930, no. of Rosenwald schools per 1,000 blacks, share of acres flooded by the Mississippi in 1928, no. of slaves in 1860, Republican vote share, New Deal spending per capita 1933-35 (loans, public works, AAA, FHA loans), and the unemployment rate in 1937. Time-invariant controls are interacted with decade fixed effects. Monetary values are deflated to 2010 U.S. dollars. Standard errors clustered at the commuting zone level in parentheses. Significance levels are denoted by * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 1.6: Triple Differences Results by Industry, 1920-1970

Outcome: $\Pr(\text{semi-skilled}_{izt}) = 1$			
	Manufacturing		
	All	Durable	Non-Durable
	(1)	(2)	(3)
Casualty rate$_z \times$ Post-war$_t$	-0.004	-0.006	0.016
	(0.007)	(0.006)	(0.012)
Casualty rate$_z \times$ Black$_{izt} \times$ Post-war$_t$	0.097***	0.086***	0.105***
	(0.005)	(0.004)	(0.006)
Observations	1,378,824	519,224	860,182
Adj. R^2	0.038	0.040	0.042
	Comparison Sectors		
	Telecom.	Retail	Public Admin.
	(1)	(2)	(3)
Casualty rate$_z \times$ Post-war$_t$	-0.003	0.000	0.002
	(0.014)	(0.004)	(0.011)
Casualty rate$_z \times$ Black$_{izt} \times$ Post-war$_t$	0.024	-0.008***	0.001
	(0.016)	(0.003)	(0.006)
Observations	39,510	469,259	361,325
Adj. R^2	0.095	0.027	0.359

Note: Difference-in-differenece-in-differences regression of a semi-skilled indicator on the commuting zone WWII casualty rate among semi-skilled whites interacted with a post-WWII dummy, and with a black indicator. The estimation sample contains data from the decennial U.S. micro Census from 1920-70 on non-institutionalized, working black and white males aged 15-65. Regression results for semi-skill (columns 1-3) and high-skill (columns 4-6) intensive sectors. All regressions include commuting zone and Census year fixed effects. Individual level controls include age, marital status, age and place of birth dummies. Commuting zone level controls are the WWII draft rate, log WWII spending per capita, share of black men, share of rural population, no. of manufacturing establishments per capita, average manufacturing firm size, log manufacturing value added per worker, share of employment in manufacturing, share of land in agricultural production, share of acres in cotton production, share of cash tenants, average value of machinery per farm, lynchings per 1,000 blacks between 1900 and 1930, no. of Rosenwald schools per 1,000 blacks, share of acres flooded by the Mississippi in 1928, no. of slaves in 1860, Republican vote share, New Deal spending per capita 1933-35 (loans, public works, AAA, FHA loans), and the unemployment rate in 1937. Time-invariant controls are interacted with decade fixed effects. Monetary values are deflated to 2010 U.S. dollars. Standard errors clustered at the commuting zone level in parentheses. Significance levels are denoted by * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 1.7: WWII Casualties and Blacks' Economic Outcomes

Outcome:	ln(wage)	Education	Owns home	ln(house val.)	Migrant
	Panel A: All U.S.				
Casualty rate$_z \times$ Post-war$_t$	-0.018**	-0.049*	-0.007**	-0.042**	0.010
	(0.008)	(0.026)	(0.004)	(0.019)	(0.011)
Casualty rate$_z \times$ Black$_{izt} \times$ Post-war$_t$	0.039***	0.266***	0.000	0.071***	-0.022***
	(0.005)	(0.030)	(0.003)	(0.012)	(0.005)
Observations	2,696,784	3,119,306	4,211,898	1,527,493	4,335,995
Adj. R^2	0.501	0.430	0.251	0.472	0.323
	Panel B: South Only				
Casualty rate$_z \times$ Post-war$_t$	-0.037***	-0.077*	-0.002	-0.051**	-0.005
	(0.012)	(0.039)	(0.005)	(0.024)	(0.007)
Casualty rate$_z \times$ Black$_{izt} \times$ Post-war$_t$	0.032***	0.311***	-0.008***	0.094***	-0.013***
	(0.007)	(0.028)	(0.002)	(0.012)	(0.003)
Observations	766,766	910,755	1,226,713	428,483	1,268,890
Adj. R^2	0.504	0.431	0.241	0.495	0.468

Note: Difference-in-differenece-in-differences regression of economic outcomes on the commuting zone WWII casualty rate among semi-skilled whites interacted with a post-WWII dummy, and with a black indicator for individuals living in 722 commuting zones in the whole U.S. The estimation sample contains data from the decennial U.S. micro Census from 1920-70 on non-institutionalized, working black and white males aged 15-65 who are not currently attending school. All regressions include commuting zone and Census year fixed effects. Owns home is a binary outcomes for whether an individual owns their home. The log house value, log wages, and education variables are only available from 1940 onward. Log house value is also missing for 1950. Individual level controls include age, marital status, age and place of birth dummies. Commuting zone level controls are the WWII draft rate, log WWII spending per capita, share of black men, share of rural population, no. of manufacturing establishments per capita, average manufacturing firm size, log manufacturing value added per worker, share of employment in manufacturing, share of land in agricultural production, share of acres in cotton production, share of cash tenants, average value of machinery per farm, lynchings per 1,000 blacks between 1900 and 1930, no. of Rosenwald schools per 1,000 blacks, share of acres flooded by the Mississippi in 1928, no. of slaves in 1860, Republican vote share, New Deal spending per capita 1933-35 (loans, public works, AAA, FHA loans), and the unemployment rate in 1937. Time-invariant controls are interacted with decade fixed effects. Monetary values are deflated to 2010 U.S. dollars. Standard errors clustered at the commuting zone level in parentheses. Significance levels are denoted by * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 1.8: Interview Questions and Outcome Coding Scheme

- **Interracial Friend**: (Var 0377)
  "Have you ever known a white (colored) person well enough that you would talk to him as a friend?"
  Coded 1 for 1 (Yes), and 0 otherwise.
- **Live in Mixed Area**: (Var 0079)
  "Racial composition of residential area of respondent"
  Coded 1 for value 3 (Mixed).
- **Favor Integration**: (Var 0374) "Are you in favor of integration, strict segregation, or something in between?"
  Coded 1 for 2 (Integration), and 0 otherwise.
- **Favor Mixed Churches**: (Var 0397)
  "Inter-racial contact: churches - Respondent favors:"
  Coded 1 for values 4 (Gradual integration), 5 (Rapid integration) and 6 (Mixed), and 0 otherwise.
- **Favor Mixed Schools**: (Var 0396)
  "Inter-racial contact: schools - Respondent favors:"
  Coded 1 for values 4 (Gradual integration), 5 (Rapid integration) and 6 (Mixed), and 0 otherwise.
- **Priest Pro Segregation**: (Var 0164)
  "Would you say that your minister believes that religion or the Bible favors segregation or integration?"
  Coded 1 for 1 (Favors segregation) and 2 (Qualified favors segregation), and 0 otherwise.

Note: Original questions from the 1961 "Negro Political Participation Study" (Matthews and Prothro, 1975) and the definitions of the outcome variables which are coded from the corresponding questions as binary variables. Outcomes are in bold font, questionnaire variable numbers are reported in parentheses, questions from the survey between in quotation marks, followed by the coding scheme for the binary variables. The code book for ICPSR study number 7255 is freely available at: http://www.icpsr.umich.edu/icpsrweb/ICPSR/studies/7255

Table 1.9: Summary Statistics - Outcome Variables by Race

	Black (n = 540)		White (n = 528)		Difference	
	mean	st. dev.	mean	st. dev.	diff.	s.e.
Interracial Friend	0.466	0.499	0.583	0.494	0.117***	0.030
Live in Mixed Area	0.161	0.368	0.085	0.279	-0.076***	0.020
Favor Integration	0.641	0.480	0.036	0.186	-0.605***	0.022
Favor Mixed Churches	0.057	0.233	0.011	0.106	-0.046***	0.011
Favor Mixed Schools	0.059	0.236	0.045	0.208	-0.014	0.014
Priest Pro Segregation	0.061	0.240	0.142	0.349	0.081***	0.018

Note: Binary outcomes of the social and political integration, standing and attitudes of blacks for black and white respondents in the "Negro Political Participation Study" of 1961 (Matthews and Prothro, 1975). Only individuals in the final estimation sample were used to produce these summary statistics. Differences in means and the corresponding standard errors were estimated with t-tests. Significance levels at 10%, 5%, and 1% are denoted by *, **, ***, respectively. The question about repercussions for political activity against blacks were only asked to African American respondents.

Table 1.10: Summary Statistics - Individual Characteristics by Race

	Black (n = 540)			
	mean	st. dev.	min.	max.
Male	0.382	0.486	0	1
Age	46.319	15.883	5	85
Years of education	4.952	3.248	1	14
Family income	2183.078	1864.756	500	11000
Veteran	0.124	0.330	0	1
Years in county	35.050	19.425	0	89
% blacks in birth county	43.222	16.309	5	85
Rural	0.205	0.404	0	1
Rural, non-farm	0.069	0.253	0	1
Suburban	0.117	0.321	0	1
City/town	0.610	0.488	0	1
	White (n = 528)			
	mean	st. dev.	min.	max.
Male	0.450	0.498	0	1
Age	45.669	15.684	5	89
Years of education	7.323	3.637	1	14
Family income	4929.061	3178.278	500	11000
Veteran	0.237	0.426	0	1
Years in county	29.638	21.130	0	83
% blacks in birth county	24.452	17.935	5	85
Rural	0.227	0.419	0	1
Rural, non-farm	0.114	0.318	0	1
Suburban	0.131	0.338	0	1
City/town	0.528	0.500	0	1

Note: Summary statistics for black and white respondents from the "Negro Political Participation Study" of 1961 by Matthews and Prothro (1975). Statistics produced for individuals from the final estimation sample. Family income is coded in income bins while for the summary statistics the midpoint of each interval was recorded as the dollar values for the corresponding bin.

Table 1.11: The Skill Upgrade and Black-White Social Relations - OLS and IV Results

	Pr(Interracial Friend)=1		Pr(Live in Mixed Race Area)=1	
	(OLS)	(IV)	(OLS)	(IV)
Δsemi-skilled blacks$_c$	0.0181	0.0180	0.0155	0.0118
	$(0.0059)^{***}$	$(0.0075)^{**}$	$(0.0046)^{***}$	$(0.0046)^{***}$
	$[0.0079]^{**}$	$[0.0103]^{*}$	$[0.0062]^{**}$	$[0.0075]$
Outcome mean	0.5235	0.5235	0.1236	0.1236
R^2	0.1213	0.1213	0.1406	0.1402
	Pr(Favor Integration)=1		Pr(Favor Mixed Schools)=1	
	(OLS)	(IV)	(OLS)	(IV)
Δsemi-skilled blacks$_c$	0.0097	0.0211	0.0105	0.0104
	$(0.0031)^{***}$	$(0.0062)^{***}$	$(0.0021)^{***}$	$(0.0032)^{***}$
	$[0.0053]^{*}$	$[0.0123]^{*}$	$[0.0039]^{***}$	$[0.0047]^{**}$
Outcome mean	0.3418	0.3418	0.0524	0.0524
R^2	0.5097	0.5079	0.0683	0.0683
	Pr(Favor Mixed Church)=1		Pr(Priest Pro Segregation)=1	
	(OLS)	(IV)	(OLS)	(IV)
Δsemi-skilled blacks$_c$	0.0027	0.0075	-0.0051	-0.0146
	$(0.0015)^{*}$	$(0.0021)^{***}$	(0.0039)	$(0.0069)^{**}$
	$[0.0021]$	$[0.0033]^{**}$	$[0.0052]$	$[0.0104]$
Outcome mean	0.0346	0.0346	0.1011	0.1011
R^2	0.0801	0.0780	0.1191	0.1160

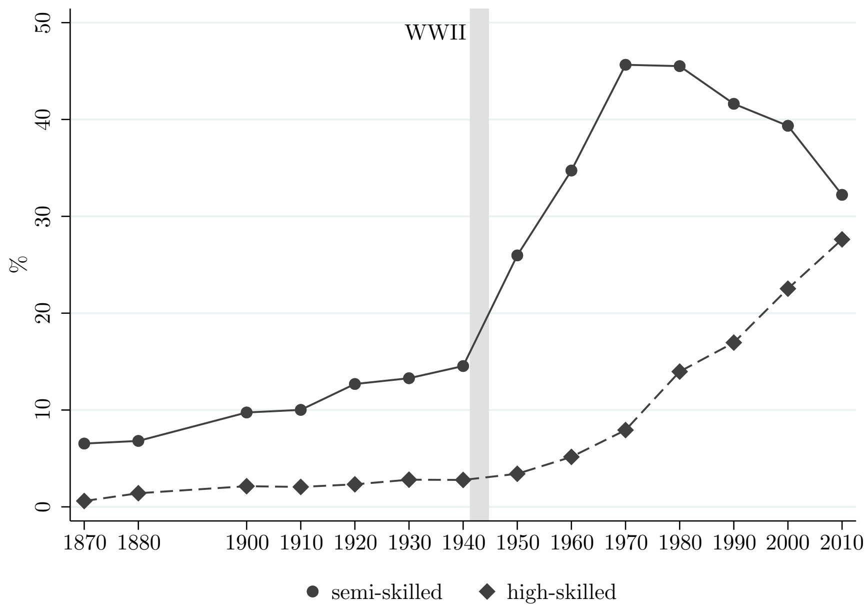
Note: The estimation sample is kept constant in all regressions with 540 black and 528 white adults in 24 counties from Southern states in 1961 using data from the "Negro Political Participation Study" (Matthews and Prothro, 1975). The change in the share of blacks in semi-skilled employment from 1940 to 1950 (Δshare of blacks$_c$) in county c is instrumented with the WWII casualty rate among semi-skilled whites in that county. The first stage F-statistic is 43.799 and the Olea and Pflueger (2013) efficient F-statistic is 45.841. Individual level controls include gender, race, age, location of dwelling (urban, suburban, rural), years lived in current county, place size, veteran status, county where a respondent grew up, and state fixed effects. County level controls used are the share of blacks in semi-skilled jobs in 1940, the share of blacks in county c, share of people not born in county c, the WWII draft rate, and variables on racial sentiment such as the number of Rosenwald schools per 1,000 blacks, the number of lynchings from 1900-30 per 1,000 blacks, and the number of black slaves in 1860. Standard errors are clustered at the county level and are reported in parentheses. Standard errors corrected for the small cluster size using the wild cluster bootstrap-t procedure for OLS models by Cameron et al. (2008) and the wild restricted efficient residual bootstrap for IV models by Davidson and MacKinnon (2010) are reported in squared brackets. Significance levels are denoted by * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

1.8 Figures

Figure 1.1: Share of Semi- and High-Skilled Employment Among Black Men, 1870 to 2010

(a) All U.S.

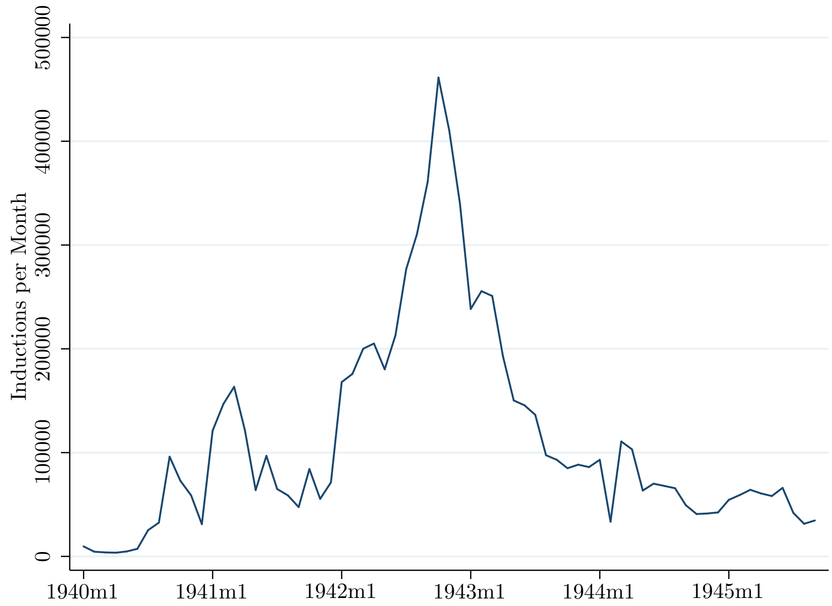
(b) Southern U.S.



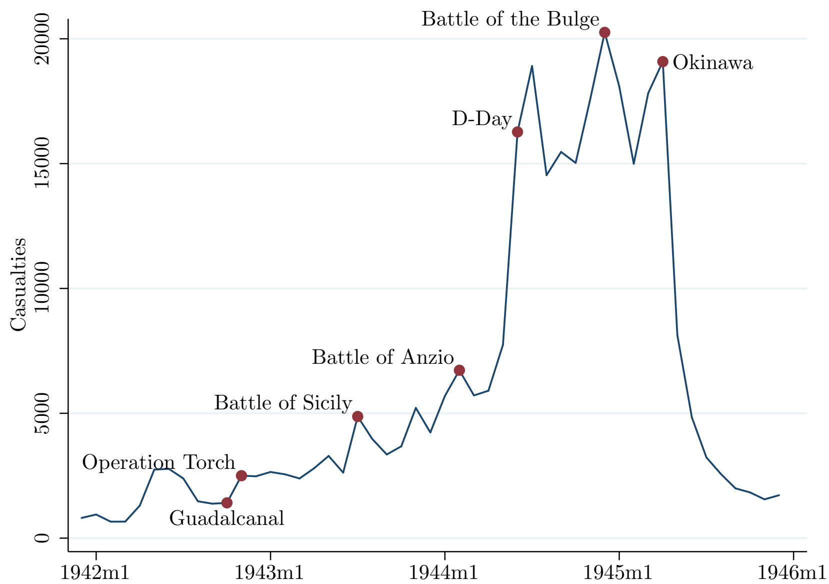
Note: Graphs are based on the public use microdata files of the 1870-2010 Decennial U.S. Censuses by Ruggles et al. (2018). The sample includes black males aged 16 to 65 of the non-institutionalized population who are not attending school at the enumeration date. Semi-skilled jobs (dots) are operatives and craftsmen, and high-skilled jobs (diamonds) are clerks, professionals, and managers. Occupations are defined according to the 1950 Census Bureau occupational classification scheme. The years of U.S. involvement in World War II are marked with light gray background shading. Data for the South includes individuals living in the states of the former Confederacy, as well as Delaware, DC, Kentucky, Maryland, Oklahoma, and West Virginia.

Figure 1.2: Number of Drafted and Fallen Soldiers by Month and Year

(a) Draft Numbers



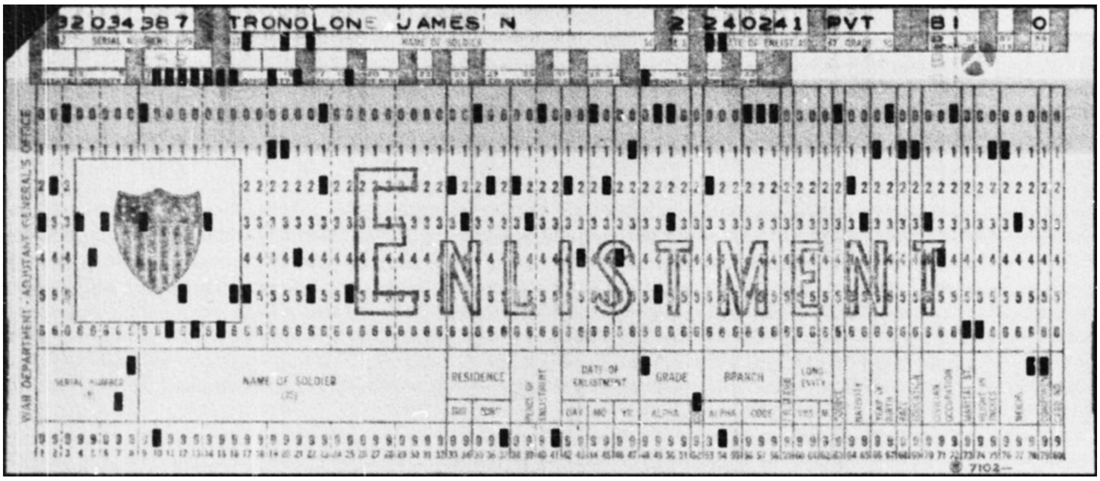
(b) Casualty Numbers



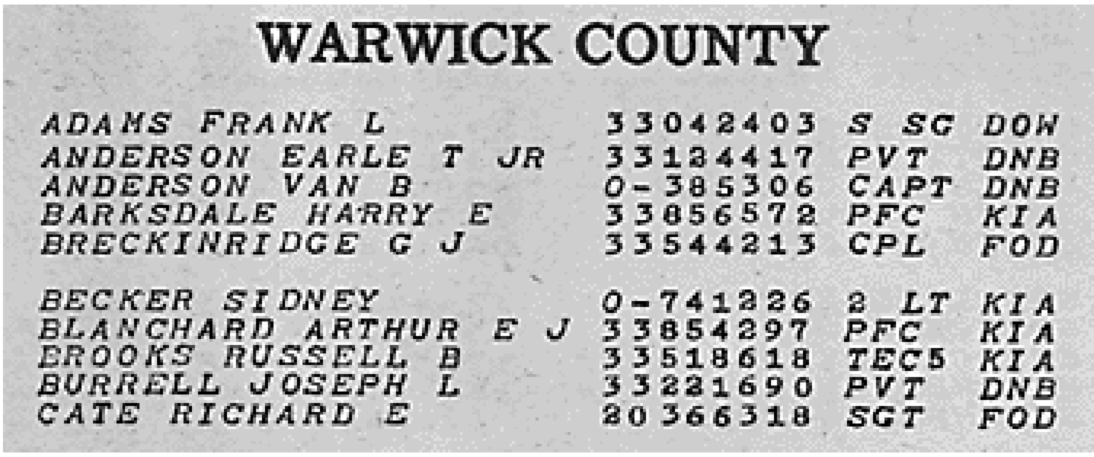
Note: Draft numbers (inductions) also include those who enlisted voluntarily prior to when voluntary enlistment was forbidden in 1942. Both draft and casualty figures are for the Army and Army Air Force only. Panel (b) shows the number of fallen soldiers per month together with major battles and operations involving U.S. Army and Army Air Force personnel. Casualties here refer to all combat and non-combat related deaths. The draft series begins with the enactment of the WWII draft in 1940 whereas the casualty series begins with the attack on Pearl Harbor. Monthly casualty counts come from the Office of the Adjutant General (1946) “Army Battle Casualties and Nonbattle Deaths in World War II - Final Report”.

Figure 1.3: Draft and Casualty Records Example

(a) IBM Draft Punch Card



(b) WWII Honor List of Dead and Missing

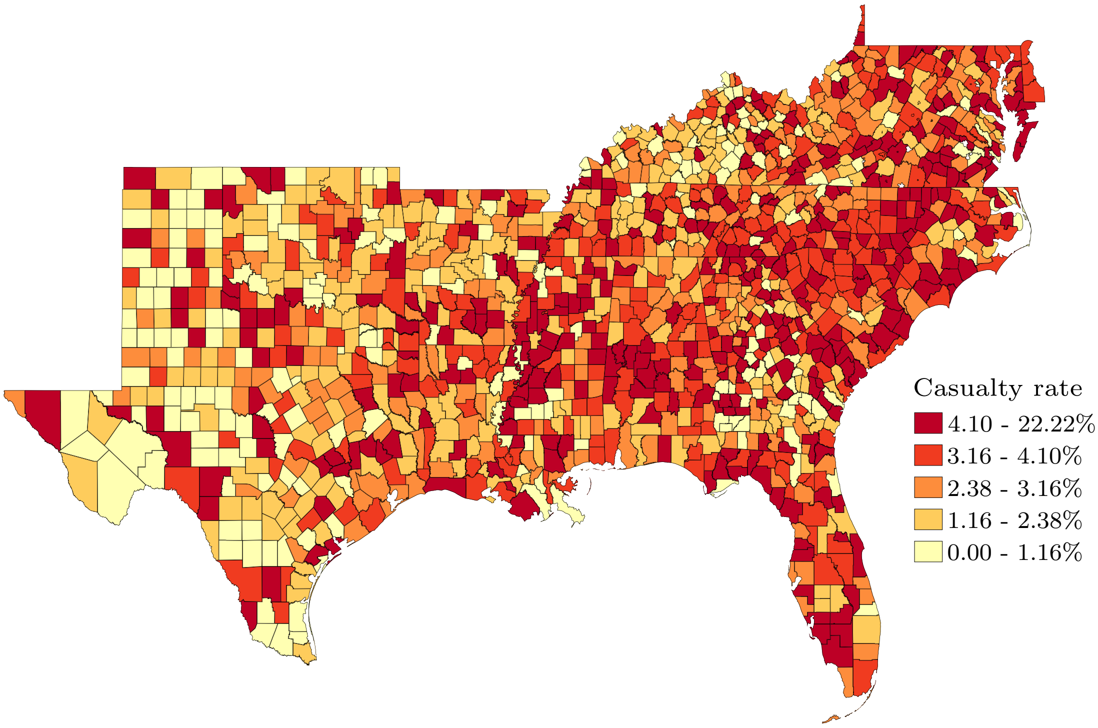


WARWICK COUNTY

ADAMS FRANK L	33042403	S SG	DOW
ANDERSON EARLE T JR	33124417	PVT	DNB
ANDERSON VAN B	O-385306	CAPT	DNB
BARKSDALE HARRY E	33856572	PFC	KIA
BRECKINRIDGE G J	33544213	CPL	FOD
BECKER SIDNEY	O-741226	2 LT	KIA
BLANCHARD ARTHUR E J	33854297	PFC	KIA
BROOKS RUSSELL B	33518618	TEC5	KIA
BURRELL JOSEPH L	33221690	PVT	DNB
CATE RICHARD E	20366318	SGT	FOD

Note: Panel a) shows the enlistment punch card for James Tronolone from Erie, New York, born in 1910. His Army serial number is shown on the top left corner of the card, his rank, date of enlistment, and service branch, among other, on the top right. Panel b) shows an excerpt from the WWII Honor List of Dead and Missing for Warwick County, Virginia. The table displays a soldier's name, their Army serial number, rank, and cause of death. Source: National Archives and Records Administration, Record Group 407: Records of the Adjutant General's Office, 1917- [AGO].

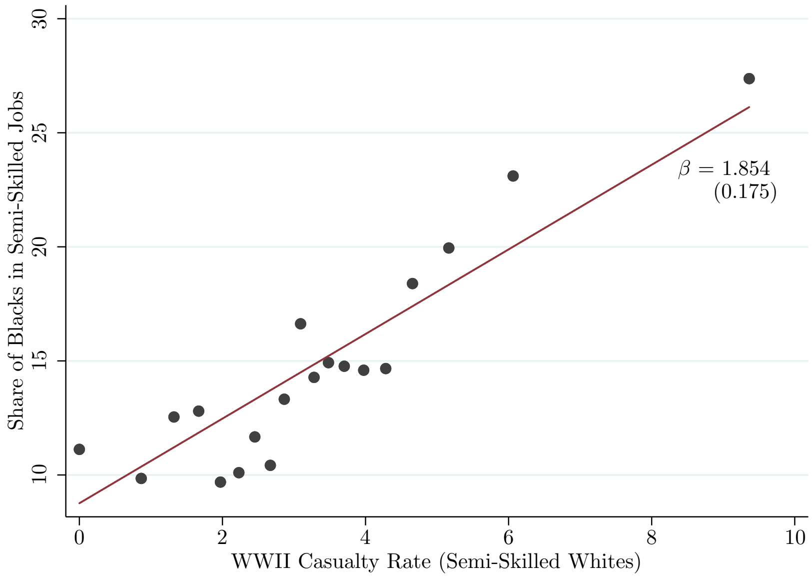
Figure 1.4: WWII Casualty Rates among Semi-Skilled Whites in the U.S. South



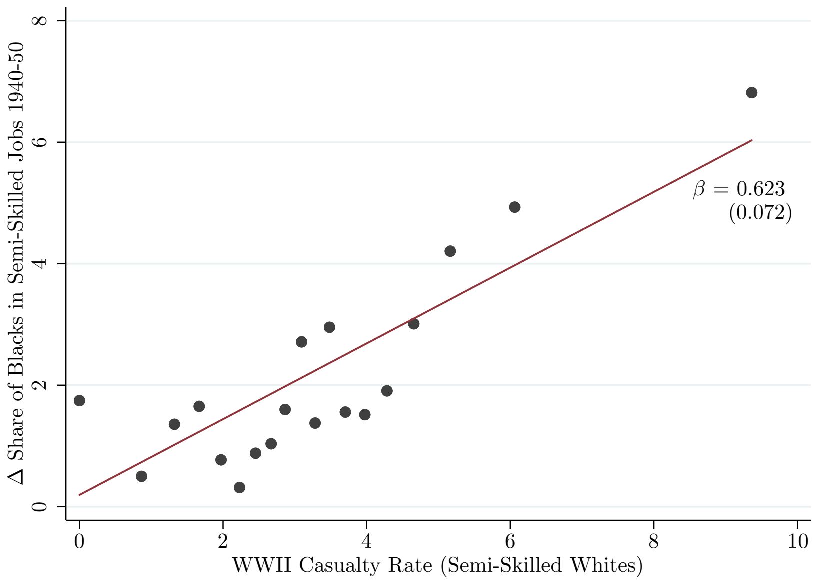
Note: Spatial distribution of WWII casualty rates among semi-skilled white men at the county level in percent. Shaded polygons display the quintiles of the casualty rate distribution with ranges being shown in the legend on the side. Southern states included here are Alabama, Arkansas, Delaware, Florida, Georgia, Kentucky, Louisiana, Maryland, Mississippi, North Carolina, South Carolina, Oklahoma, Tennessee, Texas, Virginia, and West Virginia.

Figure 1.5: Scatter Plots for WWII Casualty Rates and the Share of Blacks in Semi-Skilled Jobs in Levels and First Differences

(a) Correlation with the Semi-Skilled Share Level in 1950

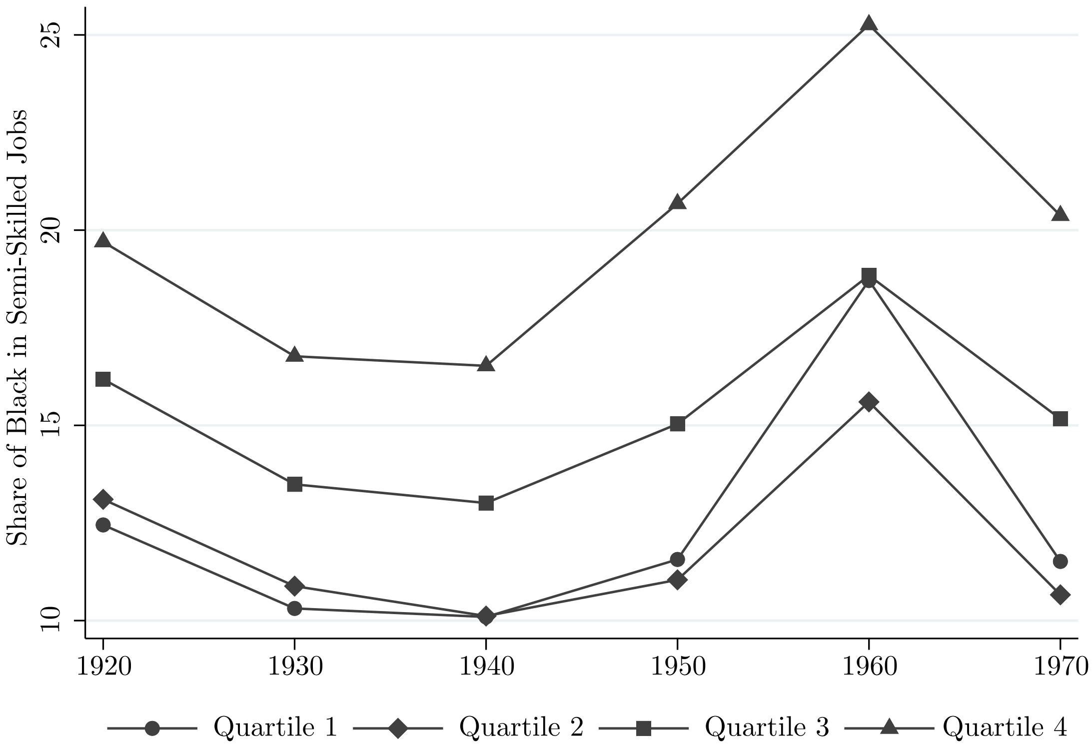


(b) Correlation with the Semi-Skilled Share 1940 to 50 First Difference



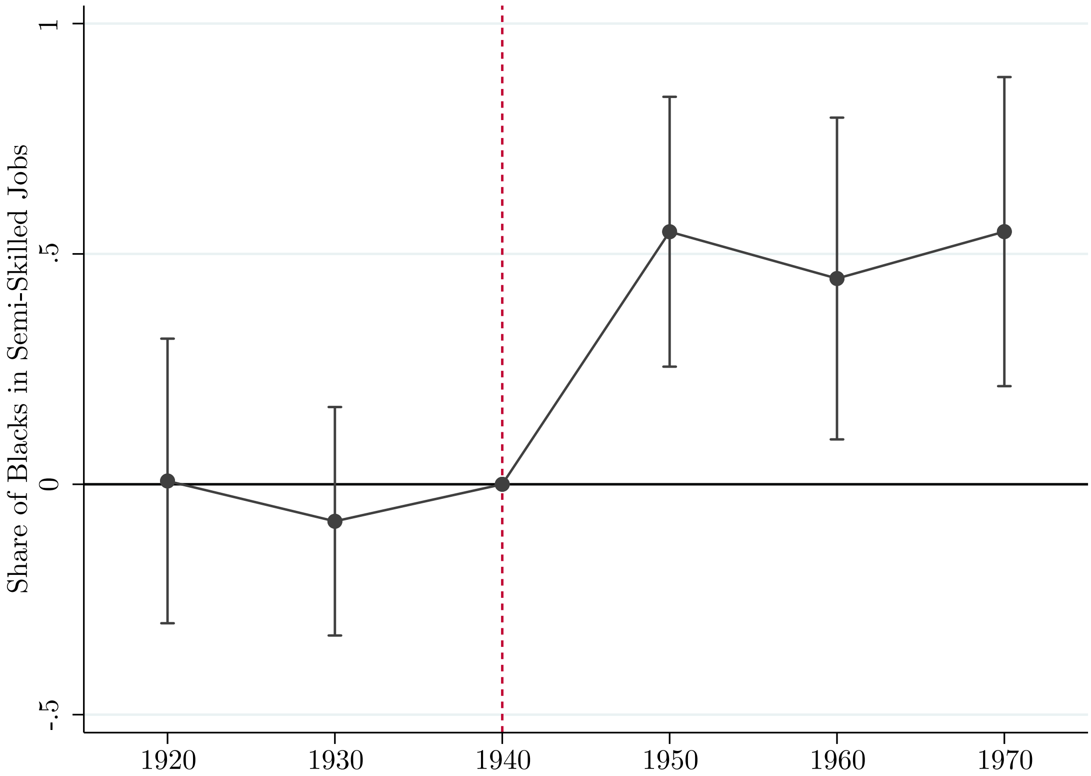
Note: Scatter plots of the relation between the WWII casualty rate among semi-skilled whites and the share of blacks in semi-skilled employment in 1950 across counties (panel a), and the change in the share of blacks in semi-skilled employment from 1940 to 1950 (panel b).

Figure 1.6: Unconditional Share of Blacks in Semi-Skilled Jobs by Casualty Rate Quartile



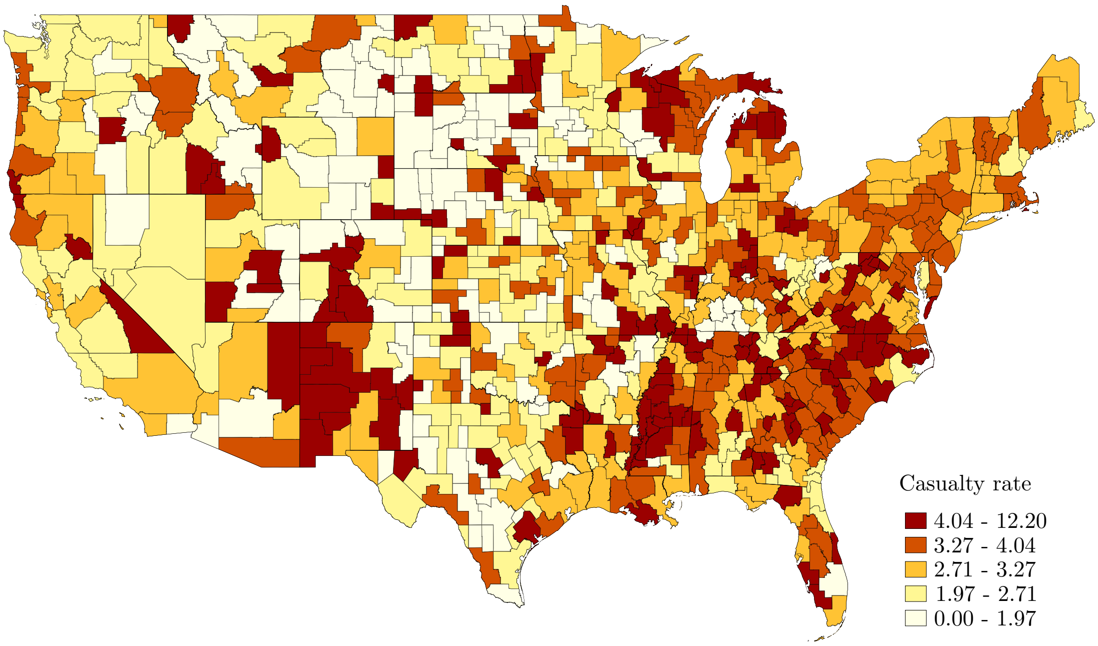
Note: The figure plots the raw outcome data for the share of blacks in semi-skilled jobs for counties in Southern states by quartiles of the WWII casualty rate among semi-skilled whites over time. This shows how the share of blacks in semi-skilled jobs evolved in a parallel fashion for all groups over time before the war. From 1940 to 1950, the increase in the outcome is stronger for higher casualty rate quartiles, after which also the gap between the top and bottom quartiles remains constantly higher.

Figure 1.7: Difference-in-Differences Coefficient Plot



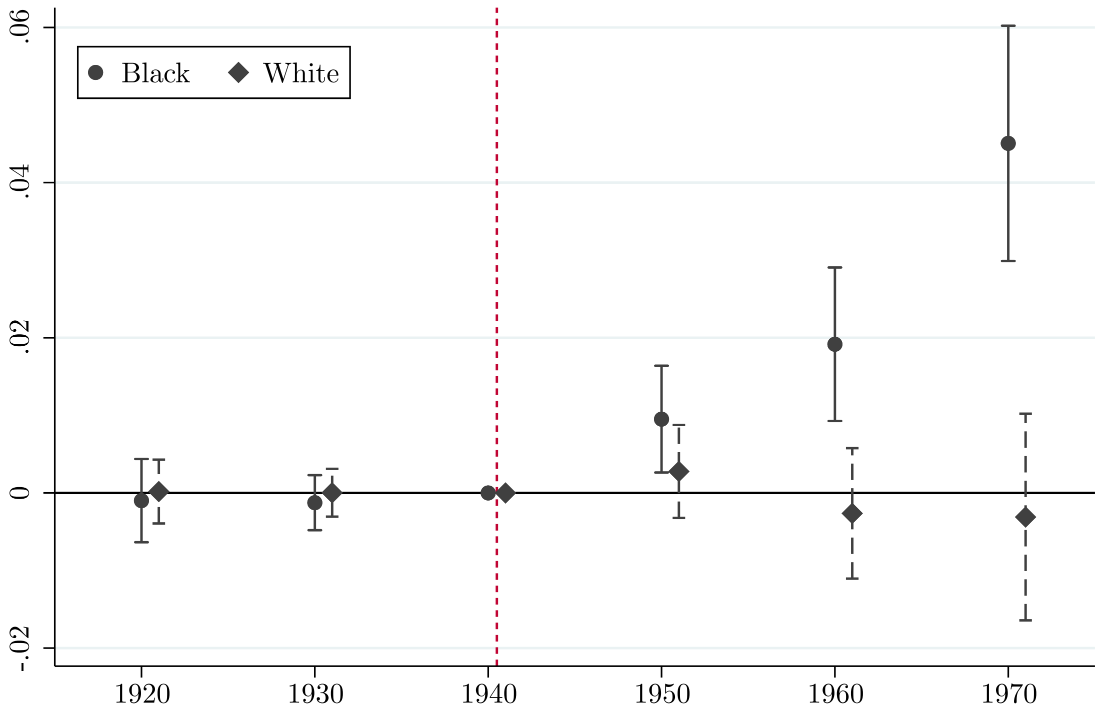
Note: Difference-in-differences regressions of the county-level share of blacks in semi-skilled occupations on the WWII county casualty rate among semi-skilled whites interacted with decade fixed effects. The omitted baseline decade is 1940 which is marked by the dashed line. This is the last pre-treatment period. The estimation sample contains counties in Southern states from 1920 to 1970. Coefficients show the effect of a one standard deviation increase in the casualty rate on the outcome in terms of percentage points. Controls include county fixed effects and flexible state-specific time trends, the county draft rate, average casualty rate in the neighboring counties, log WWII spending per capita, share of black men, share of rural population, no. of manufacturing establishments per capita, average manufacturing firm size, log manufacturing value added per worker, share of employment in manufacturing, share of land in agricultural production, share of acres in cotton production, share of cash tenants, average value of machinery per farm, lynchings per 1,000 blacks between 1900 and 1930, no. of Rosenwald schools per 1,000 blacks, share of acres flooded by the Mississippi in 1928, no. of slaves in 1860, Republican vote share, New Deal spending per capita 1933-35 (loans, public works, AAA, FHA loans), and the unemployment rate in 1937. Time-invariant controls are interacted with decade fixed effects. Monetary values are deflated to 2010 U.S. dollars. Standard errors clustered at the county level. Error bars show 95% confidence intervals around each coefficient estimate.

Figure 1.8: Spatial Distribution of WWII Casualty Rates among Semi-Skilled Whites



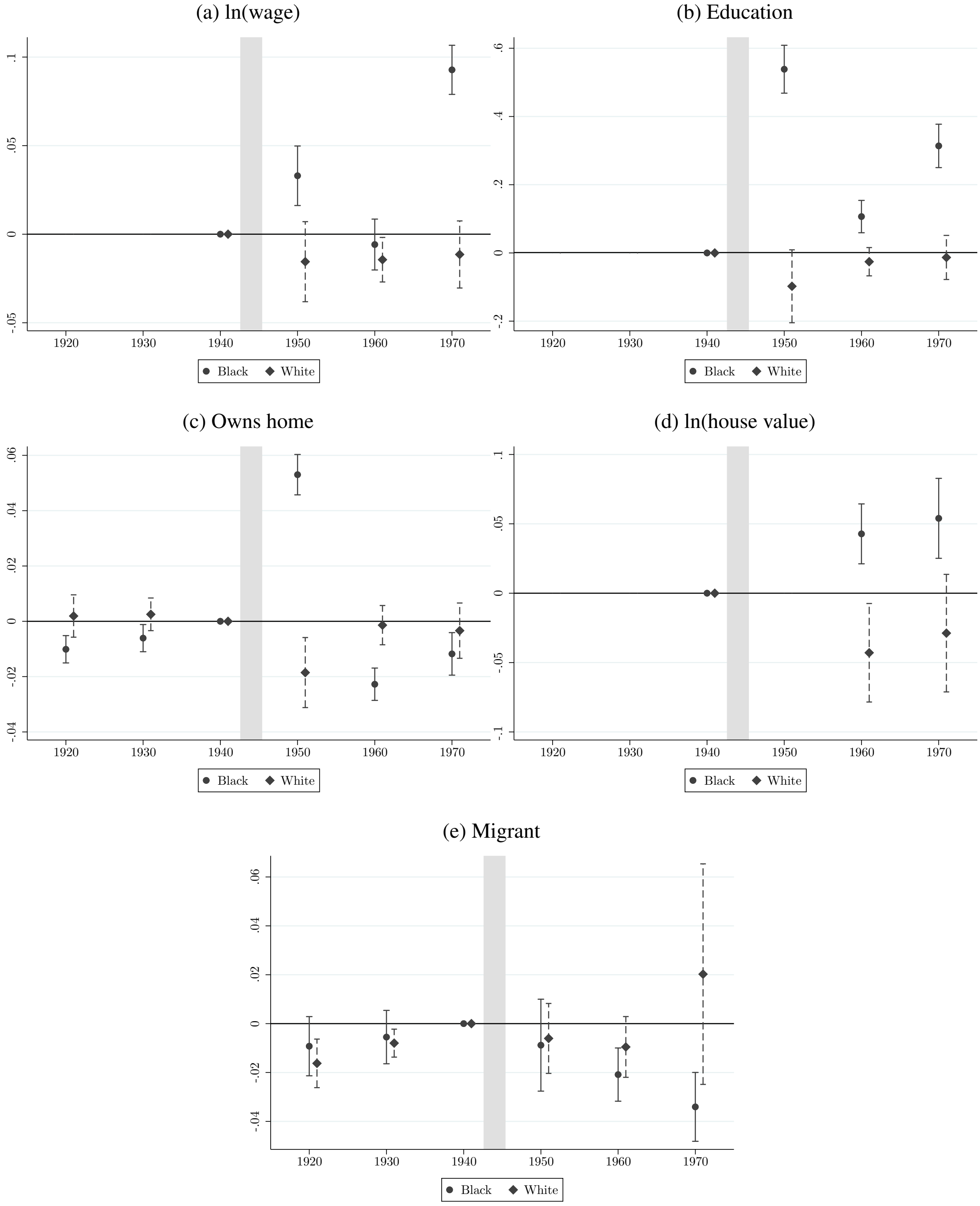
Note: Spatial distribution of WWII casualty rates among semi-skilled white men at the commuting zone level in percent. Shaded polygons display the quintiles of the casualty rate distribution with ranges being shown in the legend on the side.

Figure 1.9: Triple Differences Coefficients Plot



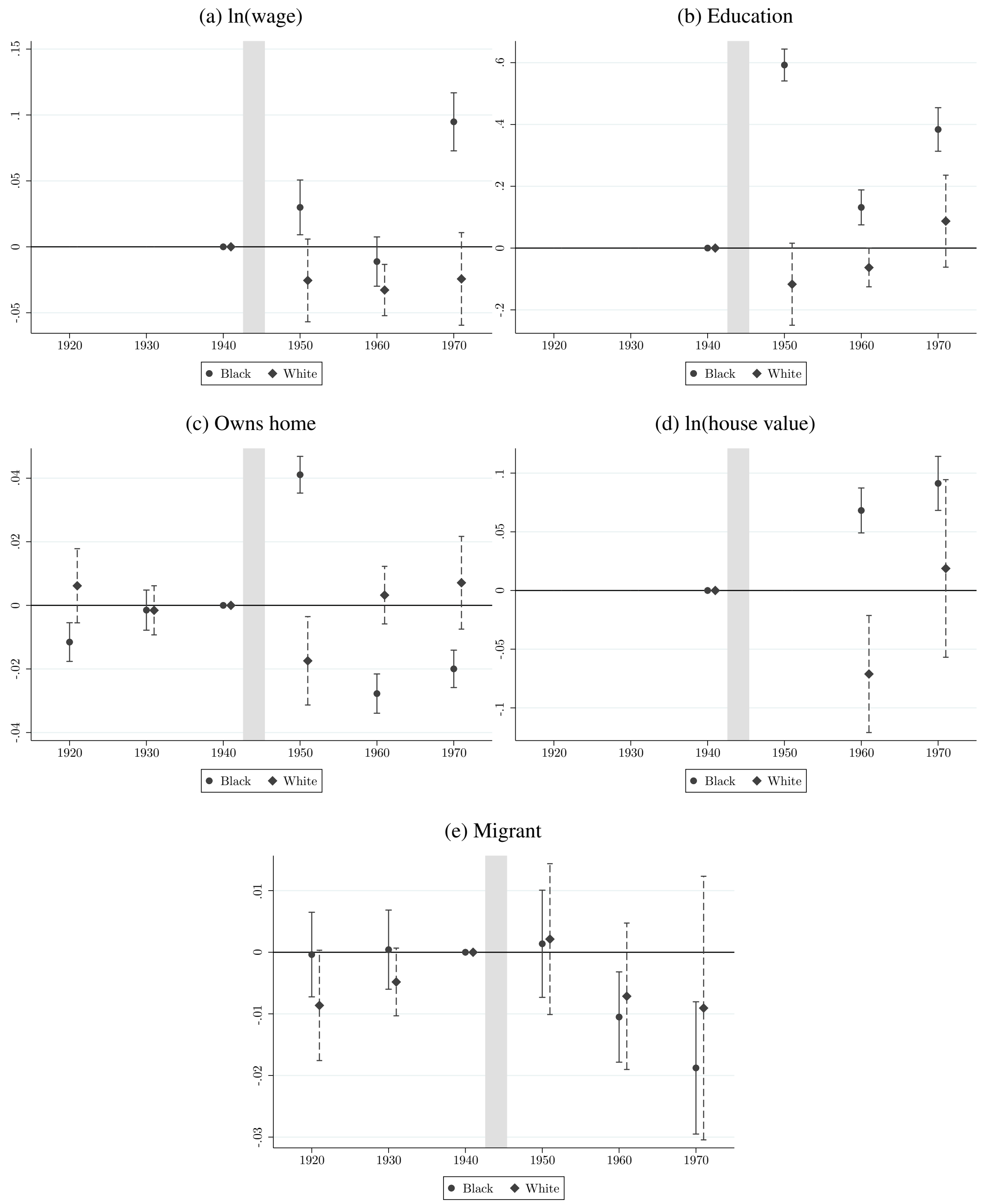
Note: Coefficients plot from a difference-in-difference-in-differences regression of a semi-skilled indicator on the commuting zone WWII casualty rate among semi-skilled whites interacted with decade dummies, and with a black indicator. White coefficients for the interaction of the casualty rate with decade dummies, plotted black coefficients are for the casualty rate interacted with decade dummies and a black indicator. The estimation sample contains data from the decennial U.S. micro Census from 1920-70 on non-institutionalized, working black and white males aged 15-65. All regressions include commuting zone and Census year fixed effects. Controls include age, marital status, year of birth, a self-employment indicator, farm status, and industry fixed effects. The vertical dashed line marks the omitted baseline year of 1940. Standard errors clustered at the commuting zone level. Error bars show 95% confidence intervals around each coefficient estimate.

Figure 1.10: Triple-Differences Coefficient Plots: WWII Casualty Treatment, all U.S.



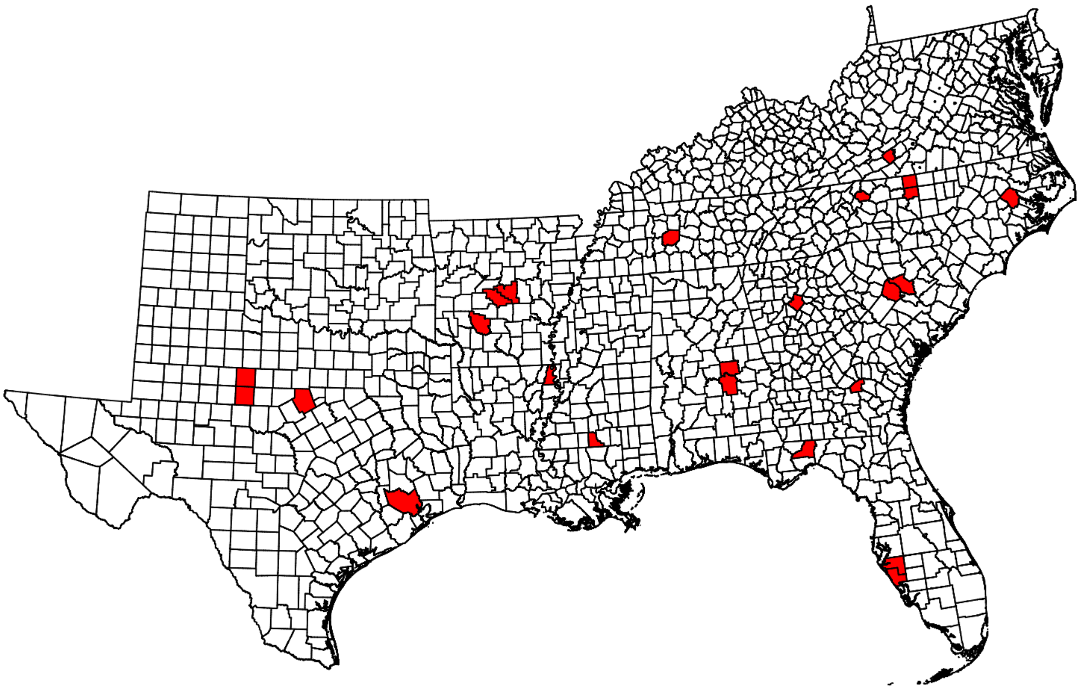
Note: Coefficient plots from the triple differences regression of each of the six outcomes on the the WWII casualty rate × year fixed effects (effect on whites), and WWII casualty rate × year fixed effects × a black indicator (effect for blacks), as well as commuting zone and year fixed effects using individual data from the U.S. Census from 1920-70. The gray area marks years of U.S. involvement in the war. Further controls include the log of WWII spending per capita, the WWII draft rate, share of black men, share of rural population, no. of manufacturing establishments per capita, average manufacturing firm size, log manufacturing value added per worker, share of employment in manufacturing, share of land in agricultural production, share of acres in cotton production, share of cash tenants, average value of machinery per farm, lynchings per 1,000 blacks between 1900 and 1930, no. of Rosenwald schools per 1,000 blacks, share of acres flooded by the Mississippi in 1928, no. of slaves in 1860, Republican vote share, New Deal spending per capita 1933-35 (loans, public works, AAA, FHA loans), and the unemployment rate in 1937. Time-invariant controls are interacted with decade fixed effects. Monetary values are deflated to 2010 U.S. dollars. Error bars show 95% confidence intervals. Standard errors are clustered at the commuting zone level.

Figure 1.11: Triple-Differences Coefficient Plots: WWII Casualty Treatment, South only



Note: Coefficient plots from the triple differences regression of each of the six outcomes on the the WWII casualty rate $\times$ year fixed effects (effect on whites), and WWII casualty rate $\times$ year fixed effects $\times$ a black indicator (effect for blacks), as well as commuting zone and year fixed effects using individual data from the U.S. Census from 1920-70. The gray area marks years of U.S. involvement in the war. The sample includes observations from Southern states only. Further controls include the log of WWII spending per capita, the WWII draft rate, share of black men, share of rural population, no. of manufacturing establishments per capita, average manufacturing firm size, log manufacturing value added per worker, share of employment in manufacturing, share of land in agricultural production, share of acres in cotton production, share of cash tenants, average value of machinery per farm, lynchings per 1,000 blacks between 1900 and 1930, no. of Rosenwald schools per 1,000 blacks, share of acres flooded by the Mississippi in 1928, no. of slaves in 1860, Republican vote share, New Deal spending per capita 1933-35 (loans, public works, AAA, FHA loans), and the unemployment rate in 1937. Time-invariant controls are interacted with decade fixed effects. Monetary values are deflated to 2010 U.S. dollars. Error bars show 95% confidence intervals. Standard errors are clustered at the commuting zone level.

Figure 1.12: Location of NPPS Respondents



Note: Counties included in the "Negro Political Participation Study" by Matthews in Prothro (1975) in 1961. Some states which were chosen for the main analysis are not included in this sample. Matthews and Prothro (1975) only included those states and counties which officially belonged to the former Confederacy. Hence border states such as Kentucky, Maryland, Delaware and West Virginia are not included. Oklahoma was Indian Territory at the time and therefore also was not included in the list of Confederate states belonging to the NPPS sampling scheme.

1.9 Appendix

A Black occupational upgrade

A1) Robustness and Heterogeneity

A1.1: Parallel Trends Assumptions

In addition to the lags and leads of the casualty treatment and their effects on the share of blacks in semi-skilled jobs in figure 1.7, figure 1.A.1 provides the same plot under different model specifications. This includes the model without covariates (i.e. the raw data less time and county fixed effects), with controls, with controls fixed at their 1940 values and interacted with time dummies, and controls selected by the Belloni et al. (2014) algorithm. The insignificance of the pre-trends and the post-war treatment effect do not hinge on any particular model specification but are indistinguishable from the coefficients plot presented in the main section.

A1.2: Selection on Observables

Table 1.A.1 estimates the DiD model in eq. (1.2) and gradually expands the covariate set. Observing the movement of the coefficient of interest shows that the casualty rate coefficient stabilizes at around 0.59 p.p. There is no one particular control which significantly alters the results after being included. The typical argument is that the treatment effect remains stable with respect to the inclusion of observed factors, it would remain stable also with respect to unobserved factors. However, as discussed in the main section with reference to the test by Oster (2017), this is not necessarily true if, for instance, observables and unobservables are unrelated to each other but separately affect the relationship between treatment and outcome.

A downside of the coefficient stability test is that invariance of the top-row coefficient might be due to measurement error in the controls. Following Pei, Pischke and Schwandt (2018), a more powerful alternative is to take the added control to the left-hand side of the equation and test for imbalances with respect to the treatment variable. This is equivalent to running regressions with and without the added control and comparing both estimates via a SUR regression. This is a generalized Hausman test. The corresponding χ^2 test statistics and p-values are reported in the bottom two rows of table 1.A.1. The test reveals no significant imbalances in the controls which are related to the casualty rate.

A1.3: Selective Migration of Blacks

Even though the casualty rate need not be random in this estimation framework, a potential threat to identification are time-varying confounding factors or systematic manipulation of individuals' treatment status. With the war period being a major episode of migration for blacks from the South (Boustan, 1961), a plausible issue could arise if blacks migrated from low- to high-casualty counties to find semi-skilled employment. In this case, the casualty rate effect picks up an additional migratory response.

To test for this possibility, I re-estimate eq. (1.3) using the share of blacks and the share of black men in a given county as dependent variable. The results for this cross-county migration test are shown in figure 1.A.2. None of the estimated coefficients are significant, neither statistically nor economically. This finding is consistent with the previous balancing test by Pei et al. (2018) in table 1.A.1 for the share of black men. The result also suggests that if blacks gained semi-skilled employment due to the war-induced lack of white workers in this skill-group, then they must have done so in their current counties of residence.

Even if the 1950 interaction in figure 1.A.2 was significantly different from zero, it would imply that the share of blacks in a given county increased by 0.05 p.p. for a one percentage point increase in the casualty rate. Relative to a pre-war average of 22.36%, such an increase would not be considered an economically significant migratory response. The result for the share of black men is the same. This is not to say that African Americans were not migrating during this period. They just did not do so differentially across high- and low-casualty rate counties. Appendix B uses data from the micro Census to provide further evidence that the findings here are not driven by migration patterns by black workers.

A1.4: Selection of Soldiers

Table 1.A.2 reports DiD results of eq. (1.2) including average soldier characteristics by county interacted with a post-war indicator. These characteristics include the average age, years of education, AGCT score (an aptitude test which is the predecessor of the AFQT), share of married, and share of voluntarily enlisted soldiers. This is to preclude the possibility that soldiers from particularly patriotic counties volunteer and die, but that these are also the types of counties where people become more attached to each other and less prejudiced on racial grounds in times of hardship.

The results are unchanged by including these variables. In addition, figure 1.A.5 shows that there are no marked differences in voluntary enlistments between a) the South and the rest of the country and b) above and below median casualty rate counties within the South. While soldiers are certainly selected (e.g. illiterates were service ineligible), the selection into the military and into death does not appear to affect the relationship between the WWII casualty rates among semi-skilled whites and the share of blacks in this skill group.

A1.5: Alternative Treatment Denominators and Denominator Bias

In this section I consider an alternative definition of the treatment variable as compared to eq. (1.1) which used the number of semi-skilled white soldiers as denominator. The rational was to account for unobservable draft deferments. Results using as denominator all semi-skilled white workers,

$$\text{Casualty rate}_c = \frac{\text{Number of fallen semi-skilled white soldiers}_c}{\text{Number of semi-skilled white workers}_c} \times 100 \tag{1.8}$$

are reported in table 1.A.3. This casualty variable has a mean of 0.55, standard deviation of 1.39, minimum of zero, and maximum of 25.54. In all specifications the casualty rate effect is positive and significant at the one percent level. Compared to the baseline specification the coefficients are larger and slightly more volatile with respect to their magnitude when county-specific linear time trends are included. The corresponding coefficients plot for the lags and leads of this treatment variable is shown in figure 1.A.4.

Another concern is that there might be a spurious relationship between the share of blacks in semi-skilled occupations the the casualty rate among semi-skilled whites due to a correlation between the denominators which is driving the estimated change. To account for this, I fix the outcome denominator in eq. (1.1) at it's pre-war level in 1940. This will result in shares that are not necessarily bound in the $[0, 1]$ interval but are indicative for whether results are sensitive with respect to changes in the denominator. Table 1.A.4 reports the estimation results. All but the last column show a positive effect which is significant at the five percent level or less.

A1.6: Sensitivity of Results by State

To test whether results are driven by any given state, I re-estimate the DiD specification in eq. 1.2 using the sample with counties from the $S - 1$ states. The results from this jackknife-type leave-one-out procedure are shown in figure 1.A.6. The figure plots the estimated WWII casualty rate DiD coefficient for each iteration with the left-out state in a given regression being displayed on the vertical axis. The resulting coefficients are indistinguishable from each other as well as from the main result in table 1.3.

A1.6: Spatial Clustering of Casualty Rates

U.S. military units were raised locally during WWII, a practice that was abandoned after D-Day. This policy as well as the patterns observed in the map in figure 1.4 may hint towards spatial dependencies in the outcome. Such spatial correlation would pose problems for inference whereby standard errors are underestimated. To test for such spatial autocorrelation, I compute the I statistic by Moran (1950) for global spatial correlation and the Getis-Ord $G_i^*(d)$ statistic (Getis and Ord, 1992) to test for local spatial correlation. Moran's I is computed as

$$I = \frac{\sum_{i=1}^{n} \sum_{j=1}^{n} w_{ij} C_i C_j}{\sum_{i=1}^{n} C_i^2} \tag{1.9}$$

where i indexes counties with a total number of n counties, j indexes all other counties with $i \neq j$, C is the WWII casualty rate among semi-skilled whites, and w is a spatial weight matrix. Like the standard correlation coefficient, Moran's I lies in $[-1, 1]$. The z score for the corresponding test statistic is given by:

$$z(I) = \frac{I - E(I)}{\sqrt{Var(I)}}$$

Results from this test are reported in table 1.A.6 for distance thresholds of 200, 400, and 600km. Columns (1) to (3) show the casualty rate has a small but statistically significant positive spatial autocorrelation at the 1% level across counties. Moran's I ranges between 0.049 and 0.078. However, once the casualty rate is demeaned by its state-specific averages, Moran's I drops to between -0.003 and -0.008 and becomes insignificant except for the 400km distance threshold where it is marginally significant at the 10% level. This implies that once state fixed effects are controlled for, the casualty rate measure is as good as randomly assigned across geographic space. In the main DiD specifications, these fixed effects would be absorbed by the county fixed effects.

Spatial correlation, however, may exist at a more concentrated level. To test for more local correlations, I provide estimates of the Getis-Ord $G_i^*(d)$ statistic:

$$G_i^*(d) = \frac{\sum_{j=1}^{n} w_{ij}(d) C_j}{\sum_{i=1}^{n} C_j} \tag{1.10}$$

where the notation is as before except that now the spatial weight matrix depends on a certain radius d within which the statistic is computed.[18] Clusters of counties with significantly higher casualty rates are referred to as hot spots. Conversely, those with significantly lower casualty rates are called cold spots.

Table 1.A.7 reports the results from the Getis-Ord test for the same 200, 400, and 600km distance bands as before. The table reports the number of counties within a given z-score interval. Casualty rates show local spatial independence if the z-score of $G_i^*(d)$ falls

[18] For both Moran's I and the Getis-Ord $G_i^*(d)$ binary spatial weights matrices were used. Changing these to exponential or power function type spatial weight matrices does not alter the results. Additional results with alternative spatial weight matrices are not reported here but are available on request. The Stata routine `getisord` by Kondo (2016) was used to compute this test.

within -1.96 and 1.96. Lower z-scores than the lower bound of -1.96 indicate cold spots while higher values than 1.96 indicate hot spots. Again, columns (1) to (3) indicate local spatial correlation with a significant number of counties displaying cold spots (365 counties) and 409 counties having hot spots, out of a total of 1,387 counties. Once state fixed effects are partialled out, almost all counties lose this local spatial autocorrelation as is shown in columns (4) to (6).

Even though spatial correlation appears to be accounted for by geographic fixed effects, I replicate the main findings in table 1.3 and compute Conley (1999) standard errors to correct for spatial dependence.[19] Table 1.A.8 reports the results and shows that the significance of previous results is not driven by spatial autocorrelation.

A1.7: Alternative Regression Specification

Studying the relationship between war casualties and semi-skilled employment for blacks in shares relates directly to the opening graph in figure 1.1. An alternative way of looking at this relation is to run the regression in eq. 1.2 using the levels and taking first differences:

$$\begin{aligned}\Delta\text{blacks in semi-skilled jobs}_{ct} = \beta &\text{white semi-skilled casualties}_c \times \text{post-war}_t \\ &+ \gamma_t + X'_{ct}\xi + \eta_{ct} \end{aligned} \tag{1.11}$$

I control for the total county population and the number of drafted men in addition to the other controls which are the same as in section 1.3. The results from estimating eq. (1.11) are reported in table 1.A.5. On average, a fallen white semi-skilled worker is replaced by four to six African Americans. This is a consistent result across all specifications and shows up with significant coefficients. The exception is column (5) which includes county-specific linear time trends.

The next question is then why there is not a one-to-one substitution between white and black workers. There are several potential explanations. A pessimistic view would be that blacks are less productive and hence it requires more workers from this group to substitute a white worker. Boustan (2009) finds that blacks who migrate North are not perfect substitutes for white workers. She estimates an elasticity of substitution between black and white males of similar skill of 8.3 to 11.1. However, this is likely not only driven by characteristics of African American workers but also by institutional factors such as wage discrimination. Her estimated elasticities are lower than those from the literature on the substitutability between natives and foreigners. This literature finds elasticities in the range of 20 to 47 (see Peri and

[19] Thiemo Fetzer's `reg2hdfespatial` Stata routine was used to run these regressions.

Sparber, 2009).[20]

A more optimistic view is provided by a learning-by-doing argument on part of the employers. Now that employers face labor shortages, they invest more into their ability to screen potential job candidates from a minority group which they had not considered for employment previously. This is the setting of Miller (2017) with the introduction of affirmative action policies. He also finds that the share of blacks keeps rising in firms that were affected by the affirmative action policies during the mid 1960s. Likewise, blacks may invest more into their education or ability to relocate to the cities. Now that manufacturing employment has become a viable option, this changes the incentives to invest on part of the workers. If this line of reasoning was plausible, we should see a gradually increasing rise in semi-skilled employment for blacks after the war. This is shown in figure 1.A.8 which plots the raw levels of black men in semi-skilled jobs over time for counties which are above or below the median number of semi-skilled white WWII casualties.

Overall the findings from this exercise confirm the main results.

[20]Source: Peri, G. and Sparber, C. (2009) "Task Specialization, Immigration, and Wages", American Economic Journal: Applied Economics, Vol. 1(3), pp. 135-169.

Table 1.A.1: Sensitivity Analysis Using Observable County Characteristics

Outcome: % blacks in semi-skilled jobs	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Casualty rate	0.518***	0.524***	0.471***	0.541***	0.559***	0.579***	0.583***	0.592***	0.591***	0.594***
	(0.117)	(0.117)	(0.119)	(0.112)	(0.114)	(0.122)	(0.120)	(0.121)	(0.124)	(0.124)
Draft Rate	-0.120***	−0.115***	−0.127***	−0.156***	−0.156***	−0.156***	−0.146***	−0.144***	−0.147***	−0.150***
	(0.036)	(0.036)	(0.037)	(0.038)	(0.038)	(0.036)	(0.036)	(0.036)	(0.037)	(0.038)
Log mil. spending p.c.		−0.216***	−0.227***	−0.128**	−0.139**	−0.133**	−0.142**	−0.140**	−0.138**	−0.135**
		(0.059)	(0.058)	(0.055)	(0.056)	(0.060)	(0.059)	(0.059)	(0.061)	(0.061)
Neighbor casualties			0.706***	1.235***	1.222***	1.198***	1.281***	1.306***	1.288***	1.284***
			(0.200)	(0.196)	(0.197)	(0.203)	(0.201)	(0.202)	(0.206)	(0.205)
% black men				0.422***	0.408***	0.420***	0.455***	0.449***	0.457***	0.458***
				(0.037)	(0.037)	(0.038)	(0.037)	(0.038)	(0.039)	(0.039)
Manufacturing firms					0.599***	0.589***	0.349*	0.350*	0.364*	0.375**
					(0.208)	(0.218)	(0.181)	(0.182)	(0.190)	(0.186)
Av. manufact. firm size						−0.007**	−0.008**	−0.008***	−0.008**	−0.007**
						(0.003)	(0.003)	(0.003)	(0.003)	(0.003)
% cotton in agriculture							−0.163***	−0.157***	−0.154***	−0.155***
							(0.023)	(0.023)	(0.024)	(0.024)
% cash tenants								0.041**	0.037*	0.034
								(0.021)	(0.022)	(0.022)
Rosenwald schools									−0.386**	−0.380*
									(0.196)	(0.197)
New Deal Relief p.c.										0.011**
										(0.005)
Observations	7,737	7,737	7,721	7,720	7,313	6,986	6,981	6,981	6,769	6,747
Counties	1,388	1,388	1,388	1,388	1,387	1,387	1,387	1,387	1,379	1,379
Balancing Test χ^2	1.890	0.121	0.576	0.346	0.112	0.790	1.014	0.452	0.469	0.050
Balancing Test p-val	0.169	0.728	0.448	0.556	0.738	0.374	0.314	0.502	0.493	0.824

Note: Difference-in-differences regressions of the county-level share of blacks in semi-skilled occupations on the WWII county casualty rate among semi-skilled whites interacted with a post-war indicator. The estimation sample uses decennial U.S. Census data on counties in Southern states from 1920 to 1970. All regressions include county and decade fixed effects. The covariate balancing test by Pei et al. (2018) is reported in the bottom two rows of the table where the null hypothesis is that a new added control does not vary systematically across high- and low-casualty rate counties. The variables on WWII military spending, WWII casualties in neighboring counties, New Deal Relief per capita, and the unemployment rate in 1937 are interacted with a post-war indicator. Standard errors clustered at the county level. Significance levels are denoted by * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 1.A.2: Difference-in-Differences Results with Average Soldier Characteristics

	Outcome: % blacks in semi-skilled jobs (pre-war mean = 12.433)					
	(1)	(2)	(3)	(4)	(5)	(6)
Casualty rate$_c\times$ Post-war$_t$	0.515***	0.530***	0.504***	0.527***	0.539**	0.465***
	(0.119)	(0.142)	(0.143)	(0.148)	(0.217)	(0.136)
Controls		Yes		Yes	Yes	Yes
1940 controls $\times$ time			Yes			
Flexible state time trends				Yes		
Linear county time trends					Yes	
Doubly-robust selection						Yes
Observations	7,737	5,713	5,692	5,713	5,713	6,429
Counties	1,388	1,320	994	1,320	1,320	1,375
Adj. R^2	0.855	0.879	0.876	0.884	0.915	0.863
Oster's δ	1.273	1.220	1.122	1.409	0.542	0.995

Note: Difference-in-differences regressions of the county-level share of blacks in semi-skilled occupations on the WWII county casualty rate among semi-skilled whites interacted with a post-war indicator. The estimation sample uses decennial U.S. Census data on counties in Southern states from 1920 to 1970. Controls include county and decade fixed effects, the county draft rate, average casualty rate in the neighboring counties, log WWII spending per capita, share of black men, share of rural population, no. of manufacturing establishments per capita, average manufacturing firm size, log manufacturing value added per worker, share of employment in manufacturing, share of land in agricultural production, share of acres in cotton production, share of cash tenants, average value of machinery per farm, lynchings per 1,000 blacks between 1900 and 1930, no. of Rosenwald schools per 1,000 blacks, share of acres flooded by the Mississippi in 1928, no. of slaves in 1860, Republican vote share, New Deal spending per capita 1933-35 (loans, public works, AAA, FHA loans), and the unemployment rate in 1937, as well as the average soldier characteristics in each county including age, education, AGCT score, share of married, and share of voluntarily enlisted. Time-invariant controls are interacted with decade fixed effects. Monetary values are deflated to 2010 U.S. dollars. The doubly-robust selection method implements the Belloni et al. (2014) machine learning covariate selection algorithm for testing the stability of treatment effects with respect to the observables. Oster's (2017) test for selection on unobservables is reported in the final row by computing the coefficient of proportionality δ for which the coefficient on the semi-skilled casualty rate among whites would equal zero. Standard errors clustered at the county level. Significance levels are denoted by * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 1.A.3: Difference-in-Differences Results with Alternative Treatment Denominator

	Outcome: % blacks in semi-skilled jobs (pre-war mean = 12.433)					
	(1)	(2)	(3)	(4)	(5)	(6)
Casualty rate$_c \times$ Post-war$_t$	1.071***	1.770***	1.568***	1.870***	2.607***	1.962***
	(0.280)	(0.386)	(0.295)	(0.392)	(0.561)	(0.349)
Controls		Yes		Yes	Yes	Yes
1940 controls $\times$ time			Yes			
Flexible state time trends				Yes		
Linear county time trends					Yes	
Doubly-robust selection						Yes
Observations	7,737	5,713	5,692	5,713	5,713	6,429
Counties	1,388	1,320	994	1,320	1,320	1,375
Adj. R^2	0.856	0.879	0.874	0.885	0.916	0.877
Oster's δ	1.946	1.514	0.953	1.487	0.853	1.568

Note: Difference-in-differences regressions of the county-level share of blacks in semi-skilled occupations on the WWII county casualty rate among semi-skilled whites interacted with a post-war indicator. The casualty rate in county c here is one hundred times the total number of killed semi-skilled whites over the number of total semi-skilled whites in 1940. The estimation sample uses decennial U.S. Census data on counties in Southern states from 1920 to 1970. Coefficients are expressed in terms of a one standard deviation increase in the casualty rate. Controls include county and decade fixed effects, the county draft rate, average casualty rate in the neighboring counties, log WWII spending per capita, share of black men, share of rural population, no. of manufacturing establishments per capita, average manufacturing firm size, log manufacturing value added per worker, share of employment in manufacturing, share of land in agricultural production, share of acres in cotton production, share of cash tenants, average value of machinery per farm, lynchings per 1,000 blacks between 1900 and 1930, no. of Rosenwald schools per 1,000 blacks, share of acres flooded by the Mississippi in 1928, no. of slaves in 1860, Republican vote share, New Deal spending per capita 1933-35 (loans, public works, AAA, FHA loans), and the unemployment rate in 1937. Time-invariant controls are interacted with decade fixed effects. Monetary values are deflated to 2010 U.S. dollars. The doubly-robust selection method implements the Belloni et al. (2014) machine learning covariate selection algorithm for testing the stability of treatment effects with respect to the observables. Oster's (2017) test for selection on unobservables is reported in the final row by computing the coefficient of proportionality δ for which the coefficient on the semi-skilled casualty rate among whites would equal zero. Standard errors clustered at the county level. Significance levels are denoted by * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 1.A.4: Difference-in-Differences Results with Fixed Outcome Denominator

	Outcome: % blacks in semi-skilled jobs (pre-war mean = 12.433)					
	(1)	(2)	(3)	(4)	(5)	(6)
Casualty rate$_c\times$ Post-war$_t$	1.167***	1.046***	0.595**	0.703**	1.218**	0.345
	(0.283)	(0.358)	(0.276)	(0.337)	(0.538)	(0.281)
Controls		Yes		Yes	Yes	Yes
1940 controls $\times$ time			Yes			
Flexible state time trends				Yes		
Linear county time trends					Yes	
Doubly-robust selection						Yes
Observations	7,737	5,713	5,692	5,713	5,713	6,429
Counties	1,388	1,334	994	1,334	1,334	1,374
Adj. R^2	0.856	0.879	0.874	0.885	0.916	0.877
Oster's δ	1.946	1.514	0.953	1.487	0.853	1.568

Note: Difference-in-differences regressions of the county-level share of blacks in semi-skilled occupations on the WWII county casualty rate among semi-skilled whites interacted with a post-war indicator. The casualty rate in county c here is one hundred times the total number of killed semi-skilled whites over the number of total semi-skilled whites in 1940. The estimation sample uses decennial U.S. Census data on counties in Southern states from 1920 to 1970. The denominator of the outcome (number of semi-skilled workers) is fixed at 1940 values to reduce denominator bias. Controls include county and decade fixed effects, the county draft rate, average casualty rate in the neighboring counties, log WWII spending per capita, share of black men, share of rural population, no. of manufacturing establishments per capita, average manufacturing firm size, log manufacturing value added per worker, share of employment in manufacturing, share of land in agricultural production, share of acres in cotton production, share of cash tenants, average value of machinery per farm, lynchings per 1,000 blacks between 1900 and 1930, no. of Rosenwald schools per 1,000 blacks, share of acres flooded by the Mississippi in 1928, no. of slaves in 1860, Republican vote share, New Deal spending per capita 1933-35 (loans, public works, AAA, FHA loans), and the unemployment rate in 1937. Time-invariant controls are interacted with decade fixed effects. Monetary values are deflated to 2010 U.S. dollars. The doubly-robust selection method implements the Belloni et al. (2014) machine learning covariate selection algorithm for testing the stability of treatment effects with respect to the observables. Oster's (2017) test for selection on unobservables is reported in the final row by computing the coefficient of proportionality δ for which the coefficient on the semi-skilled casualty rate among whites would equal zero. Standard errors clustered at the county level. Significance levels are denoted by * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 1.A.5: Difference-in-Differences Results with First Differenced Outcome

	Outcome: Δ No. of blacks in semi-sk. jobs (pre-war mean = 232.842)					
	(1)	(2)	(3)	(4)	(5)	(6)
No. semi-sk. white deaths$_c$	5.116***	4.432**	6.678**	4.295*	7.382	4.320***
$\times$ Post-war$_t$	(1.779)	(2.241)	(3.243)	(2.399)	(6.757)	(1.613)
Controls		Yes		Yes	Yes	Yes
1940 controls $\times$ time			Yes			
Flexible state time trends				Yes		
Linear county time trends					Yes	
Doubly-robust selection						Yes
Observations	6,006	4,677	4,513	4,677	4,677	4,687
Counties	1,388	1,289	994	1,289	1,289	1,289
Adj. R^2	0.377	0.375	0.383	0.388	0.280	0.390

Note: Difference-in-differences regressions of the county-level share of blacks in semi-skilled occupations on the WWII county casualty rate among semi-skilled whites interacted with a post-war indicator. The estimation sample uses decennial U.S. Census data on counties in Southern states from 1920 to 1970. Controls include decade fixed effects, county population, number of drafted soldiers, average casualty rate in the neighboring counties, log WWII spending per capita, share of black men, share of rural population, no. of manufacturing establishments per capita, average manufacturing firm size, log manufacturing value added per worker, share of employment in manufacturing, share of land in agricultural production, share of acres in cotton production, share of cash tenants, average value of machinery per farm, lynchings per 1,000 blacks between 1900 and 1930, no. of Rosenwald schools per 1,000 blacks, share of acres flooded by the Mississippi in 1928, no. of slaves in 1860, Republican vote share, New Deal spending per capita 1933-35 (loans, public works, AAA, FHA loans), and the unemployment rate in 1937. Time-invariant controls are interacted with decade fixed effects. Monetary values are deflated to 2010 U.S. dollars. The doubly-robust selection method implements the Belloni et al. (2014) machine learning covariate selection algorithm for testing the stability of treatment effects with respect to the observables. Significance levels are denoted by * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 1.A.6: Spatial Independence Test of WWII Casualty Rates

	Distance threshold					
	200km	400km	600km	200km	400km	600km
	(1)	(2)	(3)	(4)	(5)	(6)
Moran's I	0.078***	0.064***	0.049***	-0.008	-0.005*	-0.003
	[16.473]	[26.595]	[31.875]	[-1.557]	[-1.775]	[-1.235]
Observations	1,387	1,387	1,387	1,387	1,387	1,387
State FE				Yes	Yes	Yes

Note: Moran's I for testing spatial independence of the WWII casualty rate among semi-skilled whites. For each I, the z-score is reported in squared brackets using a binary spatial weight matrix. Each county is identified by the latitude and longitude of its centroid. Significance levels are denoted by * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 1.A.7: Testing for Hot and Cold Spots of WWII Casualty Rates

	Distance threshold					
Getis-Ord $G_i^*(d)$	200km	400km	600km	200km	400km	600km
z-score interval	(1)	(2)	(3)	(4)	(5)	(6)
z $\leq$ -2.58	232	347	347	0	0	0
-2.58 $<$ z $\leq$ -1.96	133	49	33	8	2	0
-1.96 $<$ z $<$ 1.96	613	371	262	1,370	1,378	1,386
1.96 $\leq$ z $<$ 2.58	130	80	59	8	7	1
2.58 $\leq$ z	279	540	686	1	0	0
Observations	1,387	1,387	1,387	1,387	1,387	1,387
State FE				Yes	Yes	Yes

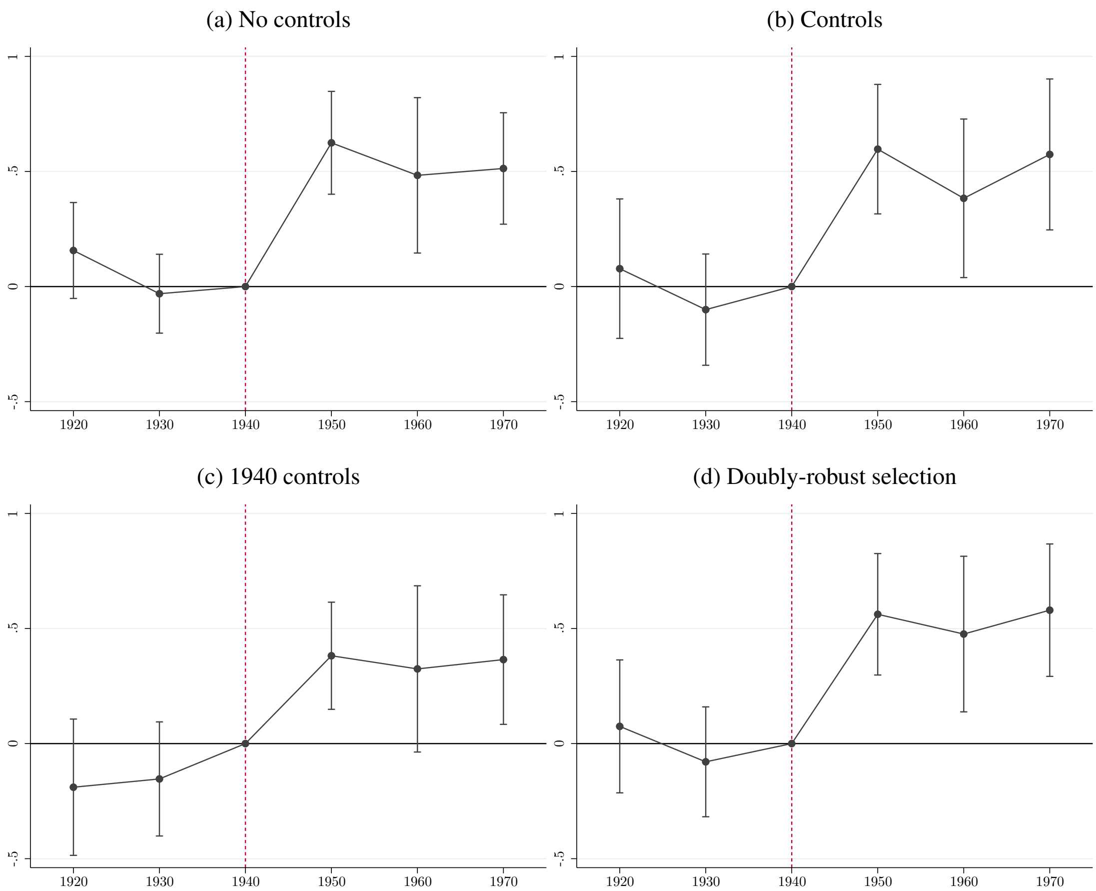
Note: Getis-Ord $G_i^*(d)$ test for testing local spatial independence of the WWII casualty rate among semi-skilled whites. Local spatial independence is given when the z-score on the corresponding test statistic lies within -1.96 $<$ z $<$ 1.96. Unusually low casualty rate clusters (cold spots) are found for counties with z-scores of z $\leq$ -1.96. Conversely, unusually high casualty rate clusters (hot spots) are found for counties with z-scores of 1.96 $\leq$ z. The number of counties in each z-score bin is provided in the rows of the table. Each county is identified by the latitude and longitude of its centroid.

Table 1.A.8: County Level Difference-in-Differences Results with Conley Standard Errors

	Outcome: % blacks in semi-skilled jobs (pre-war mean = 12.433)					
	(1)	(2)	(3)	(4)	(5)	(6)
Casualty rate$_c \times$ Post-war$_t$	0.515	0.545	0.508	0.548	0.587	0.589
s.e. (200km)	(0.072)	(0.075)	(0.078)	(0.075)	(0.080)	(0.067)
s.e. (400km)	(0.077)	(0.074)	(0.078)	(0.075)	(0.074)	(0.078)
s.e. (600km)	(0.079)	(0.076)	(0.079)	(0.078)	(0.073)	(0.077)
Controls		Yes		Yes	Yes	Yes
1940 controls $\times$ time			Yes			
Flexible state time trends				Yes		
Linear county time trends					Yes	
Doubly-robust selection						Yes
Observations	7,737	5,713	5,692	5,713	5,713	5,723
Adj. R^2	0.013	0.169	0.158	0.214	0.192	0.015

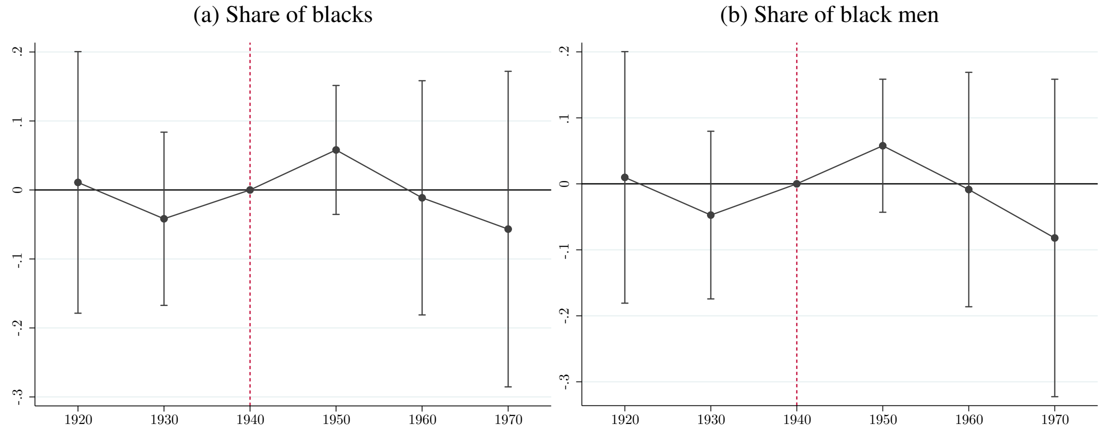
Note: Difference-in-differences regressions of the county-level share of blacks in semi-skilled occupations on the WWII county casualty rate among semi-skilled whites interacted with a post-war indicator. The estimation sample uses decennial U.S. Census data on counties in Southern states from 1920 to 1970. Controls include county and decade fixed effects, the county draft rate, average casualty rate in the neighboring counties, log WWII spending per capita, share of black men, share of rural population, no. of manufacturing establishments per capita, average manufacturing firm size, log manufacturing value added per worker, share of employment in manufacturing, share of land in agricultural production, share of acres in cotton production, share of cash tenants, average value of machinery per farm, lynchings per 1,000 blacks between 1900 and 1930, no. of Rosenwald schools per 1,000 blacks, share of acres flooded by the Mississippi in 1928, no. of slaves in 1860, Republican vote share, New Deal spending per capita 1933-35 (loans, public works, AAA, FHA loans), and the unemployment rate in 1937. Time-invariant controls are interacted with decade fixed effects. Monetary values are deflated to 2010 U.S. dollars. The doubly-robust selection method implements the Belloni et al. (2014) machine learning covariate selection algorithm for testing the stability of treatment effects with respect to the observables. Standard errors adjusted for spatial correlation using Conley (1999) standard errors with a distance threshold of 200, 400, and 600km.

Figure 1.A.1: Difference-in-Differences Coefficient Plots using Alternative Specifications



Note: Difference-in-differences regressions of the county-level share of blacks in semi-skilled occupations on the WWII county casualty rate among semi-skilled whites interacted with decade fixed effects. The omitted baseline decade is 1940 which is marked by the dashed line. This is the last pre-treatment period. The estimation sample contains counties in Southern states from 1920 to 1970. Coefficients show the effect of a one standard deviation increase in the casualty rate on the outcome in terms of percentage points. All regressions include county and decade fixed effects unless stated otherwise. If used by a given specification, controls include the county draft rate, average casualty rate in the neighboring counties, log WWII spending per capita, share of black men, share of rural population, no. of manufacturing establishments per capita, average manufacturing firm size, log manufacturing value added per worker, share of employment in manufacturing, share of land in agricultural production, share of acres in cotton production, share of cash tenants, average value of machinery per farm, lynchings per 1,000 blacks between 1900 and 1930, no. of Rosenwald schools per 1,000 blacks, share of acres flooded by the Mississippi in 1928, no. of slaves in 1860, Republican vote share, New Deal spending per capita 1933-35 (loans, public works, AAA, FHA loans), and the unemployment rate in 1937. Time-invariant controls are interacted with decade fixed effects. The 1940 controls plot fixes all controls at their level in that year and interacts them with decade fixed effects. The doubly-robust selection method implements the Belloni et al. (2014) machine learning covariate selection algorithm to select the most relevant controls. Monetary values are deflated to 2010 U.S. dollars. Standard errors clustered at the county level. Error bars show 95% confidence intervals around each coefficient estimate.

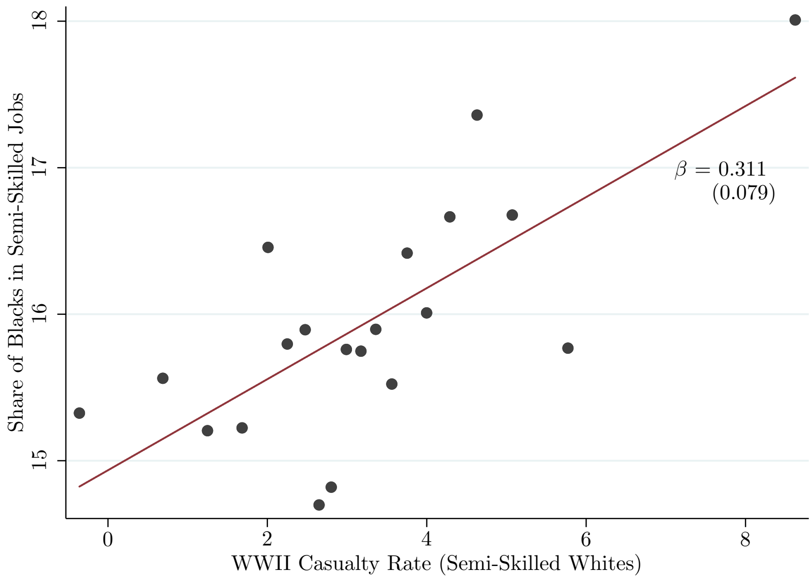
Figure 1.A.2: Difference-in-Differences Cross-County Migration Test



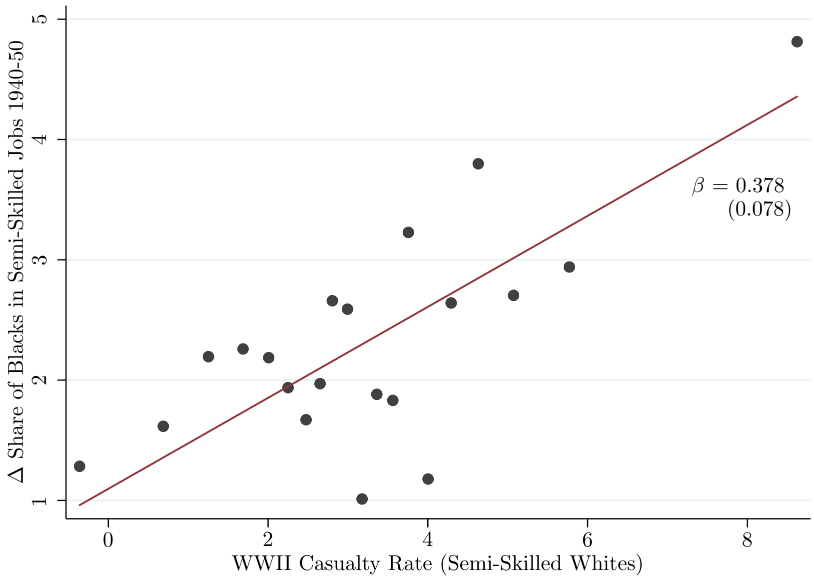
Note: Difference-in-differences regressions of the county-level share of blacks and the share of black men in percent on the WWII county casualty rate among semi-skilled whites interacted with decade fixed effects. The omitted baseline decade is 1940 which is marked by the dashed line. This is the last pre-treatment period. The estimation sample contains decennial U.S. Census data on counties in Southern states from 1920 to 1970. Coefficients show the effect of a one standard deviation increase in the casualty rate on the outcome in terms of percentage points. Controls include county fixed effects, flexible state-specific time trends, the county draft rate, average casualty rate in the neighboring counties, log WWII spending per capita, share of rural population, no. of manufacturing establishments per capita, average manufacturing firm size, log manufacturing value added per worker, share of employment in manufacturing, share of land in agricultural production, share of acres in cotton production, share of cash tenants, average value of machinery per farm, lynchings per 1,000 blacks between 1900 and 1930, no. of Rosenwald schools per 1,000 blacks, share of acres flooded by the Mississippi in 1928, no. of slaves in 1860, Republican vote share, New Deal spending per capita 1933-35 (loans, public works, AAA, FHA loans), and the unemployment rate in 1937. Time-invariant controls are interacted with decade fixed effects. Monetary values are deflated to 2010 U.S. dollars. Standard errors clustered at the county level. Error bars show 95% confidence intervals around each coefficient estimate.

Figure 1.A.3: Scatter Plots for WWII Casualty Rates and the Share of Blacks in Semi-Skilled Jobs in Levels and First Differences

(a) Correlation with the Semi-Skilled Share Level in 1950



(b) Correlation with the Semi-Skilled Share 1940 to 50 First Difference



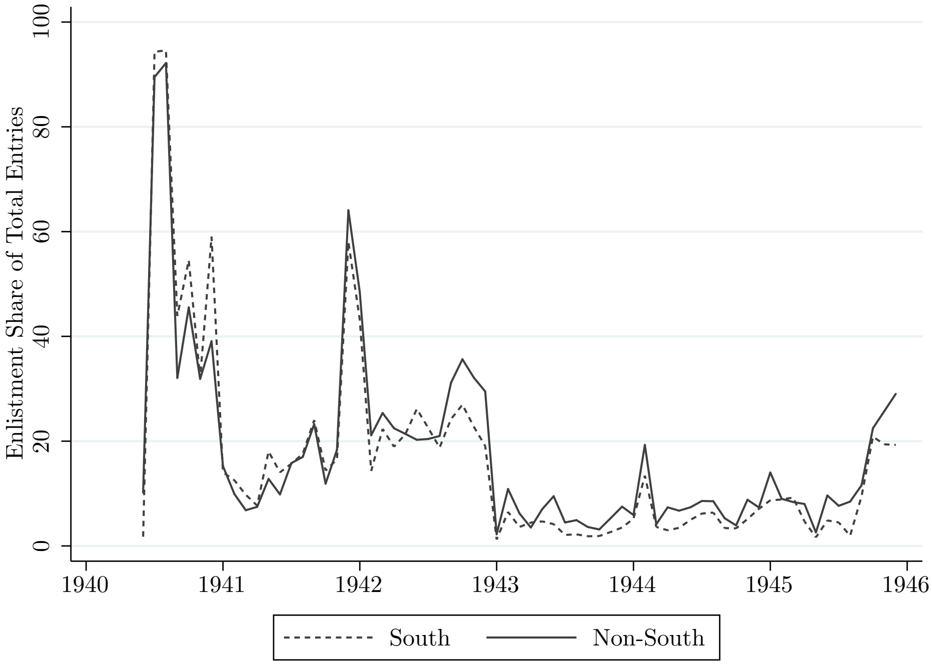
Note: Scatter plots of the relation between the WWII casualty rate among semi-skilled whites and the share of blacks in semi-skilled employment in 1950 across counties (panel a), and the change in the share of blacks in semi-skilled employment from 1940 to 1950 (panel b). Controls partial out county characteristics in 1940 including the county population, share of black men, and the shares of agricultural and manufacturing employment.

Figure 1.A.4: Difference-in-Differences Coefficient Plot with Alternative Treatment

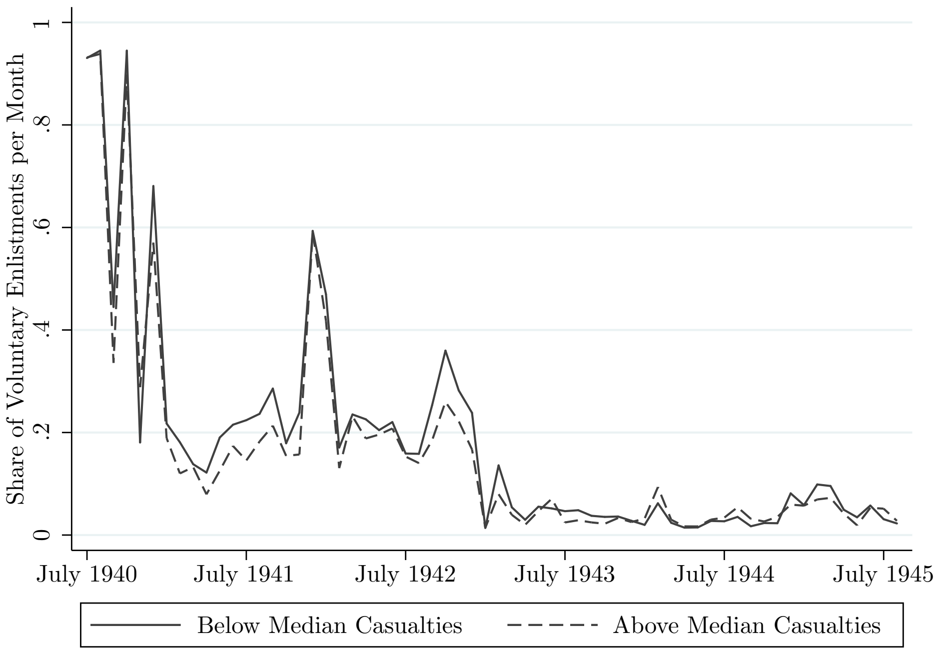
Note: Difference-in-differences regressions of the county-level share of blacks in semi-skilled occupations on the WWII county casualty rate among semi-skilled whites interacted with decade fixed effects. The denominator in the computation of the casualty rate here is the number of all semi-skilled whites in 1940 in county c. The omitted baseline decade is 1940 which is marked by the dashed line. This is the last pre-treatment period. The estimation sample contains counties in Southern states from 1920 to 1970. Coefficients show the effect of a one standard deviation increase in the casualty rate on the outcome in terms of percentage points. Controls include county and decade fixed effects, the county draft rate, average casualty rate in the neighboring counties, log WWII spending per capita, share of black men, share of rural population, no. of manufacturing establishments per capita, average manufacturing firm size, log manufacturing value added per worker, share of employment in manufacturing, share of land in agricultural production, share of acres in cotton production, share of cash tenants, average value of machinery per farm, lynchings per 1,000 blacks between 1900 and 1930, no. of Rosenwald schools per 1,000 blacks, share of acres flooded by the Mississippi in 1928, no. of slaves in 1860, Republican vote share, New Deal spending per capita 1933-35 (loans, public works, AAA, FHA loans), and the unemployment rate in 1937. Time-invariant controls are interacted with decade fixed effects. Monetary values are deflated to 2010 U.S. dollars. Standard errors clustered at the county level. Error bars show 95% confidence intervals around each coefficient estimate.

Figure 1.A.5: Voluntary Enlistment Rates

(a) South vs. Non-South

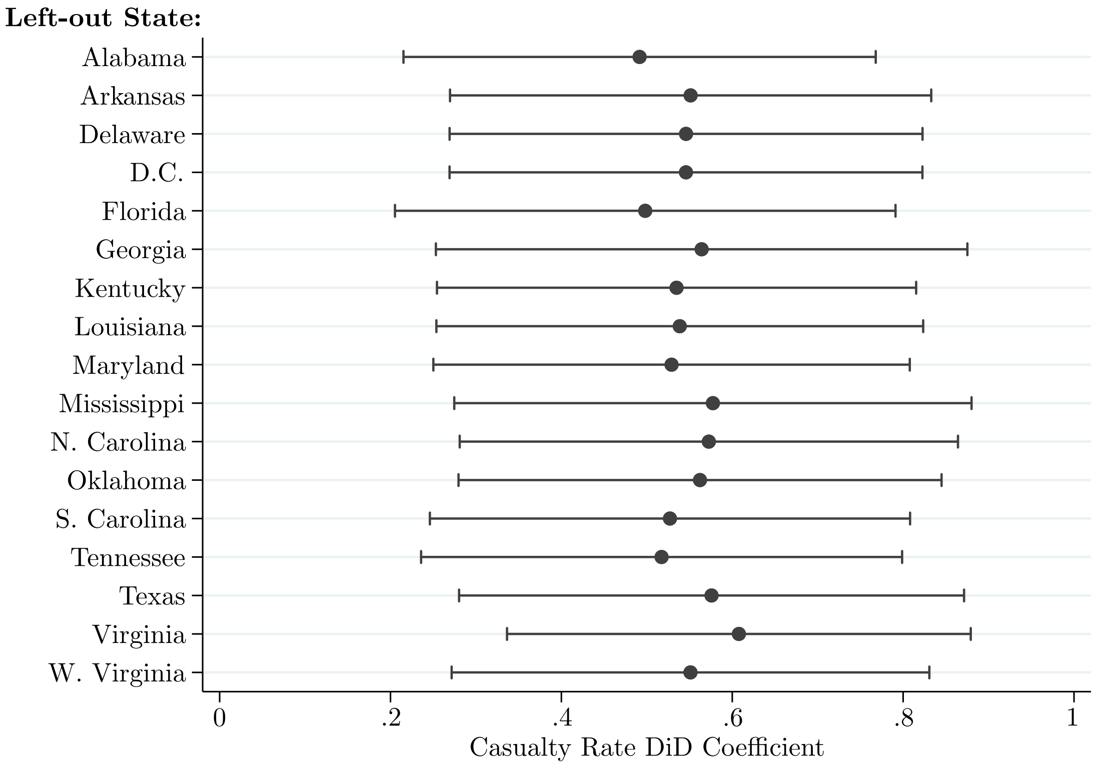


(b) Within South



Note: Share of voluntary enlistments out of total new entries into the Army and Army Air Force by month. The drop at the end of 1942 is because voluntary enlistment was forbidden to avoid hurting the war economy due to overenthusiastic enlistments as was the case in the United Kingdom. After December 1942 only men aged 38 or older were allowed to volunteer if they demonstrated their physical and mental fitness for service.

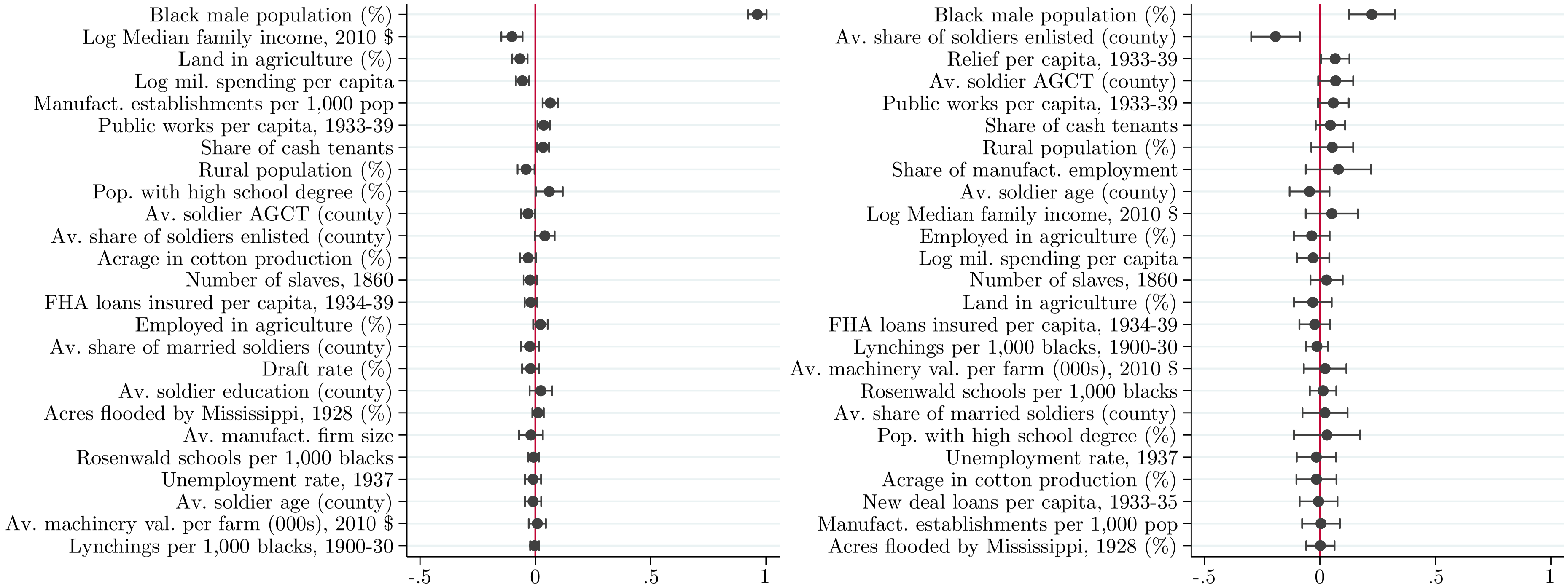
Figure 1.A.6: Leave-One Out DiD Sensitivity Check



Note: Difference-in-differences regressions of the county-level share of blacks in semi-skilled occupations on the WWII county casualty rate among semi-skilled whites interacted with a post-war indicator. The estimation sample uses decennial U.S. Census data on counties in Southern states from 1920 to 1970. Each regression leaves out all counties from a specific state at a time to assess whether results are driven by any one single state. The omitted state is listed on the left. Each regression includes county and decade fixed effects, the county draft rate, average casualty rate in the neighboring counties, log WWII spending per capita, share of black men, share of rural population, no. of manufacturing establishments per capita, average manufacturing firm size, log manufacturing value added per worker, share of employment in manufacturing, share of land in agricultural production, share of acres in cotton production, share of cash tenants, average value of machinery per farm, lynchings per 1,000 blacks between 1900 and 1930, no. of Rosenwald schools per 1,000 blacks, share of acres flooded by the Mississippi in 1928, no. of slaves in 1860, Republican vote share, New Deal spending per capita 1933-35 (loans, public works, AAA, FHA loans), and the unemployment rate in 1937. Time-invariant controls are interacted with decade fixed effects. Monetary values are deflated to 2010 U.S. dollars. Standard errors are clustered by county. Error bars show 95% confidence intervals.

Figure 1.A.7: Observable Determinants of Outcome and Treatment

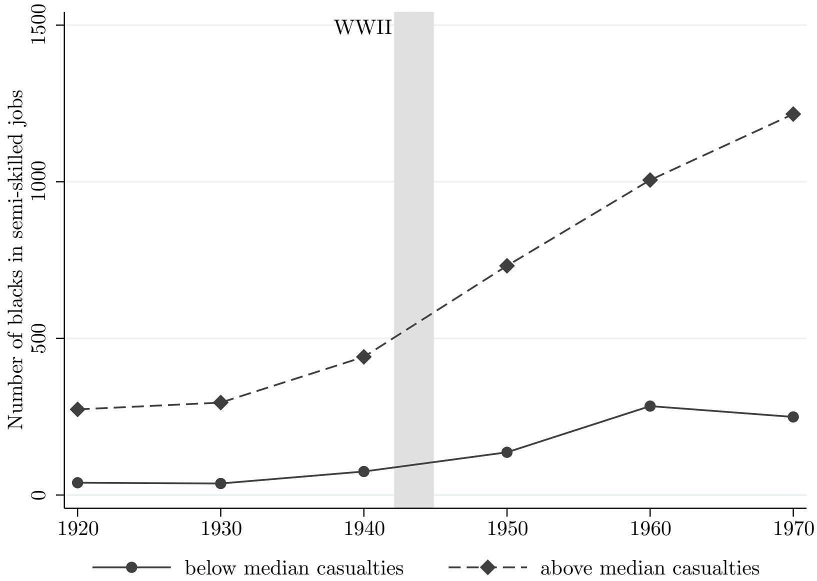
(a) **Outcome**: Share of Blacks in Semi-Skilled Jobs (b) **Treatment**: WWII Casualty Rate



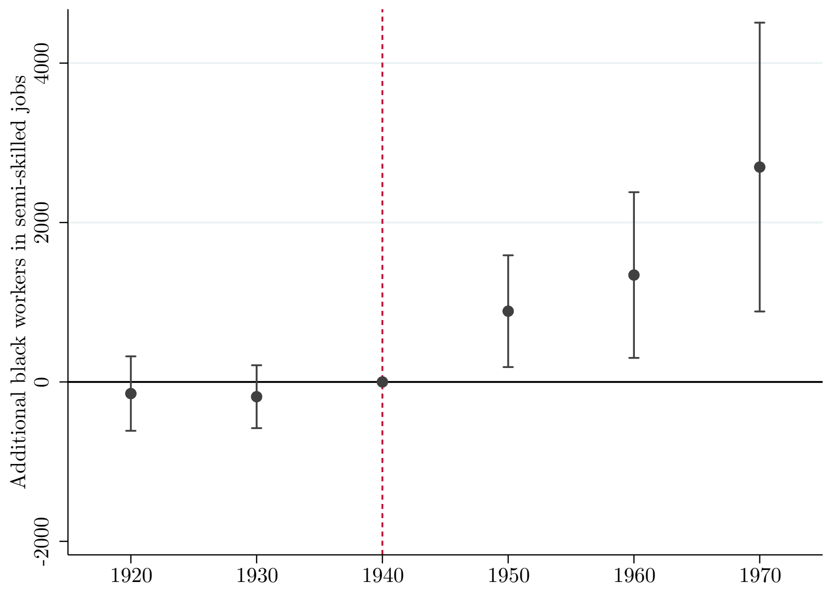
Note: Cross-sectional correlation ranking of pre-war controls from 1940 with the post-war outcome (share of blacks in semi-skilled jobs) and treatment (WWII casualty rate among semi-skilled whites) variables in 1950. All variables are de-meaned and standardized to have unit variance. Beta coefficients are ranked by the absolute value of their t-statistic to show the most important correlates from top to bottom. All regressions include state fixed effects for which coefficients have been dropped for this plot. Error bars show 95% confidence intervals.

Figure 1.A.8: Black Semi-Skilled Employment in Levels - Conditional and Unconditional

(a) Levels



(b) Coefficients Plot



Note: Panel (a) plots the number of black men employed in semi-skilled occupations for 1,388 Southern counties from 1920-70. Counties are split into two groups, those with above and below median WWII casualties among semi-skilled whites. The gray shaded area marks years with U.S. involvement in the war. Panel (b) plots the coefficients of the above median casualty indicator interacted with decade fixed effects, omitting 1940 as the baseline. The dashed line marks the last pre-treatment period. The regression controls for county and decade fixed effects, the log of WWII military spending per capita, the draft rate, average casualty rate in neighboring counties, number of manufacturing establishments per capita, average manufacturing firm size, average value added per manufacturing worker, the share of manufacturing employment, the share of black men, share of cotton production in agriculture, counties flooded by the Mississippi in 1928, Republican vote share, the share of land mass used in agriculture, the share of cash tenants, and flexible state-specific time trends. Error bars show 95% confidence intervals. Standard errors are clustered at the county level.

B Commuting Zone Appendix

B1) Semi-Skilled Employment and Economic Outcomes

While the casualty rate is arguably the more exogenous shock, it might still be instructive to examine the effect of semi-skilled employment of blacks before and after the war on other economic outcomes. A first test amounts to running the following difference-in-difference-in-differences (DDD) regression:

$$\begin{aligned} y_{izt} = & \beta_1 \left(\text{semi-skilled}_{izt}\text{post-WWII}_t\right) \\ & + \beta_2 \left(\text{semi-skilled}_{izt} \times \text{black}_{izt} \times \text{post-WWII}_t\right) \\ & + \alpha_z + \lambda_t + \delta\text{black}_{izt} + X'_{(i)zt}\gamma + \epsilon_{izt} \end{aligned} \tag{1.12}$$

where y_{izt} is the given economic outcome for individual i in commuting zone z in decade t. The regression includes fixed effects for race black$_{izt}$, commuting zone α_z, and census year λ_t, as well as individual- and commuting zone-level controls $X'_{(i)zt}$. Individual level controls include dummies for age, marital status, and place of birth. Commuting zone controls include all the controls used also in section 1.3 which are aggregated to the county- to the commuting zone-level. Standard errors are clustered by commuting zone.

Estimating a triple differences regression, using whites as additional control group, has the attraction that it also estimates the response by whites with respect to the economic upgrading of blacks. This provides an estimate for whether whites lose out relative to blacks, whether both groups are affected by the shift of blacks into semi-skilled employment, or whether black economic progress is entirely independent of the economic fortunes of white workers. Table 1.B.1 reports the results from this regression for six outcomes. The first three are indicators for urban and cross-state migration status, and home ownership. A cross-state migrant here is a person who does not reside in their state of birth.

While the post-war skill-upgrade has positive effects for African Americans, it is typically associated with negative effects for whites. This finding points towards potential selection which would be consistent with the previous literature. For instance, in both the full U.S. and Southern samples, semi-skilled post-war employment has a positive and statistically significant impact on the urban status of blacks, their wages, and house values, but affects whites in the opposite direction. Boustan (2010) shows that for every black arrival into a Northern city center 2.7 whites leave. If the more skilled or wealthy whites can more easily switch jobs or their homes, then the remaining whites are a selected part of the white population that was too constrained to satisfy their racial preferences - or that was more tolerant to begin with.

Table 1.B.1: Micro Census Triple Differences Results using the Semi-Skilled Treatment

Outcome:	ln(wage)	Education	Owns home	ln(house val.)	Migrant
			Panel A: All U.S.		
Semi-Skilled$_{izt}\times$ Post-war$_t$	-0.072***	-1.832***	-0.001	-0.246***	-0.013**
	(0.007)	(0.035)	(0.003)	(0.005)	(0.006)
Semi-Skilled$_{izt}\times$ Black$_{izt}\times$	0.250***	1.757***	0.012*	0.269***	0.015**
Post-war$_t$	(0.010)	(0.054)	(0.007)	(0.015)	(0.008)
Observations	2,696,784	3,119,306	4,211,898	1,527,493	4,335,995
Adj. R^2	0.502	0.457	0.251	0.487	0.323
			Panel B: South Only		
Semi-Skilled$_{izt}\times$ Post-war$_t$	-0.100***	-1.843***	-0.012**	-0.309***	0.014
	(0.014)	(0.069)	(0.005)	(0.011)	(0.009)
Semi-Skilled$_{izt}\times$ Black$_{izt}\times$	0.288***	1.902***	0.009	0.344***	-0.015**
Post-war$_t$	(0.013)	(0.078)	(0.006)	(0.019)	(0.006)
Observations	766,766	910,755	1,226,713	428,483	1,268,890
Adj. R^2	0.507	0.452	0.241	0.508	0.467

Note: Difference-in-differenece-in-differences regression of economic outcomes on the commuting zone WWII casualty rate among semi-skilled whites interacted with a post-WWII dummy, and with a black indicator for individuals living in 722 commuting zones in the whole U.S. The estimation sample contains data from the decennial U.S. micro Census from 1920-70 on non-institutionalized, working black and white males aged 15-65 who are not currently attending school. All regressions include commuting zone and Census year fixed effects. Owns home is a binary outcome for whether an individual owns their home. The log house value, log wages, and education variables are only available from 1940 onward. Log house value is also missing for 1950. Individual level controls include age, marital status, age and place of birth dummies. Commuting zone level controls are the WWII draft rate, log WWII spending per capita, share of black men, share of rural population, no. of manufacturing establishments per capita, average manufacturing firm size, log manufacturing value added per worker, share of employment in manufacturing, share of land in agricultural production, share of acres in cotton production, share of cash tenants, average value of machinery per farm, lynchings per 1,000 blacks between 1900 and 1930, no. of Rosenwald schools per 1,000 blacks, share of acres flooded by the Mississippi in 1928, no. of slaves in 1860, Republican vote share, New Deal spending per capita 1933-35 (loans, public works, AAA, FHA loans), and the unemployment rate in 1937. Time-invariant controls are interacted with decade fixed effects. Monetary values are deflated to 2010 U.S. dollars. Standard errors clustered at the commuting zone level in parentheses. Significance levels are denoted by * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Blacks who secure a semi-skilled job after the war are looking at a substantial wage increase of 28.4 p.p. in the full sample and 33.2 p.p. in the Southern sample. The skill-upgrade is only significantly related to the probability of home ownership in the full sample with a 1.2 p.p. rise. However, when African Americans manage to own their home, this is now of substantially higher value for those who experience the skill-upgrade. The associated home value increase is 30.9 p.p. in the whole U.S. and 41.1 p.p. in the South. For whites there is a negative effect on house values which might be due to outmigration of wealthier whites driving down home values (Boustan and Margo, 2013) or a decline in housing segregation that reduces prices for homes of whites (Logan and Parman, 2017).

B2) Further Robustness Checks for Migration Responses

Are the results here driven by migration? To test for this possibility, tables 1.B.2 and 1.B.3 repeat the DDD analysis for the sub-samples of those who do not reside in their state of birth and birth-state stayers in the country as a whole and in the South only, respectively. While wage gains are typically larger for those who move, the casualty rate effect increases the house values only for birth-state stayers in the full sample. The likely reason for this relates to blacks moving to lower quality housing in the city centers of the industrial centers in the North. When considering the Southern sample, movers also outperform stayers in terms of house value. This difference is not statistically significant though. Even though moving is an endogenous choice, the results here provide evidence that the economic benefits are not only reaped by this particular group of individuals. Also stayers gain. Even though the wage increases associated with the white WWII casualty rate are lower for stayers, the increases in house value and educational attainment are comparable across movers and stayers.

Table 1.B.2: Movers vs. Birth-State Stayers, all U.S.

Outcome:	ln(wage)	Education	Owns home	ln(house value)
	Panel A: Cross-State Migrants			
Casualty rate$_z \times$ Post-war$_t$	-0.022**	-0.030	0.000	-0.051**
	(0.010)	(0.037)	(0.005)	(0.021)
Casualty rate$_z \times$ Black$_{izt} \times$	0.055***	0.240***	0.004	0.025
Post-war$_t$	(0.006)	(0.032)	(0.005)	(0.020)
Observations	1,073,820	1,208,270	1,515,175	557,437
Adj. R^2	0.462	0.409	0.263	0.430
	Panel B: Birth-State Stayers			
Casualty rate$_z \times$ Post-war$_t$	-0.012	-0.033	-0.007*	-0.033
	(0.009)	(0.029)	(0.004)	(0.022)
Casualty rate$_z \times$ Black$_{izt} \times$	0.027***	0.311***	-0.012***	0.084***
Post-war$_t$	(0.007)	(0.028)	(0.003)	(0.010)
Observations	1,622,964	1,911,036	2,696,723	970,056
Adj. R^2	0.523	0.453	0.255	0.492

Note: Difference-in-differenece-in-differences regression of economic outcomes on the commuting zone WWII casualty rate among semi-skilled whites interacted with a post-WWII dummy, and with a black indicator for individuals living in 722 commuting zones in the whole U.S. The estimation sample contains data from the decennial U.S. micro Census from 1920-70 on non-institutionalized, working black and white males aged 15-65 who are not currently attending school. All regressions include commuting zone and Census year fixed effects. Owns home is a binary outcome for whether an individual owns their home. The log house value, log wages, and education variables are only available from 1940 onward. Log house value is also missing for 1950. Individual level controls include age, marital status, age and place of birth dummies. Commuting zone level controls are the WWII draft rate, log WWII spending per capita, share of black men, share of rural population, no. of manufacturing establishments per capita, average manufacturing firm size, log manufacturing value added per worker, share of employment in manufacturing, share of land in agricultural production, share of acres in cotton production, share of cash tenants, average value of machinery per farm, lynchings per 1,000 blacks between 1900 and 1930, no. of Rosenwald schools per 1,000 blacks, share of acres flooded by the Mississippi in 1928, no. of slaves in 1860, Republican vote share, New Deal spending per capita 1933-35 (loans, public works, AAA, FHA loans), and the unemployment rate in 1937. Time-invariant controls are interacted with decade fixed effects. Monetary values are deflated to 2010 U.S. dollars. Standard errors clustered at the commuting zone level in parentheses. Significance levels are denoted by * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 1.B.3: Movers vs. Birth-State Stayers, South

Outcome:	ln(wage)	Education	Owns home	ln(house value)
	Panel A: Cross-State Migrants			
Casualty rate$_z \times$ Post-war$_t$	-0.029**	0.016	0.005	-0.048
	(0.015)	(0.071)	(0.006)	(0.032)
Casualty rate$_z \times$ Black$_{izt} \times$	0.063***	0.357***	-0.004	0.096***
Post-war$_t$	(0.008)	(0.037)	(0.004)	(0.017)
Observations	262,134	297,978	368,036	132,068
Adj. R^2	0.491	0.439	0.265	0.466
	Panel B: Birth-State Stayers			
Casualty rate$_z \times$ Post-war$_t$	-0.035***	-0.084**	-0.004	-0.049*
	(0.012)	(0.039)	(0.005)	(0.025)
Casualty rate$_z \times$ Black$_{izt} \times$	0.018***	0.309***	-0.011***	0.087***
Post-war$_t$	(0.007)	(0.030)	(0.002)	(0.011)
Observations	504,632	612,777	858,677	296,415
Adj. R^2	0.498	0.405	0.239	0.471

Note: Difference-in-differenece-in-differences regression of economic outcomes on the commuting zone WWII casualty rate among semi-skilled whites interacted with a post-WWII dummy, and with a black indicator for individuals living in 300 commuting zones in the U.S. South. The estimation sample contains data from the decennial U.S. micro Census from 1920-70 on non-institutionalized, working black and white males aged 15-65 who are not currently attending school. All regressions include commuting zone and Census year fixed effects. Owns home is a binary outcome for whether an individual owns their home. The log house value, log wages, and education variables are only available from 1940 onward. Log house value is also missing for 1950. Individual level controls include age, marital status, age and place of birth dummies. Commuting zone level controls are the WWII draft rate, log WWII spending per capita, share of black men, share of rural population, no. of manufacturing establishments per capita, average manufacturing firm size, log manufacturing value added per worker, share of employment in manufacturing, share of land in agricultural production, share of acres in cotton production, share of cash tenants, average value of machinery per farm, lynchings per 1,000 blacks between 1900 and 1930, no. of Rosenwald schools per 1,000 blacks, share of acres flooded by the Mississippi in 1928, no. of slaves in 1860, Republican vote share, New Deal spending per capita 1933-35 (loans, public works, AAA, FHA loans), and the unemployment rate in 1937. Time-invariant controls are interacted with decade fixed effects. Monetary values are deflated to 2010 U.S. dollars. Standard errors clustered at the commuting zone level in parentheses. Significance levels are denoted by * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

C NPPS Additional Results

C1) Robustness and Heterogeneity

C1.1: Splitting the Sample into Black and White Respondents

Tables 1.C.1 and 1.C.2 re-estimate the OLS and IV regressions for eq. (1.6) for the black and white samples, respectively. Given that the sample size is essentially halved, this is reflected in the very wide standard errors. The main aim of this exercise is to explore from which group the estimated effect sizes in the main table originate. In most cases the absolute size of the coefficients is larger in the sample of black respondents. However, comparing the coefficients to the sample means within each group shows that the relative magnitudes are comparable across blacks and whites. The only outcome where black and white respondents differ is the favor integration at church outcome which yields a slightly negative but close to zero IV coefficient for whites. This is the only result which is mainly driven by black respondents.

C1.2: Weighted Regressions

Despite the attempt by the authors of the initial study to produce a representative sample of the Southern population, blacks and whites were sampled in equal proportion. This does not reflect the population shares in their counties of residence. To account for this, table 1.C.3 weights black and white respondents by their population share in their residence county. This does not overturn the previous findings.

C1.3: Alternative Treatment Definition

Another concern is that the treatment change from 1940 to 1950 is not relevant for black-white social outcomes in 1961. I therefore re-estimate eq. (1.6) by taking the change from 1940 to 1960. While the instrument does gain strength, the point estimates are not significantly different from the main results. The results from this exercise are reported in table 1.C.4

Table 1.C.1: The Skill Upgrade and Black-White Social Relations - Black Sample

	Pr(Interracial Friend)=1		Pr(Live in Mixed Race Area)=1	
	(OLS)	(IV)	(OLS)	(IV)
Δsemi-skilled blacks$_c$	0.0325	0.0525	0.0125	0.0009
	$(0.0119)^{**}$	$(0.0159)^{***}$	(0.0155)	(0.0175)
	$[0.0189]^{*}$	$[0.0274]^{*}$	[0.0245]	[0.0255]
Outcome mean	0.4657	0.4657	0.1611	0.1611
R^2	0.1377	0.1359	0.2693	0.2683
	Pr(Favor Integration)=1		Pr(Favor Mixed Schools)=1	
	(OLS)	(IV)	(OLS)	(IV)
Δsemi-skilled blacks$_c$	0.0146	0.0267	0.0078	-0.0039
	(0.0091)	$(0.0140)^{*}$	(0.0059)	(0.0060)
	[0.0139]	[0.0244]	[0.0104]	[0.0082]
Outcome mean	0.6407	0.6407	0.0593	0.0593
R^2	0.2671	0.2664	0.1110	0.1084
	Pr(Favor Mixed Church)=1		Pr(Priest Pro Segregation)=1	
	(OLS)	(IV)	(OLS)	(IV)
Δsemi-skilled blacks$_c$	0.0049	0.0209	0.0046	-0.0119
	(0.0031)	$(0.0085)^{**}$	(0.0046)	$(0.0068)^{*}$
	[0.0042]	[0.0162]	[0.0055]	[0.0100]
Outcome mean	0.0574	0.0574	0.0611	0.0611
R^2	0.1015	0.0964	0.0497	0.0446

Note: The estimation sample is kept constant in all regressions with 540 black adults in 24 counties from Southern states in 1961 using data from the "Negro Political Participation Study" (Matthews and Prothro, 1975). The change in the share of blacks in semi-skilled employment from 1940 to 1950 (Δshare of blacks$_c$) in county c is instrumented with the WWII casualty rate among semi-skilled whites in that county. The first stage F-statistic is 22.905 and the Olea and Pflueger (2013) efficient F-statistic is 24.207. Individual level controls include gender, race, age, location of dwelling (urban, suburban, rural), years lived in current county, place size, veteran status, county where a respondent grew up, and state fixed effects. County level controls used are the share of blacks in semi-skilled jobs in 1940, the share of blacks in county c, share of people not born in county c, the WWII draft rate, and variables on racial sentiment such as the number of Rosenwald schools per 1,000 blacks, the number of lynchings from 1900-30 per 1,000 blacks, and the number of black slaves in 1860. Standard errors are clustered at the county level and are reported in parentheses. Standard errors corrected for the small cluster size using the wild cluster bootstrap-t procedure for OLS models by Cameron et al. (2008) and the wild restricted efficient residual bootstrap for IV models by Davidson and MacKinnon (2010) are reported in squared brackets. Significance levels are denoted by * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 1.C.2: The Skill Upgrade and Black-White Social Relations - White Sample

	Pr(Interracial Friend)=1		Pr(Live in Mixed Race Area)=1	
	(OLS)	(IV)	(OLS)	(IV)
Δsemi-skilled blacks$_c$	0.0207	0.0129	0.0135	0.0168
	$(0.0066)^{***}$	(0.0089)	$(0.0047)^{***}$	$(0.0046)^{***}$
	$[0.0090]^{**}$	$[0.0120]$	$[0.0058]^{**}$	$[0.0072]^{**}$
Outcome mean	0.5825	0.5825	0.0852	0.0852
R^2	0.1811	0.1800	0.3912	0.3906
	Pr(Favor Integration)=1		Pr(Favor Mixed Schools)=1	
	(OLS)	(IV)	(OLS)	(IV)
Δsemi-skilled blacks$_c$	0.0053	0.0017	0.0091	0.0068
	$(0.0022)^{**}$	(0.0033)	$(0.0019)^{***}$	$(0.0033)^{**}$
	$[0.0041]$	$[0.0046]$	$[0.0032]^{***}$	$[0.0048]$
Outcome mean	0.0360	0.0360	0.0455	0.0455
R^2	0.1632	0.1617	0.1213	0.1207
	Pr(Favor Mixed Church)=1		Pr(Priest Pro Segregation)=1	
	(OLS)	(IV)	(OLS)	(IV)
Δsemi-skilled blacks$_c$	0.0014	-0.0008	-0.0081	-0.0095
	(0.0011)	(0.0014)	$(0.0044)^{*}$	$(0.0045)^{**}$
	$[0.0020]$	$[0.0025]$	$[0.0065]$	$[0.0066]$
Outcome mean	0.0114	0.0114	0.1420	0.1420
R^2	0.1298	0.1279	0.1973	0.1973

Note: The estimation sample is kept constant in all regressions with 528 white adults in 24 counties from Southern states in 1961 using data from the "Negro Political Participation Study" (Matthews and Prothro, 1975). The change in the share of blacks in semi-skilled employment from 1940 to 1950 (Δshare of blacks$_c$) in county c is instrumented with the WWII casualty rate among semi-skilled whites in that county. The first stage F-statistic is 54.895 and the Olea and Pflueger (2013) efficient F-statistic is 57.400. Individual level controls include gender, race, age, location of dwelling (urban, suburban, rural), years lived in current county, place size, veteran status, county where a respondent grew up, and state fixed effects. County level controls used are the share of blacks in semi-skilled jobs in 1940, the share of blacks in county c, share of people not born in county c, the WWII draft rate, and variables on racial sentiment such as the number of Rosenwald schools per 1,000 blacks, the number of lynchings from 1900-30 per 1,000 blacks, and the number of black slaves in 1860. Standard errors are clustered at the county level and are reported in parentheses. Standard errors corrected for the small cluster size using the wild cluster bootstrap-t procedure for OLS models by Cameron et al. (2008) and the wild restricted efficient residual bootstrap for IV models by Davidson and MacKinnon (2010) are reported in squared brackets. Significance levels are denoted by * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 1.C.3: The Skill Upgrade and Black-White Social Relations - Weighted Regressions

	$\Pr$(Interracial Friend)=1		$\Pr$(Live in Mixed Race Area)=1	
	(OLS)	(IV)	(OLS)	(IV)
Δsemi-skilled blacks$_c$	0.0202	0.0160	0.0153	0.0149
	$(0.0060)^{***}$	$(0.0074)^{**}$	$(0.0053)^{***}$	$(0.0049)^{***}$
	$[0.0081]^{**}$	$[0.0098]$	$[0.0079]^{*}$	$[0.0086]^{*}$
Outcome mean	0.5235	0.5235	0.1236	0.1236
R^2	0.1486	0.1483	0.1692	0.1692
	$\Pr$(Favor Integration)=1		$\Pr$(Favor Mixed Schools)=1	
	(OLS)	(IV)	(OLS)	(IV)
Δsemi-skilled blacks$_c$	0.0070	0.0117	0.0093	0.0091
	$(0.0030)^{**}$	$(0.0044)^{***}$	$(0.0019)^{***}$	$(0.0031)^{***}$
	$[0.0053]$	$[0.0073]$	$[0.0035]^{***}$	$[0.0044]^{**}$
Outcome mean	0.3418	0.3418	0.0524	0.0524
R^2	0.5162	0.5157	0.0796	0.0796
	$\Pr$(Favor Mixed Church)=1		$\Pr$(Priest Pro Segregation)=1	
	(OLS)	(IV)	(OLS)	(IV)
Δsemi-skilled blacks$_c$	0.0024	0.0034	-0.0068	-0.0123
	$(0.0012)^{*}$	$(0.0014)^{**}$	(0.0042)	$(0.0055)^{**}$
	$[0.0020]$	$[0.0021]^{*}$	$[0.0060]$	$[0.0084]$
Outcome mean	0.0346	0.0346	0.1011	0.1011
R^2	0.0788	0.0787	0.1525	0.1515

Note: The estimation sample is kept constant in all regressions with 540 black and 528 white adults in 24 counties from Southern states in 1961 using data from the "Negro Political Participation Study" (Matthews and Prothro, 1975). The change in the share of blacks in semi-skilled employment from 1940 to 1950 (Δshare of blacks$_c$) in county c is instrumented with the WWII casualty rate among semi-skilled whites in that county. Observations are weighted by the respondent's racial group's population share in their county. The first stage F-statistic is 43.799 and the Olea and Pflueger (2013) efficient F-statistic is 45.841. Individual level controls include gender, race, age, location of dwelling (urban, suburban, rural), years lived in current county, place size, veteran status, county where a respondent grew up, and state fixed effects. County level controls used are the share of blacks in semi-skilled jobs in 1940, the share of blacks in county c, share of people not born in county c, the WWII draft rate, and variables on racial sentiment such as the number of Rosenwald schools per 1,000 blacks, the number of lynchings from 1900-30 per 1,000 blacks, and the number of black slaves in 1860. Standard errors are clustered at the county level and are reported in parentheses. Standard errors corrected for the small cluster size using the wild cluster bootstrap-t procedure for OLS models by Cameron et al. (2008) and the wild restricted efficient residual bootstrap for IV models by Davidson and MacKinnon (2010) are reported in squared brackets. Significance levels are denoted by * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 1.C.4: The Skill Upgrade and Black-White Social Relations - 1940 to 1960 Differenced Treatment

	Pr(Interracial Friend)=1		Pr(Live in Mixed Race Area)=1	
	(OLS)	(IV)	(OLS)	(IV)
Δsemi-skilled blacks$_c$	0.0132	0.0133	0.0105	0.0088
	$(0.0049)^{**}$	$(0.0059)^{**}$	$(0.0035)^{***}$	$(0.0037)^{***}$
	$[0.0071]^{*}$	$[0.0079]^{*}$	$[0.0048]^{**}$	$[0.0059]$
Outcome mean	0.5235	0.5235	0.1236	0.1236
R^2	0.1202	0.1202	0.1380	0.1379
	Pr(Favor Integration)=1		Pr(Favor Mixed Schools)=1	
	(OLS)	(IV)	(OLS)	(IV)
Δsemi-skilled blacks$_c$	0.0099	0.0157	0.0053	0.0077
	$(0.0030)^{***}$	$(0.0043)^{***}$	$(0.0025)^{**}$	$(0.0025)^{***}$
	$[0.0053]^{*}$	$[0.0087]^{*}$	$[0.0041]$	$[0.0036]^{**}$
Outcome mean	0.3418	0.3418	0.0524	0.0524
R^2	0.5102	0.5096	0.0639	0.0634
	Pr(Favor Mixed Church)=1		Pr(Priest Pro Segregation)=1	
	(OLS)	(IV)	(OLS)	(IV)
Δsemi-skilled blacks$_c$	0.0033	0.0056	-0.0041	-0.0108
	$(0.0010)^{***}$	$(0.0012)^{***}$	(0.0033)	$(0.0049)^{**}$
	$[0.0015]^{**}$	$[0.0019]^{***}$	$[0.0040]$	$[0.0077]$
Outcome mean	0.0346	0.0346	0.1011	0.1011
R^2	0.0808	0.0802	0.1189	0.1169

Note: The estimation sample is kept constant in all regressions with 540 black and 528 white adults in 24 counties from Southern states in 1961 using data from the "Negro Political Participation Study" (Matthews and Prothro, 1975). The change in the share of blacks in semi-skilled employment from 1940 to 1960 (Δshare of blacks$_c$) in county c is instrumented with the WWII casualty rate among semi-skilled whites in that county. The first stage F-statistic is 86.147 and the Olea and Pflueger (2013) efficient F-statistic is 90.164. Individual level controls include gender, race, age, location of dwelling (urban, suburban, rural), years lived in current county, place size, veteran status, county where a respondent grew up, and state fixed effects. County level controls used are the share of blacks in semi-skilled jobs in 1940, the share of blacks in county c, share of people not born in county c, the WWII draft rate, and variables on racial sentiment such as the number of Rosenwald schools per 1,000 blacks, the number of lynchings from 1900-30 per 1,000 blacks, and the number of black slaves in 1860. Standard errors are clustered at the county level and are reported in parentheses. Standard errors corrected for the small cluster size using the wild cluster bootstrap-t procedure for OLS models by Cameron et al. (2008) and the wild restricted efficient residual bootstrap for IV models by Davidson and MacKinnon (2010) are reported in squared brackets. Significance levels are denoted by * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

C2) Sensitivity of IV Results to Small Violations of the Exclusion Restriction

The typical IV framework in eq. (1.6) assumes that the instrument does not have a direct partial effect on the outcome such that in,

$$\text{social outcome}_{ic} = \phi\Delta\text{share of blacks}_c + \gamma_z\text{casualty rate} + X_{ic}'\lambda + \epsilon_{ic} \tag{1.13}$$

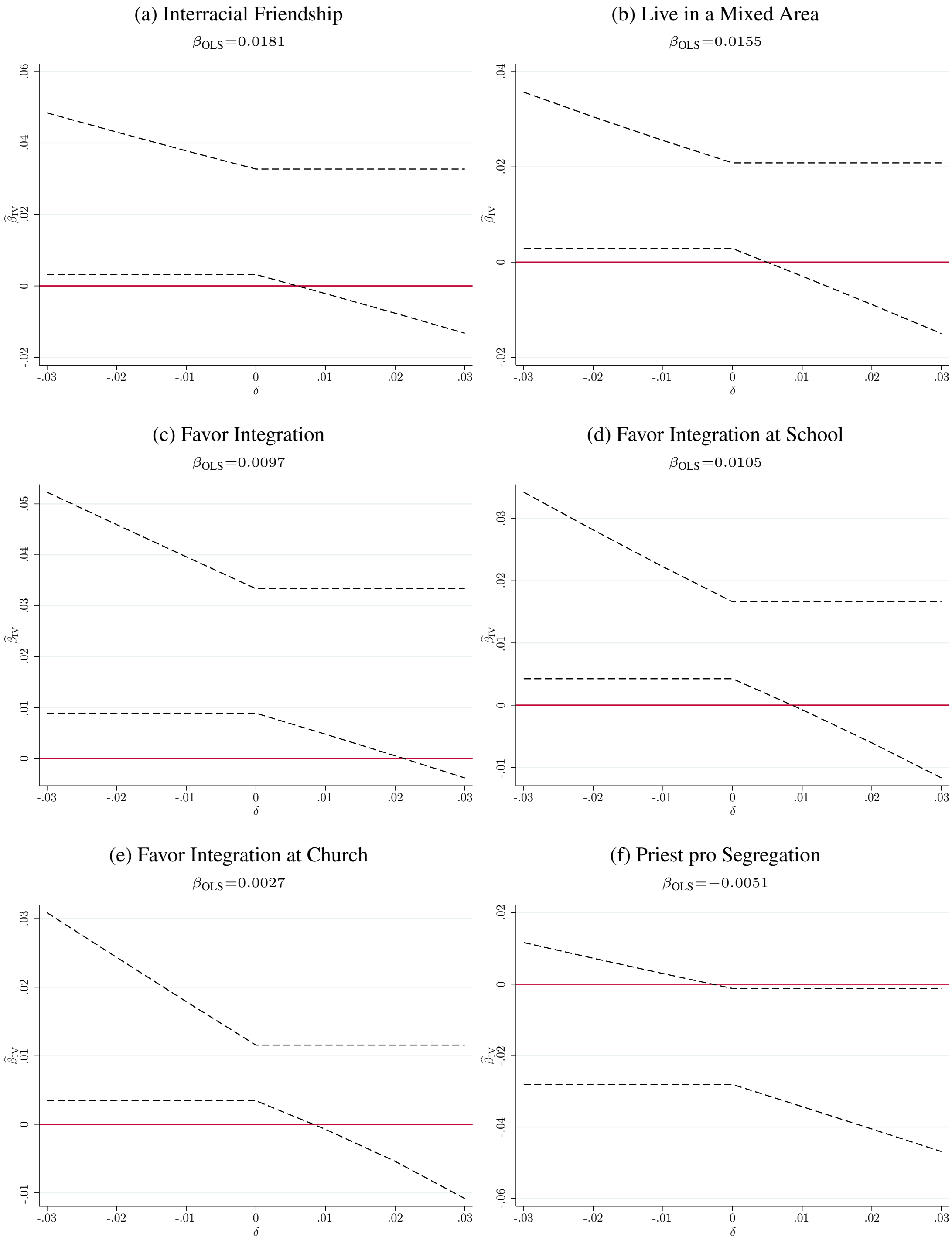
the coefficient $\gamma_z = 0$ in the structural model. While this assumption cannot be directly tested, Conley et al. (2012) construct a bounding exercise which tests the sensitivity of IV estimates with respect to small violations of the exclusion restriction. A small violation means that the instrument is not perfectly exogenous but "plausibly exogenous", i.e. $\gamma_z \neq 0$ but is close to zero.

For this test, the econometrician needs to specify a range of possible values that γ_z can take with $\gamma_z \in [-\delta, \delta]$ for some δ. Their union of confidence intervals (UCI) procedure re-estimates eq. (1.13) for every value of γ_z in the specified range which allows to place bounds on β_{IV} in eq. (1.6). These then provide 95% confidence intervals for the value that β_{IV} could take under a given size of the violation.

A main disadvantage of this method is that the bounds may be wide. In principle, they can be tightened by providing further structure on the distribution of γ_z. For the sake of this sensitivity analysis I refrain from imposing such structural assumptions and provide the most conservative bounds instead. The plots for the sensitivity analysis are shown in figure 1.C.1 for each of the considered outcomes for $\delta = 0.5$. The figure reports the corresponding OLS coefficients for comparison.

For instance, the outcome on interracial friendships tolerates a direct partial effect of the instrument on the outcome of 2.5 p.p. before the IV estimate cannot be distinguished from zero at the 95% level. A coefficient of 2.5 p.p. for the instrument would be 29% of the corresponding OLS coefficient, hence one might not regard this as "small" violation of the exclusion restriction but rather a large direct partial effect of the instrument that would be required to threaten set identification. For the outcome on interracial friendships at work the bounds are less forgiving and already make the IV indistinguishable from zero for a small positive instrument coefficient in absolute terms.

Figure 1.C.1: Conley et al. (2012) IV Bounds

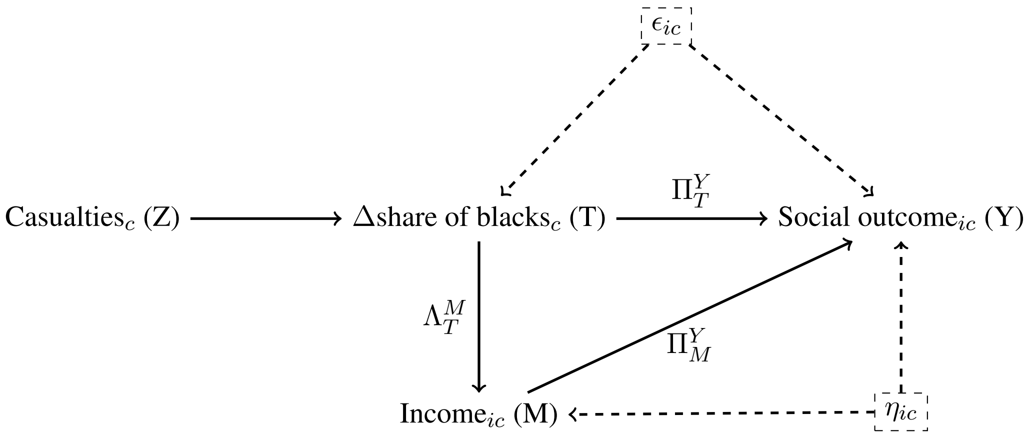


Note: Conley et al. (2012) bounds on the IV coefficients from regressing each outcome (a)-(f) on the change in the share of semi-skilled blacks in county c from 1940 to 1950 using individual level data from the "Negro Political Participation Study" (Matthews and Prothro, 1975) for 540 black and 528 white adults in 24 counties in Southern states in 1961. The change in the share of semi-skilled blacks is instrumented with the WWII casualty rate among semi-skilled whites. The bounds are constructed to allow for a non-zero direct partial effect of the instrument (γ_z) on each outcome where an interval of plausible ranges of this coefficient is chosen as $\gamma_z \in [-\delta, \delta]$ with $\delta = 0.3$. To make values of γ_z for which $\widehat{\beta}_{\text{IV}}$ cannot be distinguished from zero comparable, I report the baseline OLS coefficients under each outcome heading. The bounds provide 95% confidence intervals within which $\widehat{\beta}_{\text{IV}}$ can be estimated for small violations of the exclusion restriction. Standard errors are clustered at the county level.

C3) Mediation Effects Through Income

There are potentially several mechanisms behind the effect of the occupational upgrade of blacks on social outcomes. One channel to be considered here is the effect of increased incomes due to employment in higher paying jobs. The main analysis did not include incomes in the regressions. In the previous context, this would have been a bad control, i.e. a control variable which is also an outcome of the treatment (the black occupational upgrade). To test how much of the effect of the occupational upgrade on social outcomes comes from increases in incomes, I use the causal mediation framework introduced by Dippel, Gold, Heblich and Pinto (2017).

Figure 1.C.2: Directed Acyclical Graph for Causal Mediation Effects



Note: Causal mediation analysis schematic. The treatment T, which is instrumented with Z, has a total effect on the outcome Y which can be decomposed into its direct effect Π_T^Y, and its indirect effect through a mediator variable M. This indirect effect is the product of the effect of T on M (Λ_T^M) and the effect of M on Y (Π_M^Y). Solid lines connect observables, dashed lines unobservables such as the two error terms ϵ and η which guide the (potential) endogeneity of T and M.

The idea of the framework is illustrated in figure 1.C.2. The standard IV model is nested in this framework in which the casualty rate instrument Z affects the social outcome Y through the change in the share of blacks in semi-skilled jobs treatment T. Potential endogeneity of T comes from a correlation with the error ϵ. Unlike in the standard framework, which assumes a single causal channel, the treatment may also partially affect Y through its effect on incomes, the so-called mediator (M). A particularly appealing feature of the Dippel et al. (2017) framework is that is allows for M to be potentially endogenous through a correlation with a second error term, η.

They show that the total effect of Δshare of blacks$_c$, instrumented by the casualty rate, on the outcome can be decomposed as,

$$\underbrace{\Lambda_T^Y}_{\text{total effect}} = \underbrace{\Pi_T^Y}_{\text{direct effect}} + \underbrace{\Pi_M^Y \times \Lambda_T^M}_{\text{indirect effect}} \tag{1.14}$$

where Λ_T^M is the second stage coefficient from the IV regression of M on T using Z as instrument. Π_M^Y is the second stage coefficient from the IV regression of Y on M using Z as instrument, conditioning on T. The same regression identifies Π_T^Y which is the second stage coefficient on T.

In addition to the standard identifying assumptions, consistent estimation of the causal effect of T on Y and the causal mediation effect of M on Y requires the exclusion restriction $Z \perp\!\!\!\perp M$ and that $\epsilon \perp\!\!\!\perp \eta$. Suppose workers dislike blacks and try to keep them out of semi-skilled employment via union involvement and that factory owners dislike blacks and hence are neither friends with them, nor would they pay fair wages. This would be a case in which the two error terms are potentially correlated. Given that such a scenario is far from impossible, the required assumption on the error correlations might be very strong.

Table 1.C.5 shows the results from this causal mediation analysis. The table displays the total effect Λ_T^Y, which can be compared to previous regression results, and the share of this total effect which is mediated through the effect of the occupational upgrade on blacks' incomes, $\frac{\Pi_M^Y \times \Lambda_T^M}{\Lambda_T^Y}$. The results show that income does not matter at all in the determination of interracial friendships. The effect is therefore likely driven by other mediators which have not been explored or are unobserved. An example of another potential mediator is exposure of black and white workers in the factories or at clubs or other social activities which are available in the cities.

The mediation effect is larger for other outcomes, such as attitudes towards integration for which 46% of the occupational upgrade effect are mediated through income. The same holds for favoring integration at church with a mediation effect of 58.6% of the total effect, and for the probability that a respondent's priest preaches in favor of segregation (62.2%). However, it should also be noted that none of these mediation effects are estimated precisely enough as that they could be taken as statistically significantly different from zero.

Table 1.C.5: Causal Mediation Analysis Results

	Pr(Interracial Friend)=1	Pr(Live in Mixed Race Area)=1
Δsemi-skilled blacks$_c$	0.018^{**}	0.011^{**}
	(0.023)	(0.029)
% mediated through income	0.001	-0.442
	(0.998)	(0.344)
	Pr(Favor Integration)=1	Pr(Favor Mixed Schools)=1
Δsemi-skilled blacks$_c$	0.020^{***}	0.011^{***}
	(0.001)	(0.001)
% mediated through income	0.460	0.026
	(0.203)	(0.909)
	Pr(Favor Mixed Church)=1	Pr(Priest Pro Segregation)=1
Δsemi-skilled blacks$_c$	0.008^{***}	-0.013^{*}
	(0.000)	(0.052)
% mediated through income	0.586	0.622
	(0.186)	(0.274)

Note: The estimation sample is kept constant in all regressions with 540 black and 528 white adults in 24 counties from Southern states in 1961 using data from the "Negro Political Participation Study" (Matthews and Prothro, 1975). The change in the share of blacks in semi-skilled employment from 1940 to 1950 (Δshare of blacks$_c$) in county c is instrumented with the WWII casualty rate among semi-skilled whites in that county. The table displays the percentage share of this estimated main effect that is mediated through increased incomes of blacks due to the skill upgrade from low- to semi-skilled occupations. Controls include gender, race, age, location of dwelling (urban, suburban, rural), years lived in current county, place size, veteran status, county where a respondent grew up, and state fixed effects. County level controls used are the share of blacks in semi-skilled jobs in 1940, the share of blacks in county c, share of people not born in county c, the WWII draft rate, and variables on racial sentiment such as the number of Rosenwald schools per 1,000 blacks, the number of lynchings from 1900-30 per 1,000 blacks, and the number of black slaves in 1860. Standard errors are clustered at the county level, p-values reported in parentheses.

D Data Appendix

Merging Enlistment and Casualty Records

Merging the 8.3 million observations from the WWII Army enlistment records with the casualty records based on the Army serial number matches 78% of all casualties. These are observations which found a unique match across both data sets. For robustness I computed the soundex string distance of first- and surname and kept those matches for which it was sufficiently small in order to be sure that the match was correct. Less than one percent of these initial matches were returned to the pool of unmatched observations because of significant differences in the names that indicated a clear mismatch despite a perfect match on the serial number. The match rate is not perfect because of mistakes in the serial number made by the Optical Character Recognition (OCR) software on part of the casualty tables for which the scans are of less than ideal quality.

The remaining casualties were matched via the probabilistic string matching algorithms provided by Wasi and Flaaen (2015). A one-to-one match was used to link each casualty with a potential enlistment record based on name and serial number stratified by state of residence. Names are matched via a tokenization and serial numbers via a bigram algorithm. The match with the highest combined matching score was kept. This results in a final match rate of 94%. From a random sample of 1,000 matches the error rate was 0.6% as judged by correctness of the name, serial number, and residence. The OCR quality of the remaining 6% of casualty observations was too poor in order to clearly identify whether a given match was correct. These cases were dropped.

Sources of the U.S. Census County Data, 1920-1970

The main data source are the county aggregates of the U.S. Decennial Census of Population and Housing from 1940 to 1970 and the 100% full count micro data of the Census. For the years 1940 to 1970, the Census publishes occupational counts at the county level where Southern states report them separated for black and white workers. For instance, see table 23a on page 278 of the 1940 Census for Georgia shown in figure 1.D.1 which are the raw data from which I digitized the employment information at the county level for blacks by county and skill group. Occupations are defined according to the harmonized 1950 definition by the U.S. Census Bureau. The categories include professional, semi-professional, farmers, proprietors and managers, clerical and sales, craftsmen and foremen, operatives, domestic services, farm laborers, and laborers. Semi-skilled occupations here are taken to be the groups of craftsmen and operatives. These definitions change considerably with the 1980 Census which makes it impossible to keep a consistent measurement of the outcome variable.

Figure 1.D.1: Data Source for Semi-Skilled Employment of Blacks

278 CHARACTERISTICS OF THE POPULATION

Table 23a.—NONWHITE EMPLOYED WORKERS 14 YEARS OLD AND OVER, BY MAJOR OCCUPATION GROUP AND SEX, BY COUNTIES: 1940

COUNTY AND SEX		Total employed (except on public emergency work)	Professional workers	Semiprofessional workers	Farmers and farm managers	Proprietors, managers, and officials, exc. farm	Clerical, sales, and kindred workers	Craftsmen, foremen, and kindred workers	Operatives and kindred workers	Domestic service workers	Service workers, except domestic	Farm laborers (wage workers) and foremen	Farm laborers, unpaid family workers	Laborers, except farm	Occupation not reported
Appling	Male	646	8	-	105	3	3	17	59	7	27	93	23	300	1
	Female	252	22	-	12	1	-	-	2	160	20	7	5	-	3
Atkinson	Male	504	3	-	44	3	-	3	19	1	7	54	12	355	3
	Female	136	8	-	2	-	-	-	1	106	8	2	5	1	3
Bacon	Male	288	1	-	48	3	11	6	16	2	4	23	5	168	1
	Female	92	4	-	3	-	-	-	1	67	7	3	4	2	1
Baker	Male	1,151	1	-	543	-	-	3	6	4	1	319	203	69	2
	Female	360	15	-	45	-	-	-	1	75	3	40	179	-	2
Baldwin	Male	1,984	23	3	392	14	13	141	147	50	248	306	124	515	8
	Female	1,484	67	-	33	2	3	2	43	923	241	41	119	2	8
Banks	Male	150	-	-	90	-	-	-	1	3	3	21	28	3	1
	Female	25	1	-	5	-	-	-	-	7	-	-	12	-	-
Barrow	Male	585	2	-	202	1	-	9	26	10	33	196	64	39	3
	Female	373	9	-	4	4	-	-	1	250	13	21	71	-	-
Bartow	Male	947	11	-	213	4	3	19	151	43	70	170	59	195	9
	Female	631	28	-	9	1	2	-	5	496	38	12	36	1	3
Ben Hill	Male	1,117	15	3	204	16	10	47	153	6	59	270	65	265	4
	Female	608	26	-	13	14	2	-	10	489	33	11	4	2	4
Berrien	Male	683	4	-	74	5	5	17	49	9	9	149	23	337	2
	Female	274	3	-	4	2	-	-	2	202	12	33	15	-	1
Bibb	Male	7,379	138	14	218	115	170	640	1,924	263	947	547	53	2,311	39
	Female	6,626	231	4	18	51	58	19	443	4,858	754	29	37	101	23

Note: Raw data source from the 1940 Census of Population and Housing for the state of Georgia (p. 278). Occupational information is reported for each skill group by county and gender.

Before 1940 the county level aggregates do not report these statistics. However, it is possible to construct them from the 100% full count micro data of the Census for 1920, 1930, and 1940. Before 1920 there is no reliable employment status data. This information is important to construct the correct county aggregates. For each county, these are the sum of all currently employed workers in a given occupational group. The emphasis lies on currently employed. Given the overlap of the full count Census and the county level aggregates in 1940, this is the only definition of workers which gives a complete overlap between the two data sources with respect to the constructed and the actual county level data.

The difference-in-differences results in table 1.3 and the related tables are not driven by potential definitional mistakes. Table 1.D.1 shows that the estimated results largely unchanged when using the county level aggregates for 1940 to 1970 only. The specification with covariates fixed at their 1940 levels estimates a slightly smaller effect while inclusion of the county-specific time trends takes away more significance. This is mostly due to the reduced size of the pre-treatment time window but the coefficient remains as before.

Table 1.D.1: County Level Difference-in-Differences Results, 1940-1970

	Outcome: % blacks in semi-skilled jobs (pre-war mean = 12.433)					
	(1)	(2)	(3)	(4)	(5)	(6)
Casualty rate$_c \times$ Post-war$_t$	0.529***	0.617***	0.343***	0.586***	0.534*	0.552***
	(0.117)	(0.155)	(0.132)	(0.162)	(0.285)	(0.123)
Controls		Yes		Yes	Yes	Yes
1940 controls $\times$ time			Yes			
Flexible state time trends				Yes		
Linear county time trends					Yes	
Doubly-robust selection						Yes
Observations	4,985	3,626	3,684	3,626	3,626	4,655
Counties	1,388	1,229	985	1,229	1,229	1,377
Adj. R^2	0.885	0.901	0.905	0.908	0.919	0.880
Oster's δ	0.951	1.023	0.545	1.109	0.599	0.996

Note: Difference-in-differences regressions of the county-level share of blacks in semi-skilled occupations on the WWII county casualty rate among semi-skilled whites interacted with a post-war indicator. The estimation sample contains decennial U.S. Census data on counties in Southern states from 1940 to 1970. Controls include county and decade fixed effects, the county draft rate, average casualty rate in the neighboring counties, log WWII spending per capita, share of black men, share of rural population, log median family income, share of pop. with high school degree, no. of manufacturing establishments per capita, average manufacturing firm size, log manufacturing value added per worker, share of employment in manufacturing, share of land in agricultural production, share of acres in cotton production, share of cash tenants, average value of machinery per farm, lynchings per 1,000 blacks between 1900 and 1930, no. of Rosenwald schools per 1,000 blacks, share of acres flooded by the Mississippi in 1928, no. of slaves in 1860, Republican vote share, New Deal spending per capita 1933-35 (loans, public works, AAA, FHA loans), and the unemployment rate in 1937. Time-invariant controls are interacted with decade fixed effects. Monetary values are deflated to 2010 U.S. dollars. The doubly-robust selection method implements the Belloni et al. (2014) machine learning covariate selection algorithm for testing the stability of treatment effects with respect to the observables. Oster's (2017) test for selection on unobservables is reported in the final row by computing the coefficient of proportionality δ for which the coefficient on the semi-skilled casualty rate among whites would equal zero. Standard errors clustered at the county level. Significance levels are denoted by * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

The Census data also contain information on each county's population but also on the local economies. This includes information on the number of manufacturing establishments, number of manufacturing workers, and value added. From the I compute the following controls:

Manufacturing firms per 1,000 pop $= \frac{\text{No. manufacturing establishments}_{ct}}{\text{Total population}_{ct}/1{,}000}$

Av. manufacturing firm size $= \frac{\text{Total manufacturing workers}_{ct}}{\text{No. manufacturing establishments}_{ct}}$

Manufact. value added per worker $= \ln\left(1 + \frac{\text{Total manufacturing value added}_{ct}}{\text{Total manufacturing workers}_{ct}}\right)$

Share of manufacturing workers $= \frac{\text{Total manufacturing workers}_{ct} \times 100}{\text{Total population}_{ct}}$

Share of black men $= \frac{\text{Total no. of black men}_{ct} \times 100}{\text{Total no. of men}_{ct}}$

Share of blacks $= \frac{\text{Total no. of blacks}_{ct} \times 100}{\text{Total population}_{ct}}$

Data on the number of slaves in 1860 by county come from the 1860 U.S. Decennial Census of Population and Housing. Additionally, information on median family income was taken from the Census files. For 1940, the median family income was computed from the 1940 100% Census micro data. Whenever information on manufacturing or income variables was not available or incomplete in the Census, these were supplemented with information from the County and City Data Books from 1947 to 1972 published by the U.S. Census Bureau.

Control Variables

Agricultural Controls

Information on agricultural variables at the county level for each decade was taken from the U.S. Agricultural Census prepared by:

- Haines, M., Fishback, P.V., and Rhode, P. (2016) "United States Agriculture Data, 1840 - 2012", Study No. ICPSR35206-v3, Inter-university Consortium for Political and Social Research 2016-06-29, Ann Arbor, MI

Constructed variables from this data set are:

acres in farm land $= \frac{\text{farm acres}_{ct} \times 100}{\text{land acres}_{ct}}$

average value of machinery per farm $= \frac{\text{value of farm machinery}_{ct} \times \text{CPI}_t}{\text{No. farms}_{ct}}$

share of cash tenants$_{ct}$ $= \frac{\text{No. cash tenants}_{ct} \times 100}{\text{Total no. tenant farmers}_{ct}}$

share of cotton in agriculture$_{ct}$ $= \frac{\text{No. acres in cotton production}_{ct} \times 100}{\text{Acres in farm land}_{ct}}$

Lynchings

Data on the number of lynchings for a given county between 1900 and 1930 come from Project HAL: Historical American Lynching. Their definition of a lynching follows the conditions outlined by the National Association for the Advancement of Colored People (NAACP). The conditions for a murder to qualify as lynching are that there must be evidence that someone was killed; the killing must have occurred illegally; three or more persons must have taken part in the killing; and the murderers must have claimed to serve tradition or justice. The lynchings variable here is defined as: $\frac{\text{No. lynchings 1900-1930}_c}{\text{No. of black pop}_{ct}/1{,}000}$. The data are freely available at:

- http://people.uncw.edu/hinese/HAL/HAL%20Web%20Page.htm
  (retrieved on November 2nd, 2017)

Mississippi Flooded Acres, 1928

This data comes from the data deposit by Hornbeck and Naidu (2014) at the American Economic Review website. The variable used here is defined as: $\frac{\text{flooded acres}_{c,1928}\times 100}{\text{total acres}_{c,1930}}$. The data can be accessed at:

- https://www.aeaweb.org/aer/data/10403/20120980_data.zip
  (retrieved on November 3rd, 2017)

Party Vote Shares

Data on the Republican vote share come from:

- Clubb, J.M., Flanigan, W.H., and Zingale, N.H. (2006) "Electoral Data for Counties in the United States: Presidential and Congressional Races, 1840-1972", ICPSR08611-v1. Ann Arbor, MI: Inter-university Consortium for Political and Social Research [distributor], 2006-11-13. https://doi.org/10.3886/ICPSR08611.v1

The data report congressional and presidential vote share by party for each election between 1840 and 1972. The Republican vote share here is taken to be the share of votes obtained by the Republican party in congressional elections in a Census year. If there was no election in given Census year, the nearest election was assigned.

Rosenwald Schools

The Rosenwald School variable here is defined as: $\frac{\text{No. Rosenwald Schools}_c}{\text{No. of black pop}_{ct}/1{,}000}$.
The number of Rosenwald Schools per county was obtained from:

- http://rosenwald.fisk.edu/index.php
  (retrieved on November 2nd, 2017)

WWII Related Spending

War related spending during World War II was taken from the 1947 County and City Data Book. A digital version is provided by:

- United States Department of Commerce. Bureau of the Census. “County and City Data Book [United States] Consolidated File: County Data, 1947-1977. ICPSR07736-v2”. Ann Arbor, MI: Inter-university Consortium for Political and Social Research [distributor], 2012-09-18. https://doi.org/10.3886/ICPSR07736.v2

The war related spending per capita variable here is computed as:

$$\text{Log mil. spending per capita} = \ln\left(1 + \frac{\left(\text{\$ combat equip.}+\text{\$ other equip.}+\text{\$ ind. facilities}+\text{\$ milfacilities}\right)_{c,1940}}{\text{Total population}_{c,1940}}\right)$$

Chapter 2

The Long-Term Effects of Losing a Father in the U.S. Civil War

2.1 Introduction

In which family we were born, who our parents are, and what they do are the first determinants of the economic, social, and cultural capital we are endowed with. What happens when a parent suddenly disappears and how does this affect the transmission of these various kinds of capital? This question is particularly relevant for more than 140 million children worldwide who lost either or both parents (UNICEF Press Center, 2017). To answer this question, we use military records from the U.S. Civil War with Census data to track children of soldiers over time, comparing those who lost a father during the war to those who did not.

Childhood is a particularly crucial formative period of our lives. There is now a large literature on the benefits of early childhood interventions in education. Cunha, Heckman and Lochner (2006) provide a literature review and a conceptual framework. Overall economic conditions during childhood matter, as shown by Feigenbaum (2016). Family structure, the number of members present in the household, also plays a major role in determining a child's later life outcomes. Chetty, Hendren, Kline and Saez (2014) show that the fraction of children in single-parent households is the strongest correlate of income-mobility in the United States.

To study the economic impact of family structure, researchers rely on disruptive events such as divorce (Painter and Levine, 2000; Corak, 2001; Gruber, 2004), imprisonment (Bhuller, Dahl, Loken and Mogstad, 2018; Dobbie, Gronqvist, Niknami, Palme and Priks, 2018), or death. The global increase in AIDS related deaths during the 1990s and early 2000s prompted many development economists to study the plight of orphans (Case, Paxson and Ableidinger, 2002; Gertler, Levine and Ames, 2004; Ainsworth and Filmer, 2006; Evans and Miguel, 2007;

Beegle, De Weerdt and Dercon, 2009, 2010; Senne, 2014). The main challenge of this research is that bereaved families are often of lower socioeconomic status, making it difficult to attribute the impact on children's later-life outcomes to the loss of a parent alone when parental deaths are not random.[1]

Our paper uses the demographic shock of the U.S. Civil War, the deadliest war in U.S. history with a death toll of over 650,000 men, to identify the causal effect of losing a parent on later life outcomes. The paper most closely related to our work is Kovac (2017), who studies the orphans of the 1991-1995 Croatian-Serbian war. He uses the within military unit variation in mortality and finds strong negative effects of paternal death on children's high-school GPA, school attendance, and health outcomes.

One important advantage of focusing on a historical event more than 150 years in the past is that it allows us to study long-term socio-economic outcomes, while the literature typically focuses on educational outcomes in the short term.[2] Few studies have used historical data for this purpose, one recent exception being Maloney and Smith (2018) who use the Utah Population Database to estimate the correlates of paternal death and son's economic outcomes in the early 20th century.

In this paper, we link military records for the 2.2 million Union Army soldiers to the full-count 1860 U.S. Census to identify fathers who fought and died or survived during the Civil War. We then follow the sons of Union Army soldiers into adulthood by linking the 1860 Census to the 1880 Census. The focus on fathers and sons is due to technical reasons. Since women change their surnames upon marriage, they cannot be easily linked across Censuses without additional information from birth or marriage certificates.

After matching our military data with the 1860 Census, we identify 41,831 men who were sons of soldiers in 1860 and who we can observe in 1880.[3] 13.8% of observed children lost their father during the war and 12.4% had a father who returned home with a disability. We also obtain a sample of 3,130 grandchildren of Union Army soldiers in the 1900 Census which allows us to test for the intergenerational effects of this paternal death shock.

To limit issues relating to selection into the military, we focus on the children of Union Army soldiers and compare the children of soldiers who died to the children of soldiers who

[1] The same problem exists for the literature on divorce and father imprisonment. Corak (2001) and Gruber (2004) use changes in divorce laws in Canada and the U.S., while Bhuller et al. (2018) and Dobbie et al. (2018) take advantage of the random allocation of judges in Norway and Sweden.

[2] The longest of these studies follows children for 13 years (Beegle et al., 2010), hence orphan studies using modern data mostly focus on educational and health outcomes as other outcomes such as fertility, occupational choice, and others have not yet materialized.

[3] The lower number of soldier-fathers relative to the total soldier population is explained by the fact most soldiers were young. The median age at the time of enlistment was 23 years. Hence most of these soldiers simply were too young to already have a family when they joined the Union Army.

survived. We start by exploring the socioeconomic correlates of soldiers' death and disability probabilities. Socioeconomic variables and military information such as date of enlistment, rank, and regiment type predict the probability to die or exit the military with a disability, but they leave most of the variation in death and disability unexplained.

In a first empirical strategy, we regress socioeconomic outcomes in 1880 on indicators for a child's father's death and father's disability controlling for a rich set of father socioeconomic characteristics measured in 1860 (including literacy, occupational score, wealth and whether the father was born abroad), as well as a polynomial in enlistment date, enlistment rank fixed effects, regiment type fixed effects, and state fixed effects. In a second empirical strategy, we instrument death by participation of the father's regiment in one of the 10 bloodiest battles of the war.

We find that losing a father in the Civil War decreased sons' occupational score in 1880 by 5% of a standard deviation. The result is mainly driven by the increased downward mobility of the sons of semi-skilled fathers, who are more likely to have a low-skilled occupation when they lose their father. We also assess the effect of having a disabled father. A returning but disabled father has a negative but smaller effect on their sons' socioeconomic status as compared to fathers who died. We find that effects are concentrated in the bottom half of the 1860 wealth distribution. Finally, the negative effects of a father's death or disability are transmitted to the generation of the grandchildren.

Our paper contributes to several strands of the literature relating to the economic consequences of parental loss or absence. Results on the effects of divorce on children are mixed: Painter and Levine (2000) and Gruber (2004) find that children of divorced parents have lower incomes, while Corak (2001) finds no effect of divorce on the future earnings of children. The same holds for the intergenerational effects of father incarceration. Dobbie et al. (2018) take advantage of the random allocation of judges of varying severity in Sweden and find that parent incarceration leads to an increase in teen crime and a decrease in early life unemployment, while Bhuller et al. (2018), using a similar strategy for Norway, cannot reject the hypothesis that father incarceration does not affect performance in school. The literature on the effect of parental loss on schooling usually finds negative effects (Gertler et al., 2004; Evans and Miguel, 2007; Kovac, 2017), though Ainsworth and Filmer (2006) find that the difference in enrollment rates between orphans and non-orphans is small in many countries. Many find that, for education, a mother's death matters more than a father's death (Beegle et al., 2009, 2010). Even though Kovac (2017) studies paternal orphans, he underlines maternal stress in utero as a key mechanisms. This literature typically focuses on human capital or health outcomes in childhood and early adulthood, though human capital is far from the only channel through which the loss a father could affect a child's economic outcomes as an adult. Our historical setting allows us to study the long run consequences of parental loss on economic outcomes, including the generation of the grandchildren. Very few studies have used historical data for this purpose, one recent exception is Maloney and Smith (2018) who use the Utah Population Database. They do not find evidence of a penalty of father's death on the socioeconomic status of their children.

We also contribute to the literature on the economic consequences of conflicts. A first group of papers use geographical variation in conflict intensity to study the long-run consequences of fighting (Miguel and Roland, 2011; Chamarbagwala and Moran, 2011; Galdo, 2013; Domingues and Barre, 2013; Serneels and Verpoorten, 2015). Our paper does not rely on geographical variation in war intensity. In fact, Civil War battles occurred far from where Union soldiers enlisted. This allows us to focus on the long run consequences of conflict through the channel of parental loss. A more limited number of works study the effect of participation in wars for the soldiers, and find large negative income and health effects (Angrist, 1990, 1998; Angrist and Krueger, 1994; Imbens and van der Klaauw, 1995). Costa and Kahn (2010) find that Union Army veterans exposed to greater wartime stress have higher mortality rates later in life. We contribute to this literature by considering the long run consequences of participation in wars for the children of the soldiers.

We also add to the literature on the economic history of the Civil War and the Union Army. A random sample of Union Army regiments based on Fogel (2000) has been used to estimate the income effect of the Union Army pension program (Costa, 1995), and its effect on living arrangements (Costa, 1997), the impact of combat unit homogeneity on desertion (Costa and Kahn, 2003), the short and long run impact of diversity for black Union soldiers (Costa and Kahn, 2006), and the importance of social networks in survival in POW camps (Costa and Kahn, 2007). Since most soldiers in the Union Army were not fathers and because we want to be able to explore heterogeneity, our study relies on a new and much larger data set of linked Union Army soldiers using newly digitized information from all of the 2.2 million Union Army soldiers.

2.2 Historical Background

The U.S. Civil War (1861–65) was not only a defining moment for the United States, it was also the deadliest conflict in U.S. military history, with over 650,000 soldiers losing their lives. This was a substantial shock to a population of only 31 million according to the 1860 Census.

While the war was primarily fought over the abolition of slavery and for the preservation of the country's unity, several other political and economic factors played a role. We refer the reader to the work of historians for a comprehensive review of the history of the American Civil War (McPherson, 1988, for example). Instead this historical background section focuses primarily on the military and institutional setting which will help in framing the empirical analysis. We also discuss Union Army pensions and the extent to which Civil War veterans, widows, and orphans were compensated.

2.2.1 The Union Army: Recruitment and Organization

The Civil War started officially with the Confederate attack on Fort Sumter on April 12, 1861. The regular Army at the time was small to the point of insignificance with 16,000 personnel. The Union Army was instead raised after Lincoln's call for 75,000 volunteers in the Summer of 1861. With 600,000 initial enlistments, this call was exceeded by a wide margin (Chambers, 1987). Volunteer regiments were raised and organized by the individual states with little centralized intervention from the government. Participation was high. Of those born between 1838-45, up to 98% were examined for service and up to 81% of these cohorts ended up serving in the war (Costa and Kahn, 2008). The only time when participation rates were higher was during World War II. Ultimately, 2.2 million soldiers would serve in the Union and, even though the regular Army also increased in size, 92% of all soldiers were volunteers. The draft lottery, which was introduced in 1863, was largely ineffective and only raised 10,000 additional soldiers. The main purpose of the draft was to act as potential threat that could be expanded if not enough volunteers could be found (Chambers, 1987).

The structure of the Union Army closely resembled those of modern Armies. The organizational structure of the infantry is shown in figure 2.B.3. The most important unit in terms of both recruitment and fighting was the regiment. A typical infantry regiment had 1,000 soldiers and was composed of ten companies of a 100 men each. A Colonel would lead the regiment, and companies were commanded by a captain and two lieutenants. Military leaders were often not trained soldiers themselves but prominent men from the community in which a regiment was raised, such as politicians, factory owners, or other prominent figures, and "no company had the ability to pick the best officers and soldiers" (Costa and Kahn, 2008, p. 57). This is true even for the higher unit commanders as half of the Union generals were military amateurs rather than professionally trained soldiers (Chambers, 1987).

Companies were not replenished during the war even if a unit had to fight with less than half of its original strength. It rarely happened that in extreme cases the remaining soldiers of a company were transferred to other units (Costa and Kahn, 2008). We do observe such cases in our data, which happen for less than 6% of all records for which we also observe the new unit. The vast majority of soldiers remained in the units in which they originally enlisted and therefore we can accurately allocate the battles and casualties that were experienced by these units.

The typical contract length for a soldier was 3 years. 61% of the regiments we observe were recruiting on the basis of 3-year contracts. The next common contract lengths were 1- and 2-year contracts, which led to increased pressure on recruiters in 1863. This led to the passing of the Enrollment Act of 1863 which provided the basis for introducing a national draft in case not enough volunteers could be found. Unlike the South, where a very efficient draft had been established early in the war, the North mainly sought to promote volunteering by the threat of establishing a national draft and with generous enlistment bounties. In general, the early recruits of 1861 and 1862 were of higher quality than later enlistees. The early recruits tended to be positively selected as they were on average taller, richer, less likely to be low-skilled laborers, or married, and they were also more likely to be born in the U.S., Germany, or Ireland (Costa and Kahn, 2008).

The massive organizational scale as well as the size of the conflict generated unexpected problems. In the beginning, politicians of both sides estimated a total duration of no more than six months for the war (Chambers, 1987). Likewise the first battles were unexpectedly bloody as unusually large forces clashed and where old-style Napoleonic military tactics of closed-rank formations were confronted with the much increase range and fire power of modern weaponry (McPherson, 1988). In addition, the size of the military posed all sorts of logistic issues. Experience with field hygiene and medical treatment on such a scale was lacking. A little under half of all deaths were therefore due to disease and illness, closely followed by battle deaths.

2.3 Data Sources and Record Linking

The main data sources used here are the full-count U.S. Census files for 1860, 1880 and 1900, and military records from the Union Army. The Census was provided by IPUMS while the military information was digitized from various reports of each state's Adjutant General.[4] The Adjutant Generals' reports were compiled after the war to keep a record of veterans and deceased soldiers for accounting purposes, especially for the payment of bounties and pensions.[5]

Even though similar data exists for the Confederacy, the Union records are of much

[4]A full list of the different Adjutant General's reports is provided in the appendix in table 2.A.1. We are grateful to Christian Dippel and Stephan Heblich for joint efforts in digitizing and collecting the military data.

[5]An overview of the pension system is provided in appendix C.

higher quality and completeness. This motivated our focus on the Union. Quoting the Adjutant General of Massachusetts in his final report (1866): "*[M]ost of the regiments and batteries are perfect, every man accounted for; of the whole number there are but 1,205 who are not accounted for*" (p. 121). These unaccounted soldiers make up 1.1% of the overall number of enlisted men from Massachusetts which totaled 106,330. The total number of 2.9 million military records covers almost all of the 2.2 million Union soldiers. The number of records is larger than the number of soldiers due to re-enlistments.[6]

Information on each individual soldier includes their full name, enlistment and discharge date, military rank at enlistment and discharge, regiment and company, duration and terms of enlistment (commissioned, drafted, volunteered), and state of enlistment. We also have information on soldiers who died during the war and the reason of death (of disease, in battle, of wounds or injuries, in a Prisoner of War camp), and on those who survived but were severely wounded or injured, or disabled.

Other information is less systematically available across states, like place of residence, place of enlistment, and age at enlistment. Taking advantage of the fact that recruitment in the Union Army was very local, we infer the county of residence from the county of enlistment when the county of enlistment is missing. An example for the records of the 22nd Massachusetts volunteer infantry regiment is shown in figure 2.4.

From the Adjutant Generals' reports we also coded aggregate regiment information. For each regiment we know its unit type (Infantry, Cavalry, Artillery, regular Army, U.S. Colored Troops, Navy, Marines), battles fought, and its composition. This includes the total number of soldiers, number of fighting soldiers, share of enlisted, drafted, and transferred-in soldiers, the percentage of those who died, died of disease, died in POW camps, deserted, transferred out, as well as the percentage of discharged and mustered out soldiers. A minor share of soldiers died from accidents, criminal acts, or suicide. Summary statistics for the individual and regimental variables are reported in table 2.1 and table 2.2, respectively.

The data provides us with the same level of detail which is available to Costa and Kahn (2003, 2007) but at a significantly larger scale. Instead of their random sample of regiments based on Fogel (2000), we can observe the entirety of Union units and their characteristics, and almost the entire universe of Union soldiers. The main reason as to why information on soldiers could not be recovered is the unprecedented scale of the war. Not everyone could receive a proper burial under those circumstances. Sometimes individuals could not be identified

[6]We try to identify duplicate soldiers (who reenlisted) based on their first and last name, date of enlistment and age, but we manage to reduce the number of records by about 200,000 only. Soldiers who appear several times in the military records will likely not be linked to the Census because resolving linking ambiguities for them will be difficult.

anymore due to the severity of the injuries inflicted by the new weaponry available or because of the weather conditions.

2.3.1 Linking Census and Military Records

Linking the military records to the 1860 Census allows us to identify fathers who would eventually fight in the Civil War, as well as their children who we then track through later Census years. This is because members of the same household have a common identifier. Tracking children whose fathers fought, comparing those who lost their father to those who did not, limits the problem of selection into the Union Army because more wealthy individuals tended to more frequently opt out of service.

To build an inter-generational data set on the family members of the Civil War soldiers, we proceed in the following way: 1) we start by linking the Union Army military records with the U.S. population Census of 1860; 2) we then link men younger than 20 in 1860 to the 1880 Census; 3) our final data set consists of children observed in 1860 and linked to the 1880 Census who had a father linked to the Union Army Records. For further analysis, we also link the children of Union Army soldiers to the 1900 Census. When we observe the children of Union Army soldiers in 1880, we also observe the grandchildren, if any, at that point which we track into the 1900 Census to obtain a sample of grandchildren. This allows us to investigate the intergenerational effects of a soldier's death or disability. Figure 2.1 provides a schematic of this linking procedure.

Several algorithmic methods for linking historical records exist. Bailey, Cole, Henderson and Massey (2017) compare the performance of several of these methods with respect to the percentage of links generated and the associated type-I error (percentage of wrong links). We can think of the choice of these methods as presenting a trade-off between statistical power and accuracy. More restrictive methods produce more accurate, but fewer matches. In this paper, we use the algorithm of Ferrie (1996), also used by Abramitzky, Boustan and Eriksson (2012) and Abramitzky, Boustan and Eriksson (2014): we link individuals exactly on first name, last name, and state of residence or state of birth.[7] We then keep links that have an absolute birth year difference of < 5 years. The threshold can be tightened but there is significant age-heaping, i.e. individuals (in)voluntarily rounding their age, which would lead us to miss many potential links. In case of multiple links, we keep the link with the smallest age difference. If this does not resolve ties, we decide we cannot link the record.

Linking Union Army record is complicated by the fact that they do not systematically

[7] State of residence is used for linking Union Army records to the 1860 Census. We use state of birth for linking observations from Census to Census because residence might have changed over time.

record information on birth year (missing in 59% of cases). However, they sometimes record county of residence/enlistment (missing in 24% of cases). We therefore adapt Ferrie's algorithm to treat multiple links in the following way: if age is never missing for any of the potential links, we keep the links with the smallest age difference; if county of residence is never missing for any of the potential links, we keep the links with the smallest county distance; finally, even if county of residence is missing for some of the links, we keep links with a county distance of zero (links with an exact match on county of residence). Any unresolved tie results in the exclusion of the record.

Bailey et al. (2017) show that Ferrie-type algorithms perform well in terms of minimizing the type I error when we restrict potential links to rare names. Unfortunately, such a restriction considerably reduces our final sample size and the associated statistical power. We however show the robustness of our main results to Ferrie-type linking with an uncommon name restriction, considering only individuals whose combination of first and last name appears less the 10 times in the entire 1860 Census.

2.3.2 Final Data Set

Linking military records to the 1860 Census produces 545,022 links (24.8% linkage rate). We have information on survival, injuries and disability for 494,048 of them. 14.8% died during the war, 9.52% returned home with a major disability, and 6% had the mention of an injury on their military records without being classified as disabled.

Figure 2.2 shows the geographical distribution (by 1860 county) of the soldiers in our data set. This corresponds to the geography of enlistment in the Union Army: the vast majority of Union Army soldiers came from the Northeastern states and the Great Lakes region. Some recruits were residing in the Southern states and came to enlist in the North, but they are a small minority. Finally, there are few recruits in the newly settled territories of the West and the Pacific coast. It is worth noting that the Civil War was fought predominantly in the Southern and border states (appendix figure 2.B.2), which allows us to estimate the effect of parental loss separately from the other short and long-term consequences of combat such as . Figure 2.3 displays the geographical distribution of death rates. There is no obvious geographical pattern in the distribution of death rates across counties — counties were few soldiers enlisted display a larger variation in death rates, as expected.

Linking men younger than 20 in the 1860 Census to the 1880 Census produces 1,164,706 links (31.7% linkage rate). Combining the linked soldier file to the linked Censuses file, we obtain a sample of 41,831 men observed in 1880 and 1860 whose father fought in the War. 13.8% lost their father during the war, and 12.4% had a father who came back from the war with a severe disability.[8]

Because Civil War soldiers were young men (the average age at enlistment was 25, see table 2.1), we obtain a much larger sample of individuals observed in 1860 and 1880 who had a brother in the Union Army: 86,096 individuals, 15.9% of which lost a brother during the war. 4,103 individuals had both a father and at least one brother in the Union Army: 14% of them lost their father, 16.4% lost a brother, 2.8 % lost both.

Finally, we also link the grandchildren of Civil War soldiers, observed in the 1880 Census, to the 1900 Census to investigate very long-run inter-generational effects. Because we are combining three linkage procedures, we end up with a small data set of 3,130 grandchildren, 14.86% of which had a grandfather who died during the war.

To investigate the long-run intergenerational effects of the war, we use the socio-economic variables available in the U.S. Censuses of 1860, 1880 and 1900. Variables in the 1860 Census are used as baseline controls, while variables in the 1880 and 1900 Censuses are used as outcomes. Our main variable of interest is occupation, given as a string variable in the Censuses. From this string variable, we follow the classification of IPUMS USA to build occupation categorical variables and an occupational score. The occupational outcomes are: 1) a dummy for having a high skilled occupation (professional, technical, manager, officials, and proprietors), 2) a dummy for having a semi-skilled occupation (sales, craftsmen, operatives), 3) a dummy for having a low-skilled occupation (service workers, laborers, including farm laborers), 4) a dummy for being a farmer. The occupational score is the IPUMS 1950 occupational score: it gives an individual the median 1950 income of his occupation.[9] We normalize it to have mean zero and unit variance. In the generation of fathers, farmer is by far the most common category: 41.4% of the soldiers-fathers in our database were farmers in 1860, 18% had a low skilled occupation. 25.3 % of their sons were farmers, 32.2% were low-skilled workers.

In 1860 only, the Census provides information on real estate and personal wealth measured in dollars. Because wealth tends to be log-normally distributed and because a large number of soldiers (69.4%) had no wealth, we take the inverse hyperbolic sine transform of wealth (Friedline, Masa and Chowa, 2015).[10] Literacy is given in 1860 — it is already very high, as 96.5% of soldiers can read. We can also measure migration in 1880 (whether the

[8]We actually obtain a sample of 47,183, but survival information is missing for 5,352.

[9]Of course, wages in 1950 and in 1860 were different, but this is a straightforward way of obtaining a continuous ordered measure from string occupation. The 1950 Census contains information about wages, but not the 1860, 1880 or 1900 Censuses. Building an occupational score for 1860 requires the collection of detailed wage data for 1860.

[10]The inverse hyperbolic sine transform of y is $\log(y + \sqrt{y^2 + 1})$. Except for small values of y, the inverse hyperbolic sine is approximately equal to $\log(2y)$, so that it can be interpreted as a log variable.

individual's county of residence changed between 1860 and 1880), as well as marital status.

2.4 Empirical Strategy

2.4.1 Selection into the army and selection into death

We are interested in the effect of losing a father, or having your father permanently disabled, on later life outcomes. if military deaths and injuries were happening at random, it would be as simple as estimating by OLS the following equation:

$$y_{ijt} = \alpha + \beta_1 died_j + \beta_2 disabled_j + \varepsilon_{ijt} \tag{2.1}$$

where y_{ijt} is a socioeconomic outcome measured in Census year t for individual i, whose father is j. $died_j$ is an indicator equal to 1 if j (the father of i) died in the war, $disabled_j$ is an indicator equal to 1 if j was disabled.

The Union Army was primarily composed of volunteers, most of whom enrolled in the months following the capture of Fort Sumter in April 1861. After 1862, as recruiting men was becoming harder, states and towns began offering bounties for enlistment. The Enrollment Act of 1863 created a conscription system, but overall draftees were relatively rare (Chambers, 1987; Costa and Kahn, 2007). In our sample of soldier, the share of enlisted is 94% (table 2.1). To address the problem of selection into the army, we restrict the analysis to children of men who fought in the Union Army, comparing the children of those who fought and died to the children of those who fought and returned. This means our effect is estimated on a selected sample of families, that might not be representative of the general population. That said, a large number of military age men fought in the war — 37% of men aged 15-44 in 1860, and the Union Army was overall representative of the northern population of military age (Costa and Kahn, 2007).

Given participation in the conflict, there is of course no reason to think that the probability of dying was as good as random. The majority of deaths during the American Civil War were due to diseases like dysentery, typhoid fever and pneumonia that spread in military camps with poor sanitation. In our database of Union soldiers, only 36% of those who died died in combat. Soldiers from lower socioeconomic groups are likely to be more at risk of dying during an epidemic, if they were already in worse health when they enlisted. Battle death and injuries could also be correlated with a soldier's wealth, negatively if poorer units are more easily sent to the front-line, positively if we think richer soldiers who are in better health are more able to fight and take risks.

Our database of Union Army soldiers linked to the 1860 Census allows us to assess the socioeconomic and military correlates of the probability to die, of disease and in combat, to be seriously wounded (disabled), and to be injured. In table 2.3, we regress the probability of each event on a vector of socioeconomic variables in the 1860 Census, and military variables, as well as state fixed effects. Death in combat, death of disease, disabled and injured are mutually exclusive, so that disabled means that a soldier's wounds resulted in disability, but that he survived (at least until the moment he was discharged). Injured means a soldier was injured and survived, but that there was no mention in the army records of his injuries resulting in disability or death.

Table 2.3 shows that the probability to die or be wounded is correlated with socioeconomic characteristics of soldiers, but the explained variance is low. Death of disease is strongly related to socioeconomic characteristics. Being foreign born decreases the probability of dying of disease by 2.4 percentage points. Illiterate workers are 0.5 percentage points more likely to die of disease. High skilled and semi-skilled are 2.6 and 2.4 percentage points less likely to die of disease than farmers, the reference group. Wealth is, surprisingly, positively correlated with the probability to die of disease (conditional on the other variables in the regression), but the coefficient is very small at 0.0003.

Socioeconomic characteristics seem to matter less for the probability to die in combat, be disabled or injured, but they matter. To give an idea of the extent to which socioeconomic characteristics matter without commenting each individual coefficient, let's consider the difference between a farmer with median wealth (no wealth) and a high-skilled worker with wealth at the 90th percentile ($ 1,200). According to our estimates, the farmer with no wealth is, all other things equal, 2.42 percentage points more likely to die of disease, 1.1 percentage points more likely to die in combat, 0.7 percentage points more likely to be disabled and 0.85 percentage points more likely to be injured without being disabled.[11]

Military variables explain more of the variation in death, disability and injuries. Mechanically, enlisting later in the war decreases the probability to be hurt as the amount of days spent fighting is reduced. For example, each 100 days delay in the date of enlistment decreases the probability to die in combat by 0.5 percentage points. Military rank is also an important determinant, as private soldiers (representing 83.8% of the total) are more likely to die of disease and in combat, and to be disabled and injured. Finally, the type of regiments matters. Being part of a cavalry, artillery or special regiment rather than an infantry regiment decreases the probability to die in combat or be injured. Effects are a little bit harder to interpret for death of

[11] Computations are the following: $-7.8 \times (0.0003) + 0.0265 = 0.02416$, $-7.8 \times (-0.0006) + 0.0063 = 0.01098$, $-7.8 \times (0.0004) + 0.0097 = 0.00658$, $-7.8 \times (-0.0002) + 0.0069 = 0.00846$.

disease and disability. It is worth keeping in mind that the majority of soldiers (76.6%) were part of an infantry regiment.

Finally, the adjusted R-squared of these four regressions is low (between 0.016 and 0.050), which indicates that socioeconomic and military variables explain only a small part of the overall variation in death, disability and injury rates. The remaining variation in death and disability rates was likely largely explained by the battles a military unit took part in, the spreading of epidemics in military camps, and the positioning of units on the battlefield during a given battle.

Appendix B shows that Union regiments that were on the front line during important battles were not different in their socioeconomic composition. We collected and digitized 128 battle maps from the Civil War Preservation Trust and we show that there is no correlation between a Union regiment's socioeconomic characteristics and its distance from the nearest enemy unit at various stages of the battle.

2.4.2 First identification strategy: OLS

Our first identification strategy consists in regressing 1880 outcomes on indicators for father's death and disability controlling for a rich set of baseline socioeconomic controls measured in 1860, and military variables. We estimate by OLS the following equation

$$y_{i,j,1880} = \alpha + \beta_1 died_j + \beta_2 disabled_j + x'_{j,1860}\theta_1 + z'_{i,1860}\theta_2 + s_{ij} + \varepsilon_{i,j} \tag{2.2}$$

where $x_{j,1860}$ is a vector of baseline controls for father j measured in 1860: age and age squared, a nonwhite dummy, occupational score, occupational skill dummies, the inverse hyperbolic sine of wealth, literacy, and being foreign born. $x_{j,1860}$ also contains military variables for father j: enlistment date, enlistment rank fixed effects, and enlistment regiment type fixed effects.[12] $z'_{i,1860}$ is a vector of 1860 baseline controls for the son i: age, age squared and literacy. s_{ij} is a vector of 1860 state of residence fixed effects. Standard errors are clustered by father (we sometimes observe two brothers in 1880).

2.4.3 Second identification strategy: battle IV

In a second identification strategy, we instrument for father's death with an indicator for whether the father's regiment took part in one of the top 10 bloodiest battles of the Civil War while the father was serving in the regiment. The reasoning is that which regiments get send to certain battles depends mainly on military strategy that does not consider the later life out-

[12] The enlistment ranks are. The regiment types are

comes of soldiers' children.[13] The list of the major land battles in the war come from Selcer (2006) and is reproduced in table 2.4; we order them with respect to total casualties.[14] Appendix figure 2.B.2 shows the location of these battles. 50.46% of soldiers in our data set took part in at least one of the top 10 bloodiest battles. We consider the following first stage equation:

$$died_{ijr} = \gamma + \delta deadlybattle_r + x_j'\theta_1 + z_i'\theta_2 + s_{ij} \tag{2.3}$$

where $died_{ijr}$ is a dummy equal to 1 if j, the father of i, was disabled or died during the war; $deadlybattle_r$ is a dummy equal to 1 if the father's enlistment regiment r took part in one of the top 10 bloodiest battles of the war while the father served in the regiment; x_j' and z_i' are defined as in equation (2.2); s_{ij} is a vector of state of residence fixed effects. Standard errors are clustered by regiment, which is the unit of treatment in this case.

Table 2.5 presents the result of the first stage for the battle IV on the sample of fathers who fought. Column (1) presents the raw correlation between having taken part in one of the top 10 bloodiest battles of the war and the probability of dying: soldiers whose regiment took part in one of these battles are 8.8 percentage points more likely to have died. Controlling for regiment type fixed effects in column (2) does not affect the first stage coefficient, but controlling for a polynomial in enlistment date reduces it to 7.2 percentage points. This is expected, as date of enlistment is one of the major determinants of which battles a soldier took part in. Additional controls hardly affect the coefficient, which reassures us that our instrument is not picking up the effect of selection of certain types of soldiers. The coefficient is 0.068 when we add state fixed effects (column 5), it is not at all affected when we add the full set of father baseline controls (column 6), nor when we add 1860 county of residence fixed effects (column 7).

2.4.4 The impact of linkage errors

Linking historical records in the absence of individual identifier inevitably produces some linkage errors. To assess their impact on our estimates, consider the following setting. Denoting ν as the share of false positives (children who did not lose their father but who were marked as fatherless), and η as the share of false negatives (children who lost their father but were not

[13]We show in the appendix using digitized battle maps that the socioeconomic composition of regiments is unrelated to battle field location and distance to the nearest enemy, i.e. it is not the case that the poorest regiments were put in the front lines.

[14]The top 10 bloodiest battles are: Gettysburg, Wilderness, Spotsylvania, Chancellorsville, Chickamauga, Seven Days, 2nd Bull Run, Shiloh, Stones River, and Cold Harbor.

marked as fatherless), the biases of OLS and IV are:

$$\text{plim}\widehat{\beta}_{OLS} = \beta(1 - \nu - \eta) \tag{2.4}$$

$$\text{plim}\widehat{\beta}_{IV} = \beta\frac{1}{1 - \nu - \eta} \tag{2.5}$$

which, for $\nu + \eta < 1$ leads to an attenuation bias for OLS and an inflation bias for IV.[15] Bailey et al. (2017) find that the rate of linking error when using Ferrie's algorithm with common names is 30%. In the extreme case where a linking error always reverses the treatment status of an individual, we have $\nu + \eta = 0.3$. In this case, the attenuation bias of OLS is 70% and the inflation bias of IV is 143%. In the absence of endogeneity problems affecting OLS, the OLS and IV estimates can be used to bound the true effect.

2.5 Results

Table 2.6 presents the results of the OLS with baseline father controls. Having lost a father in the Civil War decreases the 1880 occupational score by 0.05 standard deviations (column 1). It decreases the probability to have a high-skilled occupation by 0.7 percentage points and the probability to have a semi-skilled occupation by 1.8 percentage points (columns 2 and 3). The probability to be low-skilled increases by 1 percentage points and the probability to be a farmer increases by 1.3 percentage points (column 5). Geographical mobility might be one of the channels explaining lower social mobility, but we do not estimate any effect of the loss of a father on the probability to have changed county between 1860 and 1880 (column 6). Finally, having lost a father during the war increases the probability to be married in 1880 by 1.6 percentage points. The average age of Civil War soldier sons in 1880 is 26, so that this increase in marriage does not reflect an overall increase in the probability of ever marrying, but rather a decrease in age at marriage for sons who lost their father during the war. This might reflect the fact they left school at a younger age to work and support the family.

The effect of having a disabled father is always lower than the effect of losing a father, but having a disabled father had an effect on later-life socioeconomic outcomes, which is indicative of income being an important channel in explaining our results. While disabled fathers have decreased earning potential, they can still transmit their human and cultural capital. Having a disabled father decreases the probability to be semi-skilled by 1.2 percentage points and increases the probability to be low skilled by 1.5 percentage points. We can not quite reject the

[15] The full derviations can be found in appendix D.

possibility that losing a father is worse than having a disabled father in terms of occupational scores (p-value of 0.13, column 1), but we can reject it in the case of the probability of being a farmer and the probability of ever marrying. This points toward the transmission of the family farm as an important factor in understanding the effect of the loss of a father in childhood.

In table 2.7, we add the effect of having an injured father (but not disabled). Being injured can be thought of as being determined by the same kind of factors as death or disability. On the other hand, being injured in war is not a banal event, and might have very long term physical and psychological consequences. One concern is that the effect of father loss and disability is an underestimate because the control group was actually also "treated" by the war. Adding whether the father was injured hardly affects the effect of father loss and disability. The variable is never statistically significant, though we do not have enough power to reject the equality of father loss and father injury, except in the case of the probability of being high-skilled (p-value=0.08) and marginally in the case of the probability of being low-skilled (p-value=0.14).

We obtain similar point estimates of the effect of losing a father when using the sample linked with Ferrie's algorithm using uncommon names only (appendix table 2.F.3), but the sample size is reduced to 7,418 sons, and the loss in statistical power results in coefficients not always being statistically significant.[16]

Table 2.8 reports the effects of losing a father estimated using the battle IV instrument. The effects go in the same direction as with OLS, except that they are larger and noisier. Losing a father decreases the 1880 occupational score by 28.5% of a standard deviation and reduces the probability to be semi-skilled by 18 percentage points. Surprisingly, the probability to have migrated between 1860 and 1880 increases by 36 percentage points with the loss of a father, and the probability to have ever married is not affected. The large effects obtained when instrumenting father death by regiment participation in one of the top bloodiest battles of the war might be explained by the fact that the instrument does not only affect the probability to die, but also the probability to be disabled. In appendix table 2.F.2, we control for father disability (without instrumenting), and we find that the effects are reduced, but remain large, at -24.8% of a standard deviation in occupational score and -14.5 percentage points in the probability of being semi-skilled. Another possibility is that, because of the local nature of regiments, the participation of a father's regiment in one of the bloodiest battles of the war also affects sons through the loss of other members of the larger familial network or due to the adverse effect on the local economy if many men from the same town died.

[16]Ferrie's algorithm with uncommon names restricts the set of potential links to individuals with rare first name last name combinations, appearing less than 10 times in the entire United States in 1860.

Another potential explanation for the difference in the OLS and IV estimates are errors resulting from wrong links. We derive the bias terms of the OLS and IV estimators in appendix D, when linkage errors reverse the treatment status of individuals. In the extreme case of linkage errors reversing the treatment status of individuals, and assuming a 30% share of wrong links, the type-I error rate with Ferrie common names in Bailey et al. (2017), we find that the OLS estimate is attenuated by 70% and the IV estimate inflated by 143%. In the absence of other endogeneity problems affecting the OLS, a linkage error rate of 30% would therefore explain 45% of the difference in the OLS and IV coefficients for the occupational income score.[17] The remaining difference must then be due to other endogeneity problems affecting OLS and solved by IV, to the fact that the instrument affects sons through other channels than the death of the father (for example mortality in the larger family network), or to heterogeneous treatment effects — IV estimates a Local Average Treatment Effect, and the effect of father death might be stronger for the takers of the instrument, i.e. children whose fathers died because they were assigned to one of the top 10 bloodiest battles and died there and who would have survived if they had not been assigned to one of these battles.

2.6 Heterogeneity and discussion of results

Does the loss of a father negatively affects occupational score in 1880 because the sons of high-skilled and semi-skilled father become farmers and low-skilled workers (increased downward mobility), or because the sons of farmers and low-skilled workers are less likely to become semi-skilled and high-skilled (increased downward mobility)? In table 2.9, we interact father death and disability with the occupational category of the father. The negative effect of father loss on occupation seem to be driven mainly by the downward mobility of the sons of semi-skilled fathers. Sons who lost their semi-skilled fathers are 5.4 percentage points less likely to be semi-skilled (significant at the 1% level) and 4 percentage points more likely to be low-skilled (significant at the 1% level). Sons whose semi-skilled father returned home disabled are less likely to be high-skilled and semiskilled, and 2.7 percentage points more likely to be low-skilled.

Next, we consider whether wealth is a mitigating factor. In table 2.10, we interact father death and disability with dummies indicating whether the father was in the top or bottom half of the wealth distribution in 1860. The effect of father death is clearly driven by the bottom half of the wealth distribution. Losing a father decreases the occupational score by 7.2% of a standard deviation in the bottom half (significant at the 1% level), and decreases the occupational score

[17]This comes from computing $1 - \frac{(1-0.3)\widehat{\beta}_{\text{IV}} - \frac{1}{1-0.3}\widehat{\beta}_{\text{OLS}}}{\widehat{\beta}_{\text{IV}} - \widehat{\beta}_{\text{OLS}}}$.

by 2.4% of a standard deviation in the top half (not statistically significant). The two effects are statistically different from each other (p-value=0.08). Wealth also seems to be a mitigating factor in the case of father disability: while having a disabled father decreases the occupational score by 4% of a standard deviation in the bottom half, it decreases it by 0.1% in the top half — though because the effect of father disability is smaller, we cannot reject that the two coefficients are equal (p-value of 0.20).

In appendix table 2.F.4, we investigate heterogeneity with respect to child's age when his father enlistment. Appendix figure 2.F.1 displays the distribution of age at father enlistment: median age at father enlistment is 7, and a sizable number of sons saw their father leave for the war when they were well in their teenage years. We expect children who lost their father at an early age to be the most affected by the loss of a father, whether we consider the human capital transmission channel or the income channel. At the same time, younger children had younger mothers for whom it was easier to remarry if they lost their husband. In table 2.F.4, we interact father death and disability with indicators indicating whether the son was younger than 7 or older than 8 at the time of father enlistment. We find that effect are quite homogeneous along the age dimension.

What was the effect of the loss of other family members, in particular brothers? To answer this question, we focus on the sample of 3,456 individuals who had both a father and a brother in the Union Army.[18] 497 lost their father, 477 had a disabled father, 517 lost a brother and 220 had a disabled brother.[19] Table 2.F.1 shows that, as expected, the loss of a father has stronger consequences than the loss of a brother. On this sample, losing a father decreases the occupational score by 15.5% of a standard deviation while losing a brother had no effect on the occupational score. Losing a brother, though, decreases the probability to be high-skilled by 2.8 percentage points and increases the probability to be a farmer by 3.9 percentage point. One explanation is that individuals who lost a brother are more likely to inherit the family farm, and as a consequence less likely to pursue their education.

Were the negative effects of the loss of a father transmitted to the next generation? To answer this question, we use our sample of 3,130 grandchildren of union army soldiers observed as adults in the 1900 Census. 14.9% of had a paternal grandfather who died during the war and 13.61% of them had a paternal grandfather disabled during the war. This is a much

[18] We could also focus only on the sample who had a brother in the Union Army, but the problem is that, because families often enlisted together, having a brother who died or was disabled is likely correlated with having a father who died or was disabled. If we manage to link a brother, but not a father, in the Union Army records, it does not mean the father did not fight, but simply that we could not link the father.

[19] 93 lost both a brother and a father, 44 lost their father and had a disabled brother, 84 lost a brother and had a disabled father, 43 had a disabled father and a disabled brother. We do not have enough statistical power to consider all interaction.

smaller sample size, and hence statistical power is greatly reduced, but we do estimate negative effects of father loss and disability being transmitted to the next generation (table 2.11). Having a grandfather who died in the Civil War reduced occupational score by 8% of a standard deviation (not statistically significant). It reduces the probability to have a semi-skilled occupation by 4.2 percentage points (statistically significant at the 10% level). Effects size are quite similar for having a disabled grandfather (and of course, sample size does not allow to carry meaningful equality of coefficients tests). Having a disabled grandfather decreases the occupational score by 10.6% of a standard deviation.

2.7 Conclusion

Using the US Civil War as a shock to family structure, we estimate negative effects of the loss of a father on long-run socioeconomic outcomes. Losing a father decreases occupational score in 1880 by 5% of a standard deviation. Several elements indicate that the income channel is an important one: first, we estimate smaller, but negative effects of father disability. Second, we show that father wealth in 1860 is an important mitigating factor: the effect of father loss is 7% of a standard deviation of occupational score in the bottom half of the wealth distribution versus 2% in the top half. Third, the death of a brother does not matter as much as the death of a father. On a very small sample, we estimate that these negative effects are transmitted to the generation of the grandchildren observed in 1900.

Future research will investigate further the mechanisms driving our result in two directions: first by exploring heterogeneity with respect to family size and family structure in 1860, second by exploring the effects of Civil War mortality in the larger family and community, as well as its labor market effects.

2.8 Tables

Table 2.1: Military Records Summary Statistics

	Obs.	Mean	St. Dev.	Min.	Max.
Age at enlistment	1,116,563	25.422	7.364	11	70
Date of enlistment	2,554,359	Jan 15 1863		Jun 10, 1801	Jul 22, 1865
Enlisted	2,660,196	0.940	0.237	0	1
Commissioned	2,660,196	0.030	0.170	0	1
Drafted	2,660,196	0.016	0.124	0	1
Substitute	2,660,196	0.014	0.119	0	1
Died	2,150,953	0.126	0.332	0	1
Died (battle)	2,150,953	0.045	0.208	0	1
Disabled	2,150,953	0.082	0.275	0	1
Injured	2,150,953	0.049	0.215	0	1
Private	2,700,600	0.841	0.365	0	1

Note: Summary statistics for the 2.7 millions Union Army Military Records. The total number of records remains larger to the total number of soldiers in the Union Army because of re-enlistments. The date of enlistment for some soldiers is before 1861 because they belong to the regular Army. Substitutes are soldiers who took the place of another drafted man for payment.

Table 2.2: Union Regiment Statistics

	Mean	St. Dev.	Min	Max
Unit Types:				
Cavalry	0.127	0.333	0	1
Infantry	0.684	0.465	0	1
Heavy Artillery	0.018	0.135	0	1
Light Artillery	0.148	0.355	0	1
Sharp Shooters	0.008	0.086	0	1
Unit Characteristics:				
Total men	990.008	927.459	2	31,763
Fighting men	976.226	914.975	1	31,692
Enlisted (%)	95.316	12.027	0.110	100
Drafted (%)	1.654	5.620	0	70.500
Killed (%)	2.376	3.239	0	19.920
Died as POW (%)	0.508	1.382	0	28.370
Died of disease (%)	4.416	4.876	0	39.270
Disabled (%)	5.515	6.700	0	41.600
Deserted (%)	4.395	6.742	0	60.390

Note: Summary statistics for the 2,922 Union Army regiments. Statistics are final values at the end of the war.

Table 2.3: Correlates of Death, Disability, and Injury Probabilities

	Cause of Death				
	Any	Disease	Combat	Disabled	Injured
Socioeconomic variables					
Age	0.0002^{***}	0.0002^{***}	-0.0002^{***}	0.0027^{***}	-0.0009^{***}
	(0.0001)	(0.0000)	(0.0000)	(0.0001)	(0.0000)
Non-white	-0.0044	0.0000	-0.0036	-0.0010	-0.0014
	(0.0060)	(0.0035)	(0.0031)	(0.0049)	(0.0040)
Foreign born	-0.0218^{***}	-0.0165^{***}	0.0009	-0.0143^{***}	0.0054^{***}
	(0.0014)	(0.0009)	(0.0008)	(0.0012)	(0.0010)
Illiterate	0.0071^{**}	-0.0006	0.0008	-0.0060^{**}	0.0035^{*}
	(0.0029)	(0.0017)	(0.0016)	(0.0024)	(0.0019)
IHS wealth	-0.0003^{*}	0.0003^{***}	-0.0005^{***}	0.0004^{***}	-0.0002^{*}
	(0.0002)	(0.0001)	(0.0001)	(0.0002)	(0.0001)
High-skilled	-0.0327^{***}	-0.0169^{***}	-0.0068^{***}	-0.0097^{***}	-0.0069^{***}
	(0.0027)	(0.0017)	(0.0015)	(0.0022)	(0.0018)
Semi-skilled	-0.0279^{***}	-0.0149^{***}	-0.0036^{***}	-0.0021^{*}	-0.0004
	(0.0014)	(0.0009)	(0.0008)	(0.0011)	(0.0009)
Low-skilled	0.0081^{***}	0.0028^{***}	0.0035^{***}	-0.0032^{***}	0.0047^{***}
	(0.0013)	(0.0008)	(0.0008)	(0.0011)	(0.0009)
Military variables					
Enlistment date	-0.0094^{***}	-0.0026^{***}	-0.0047^{***}	-0.0119^{***}	-0.0048^{***}
(in 100 days)	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)
Private	0.0392^{***}	0.0234^{***}	0.0023^{***}	0.0137^{***}	0.0044^{***}
	(0.0015)	(0.0009)	(0.0009)	(0.0012)	(0.0010)
Professional	-0.0270^{***}	-0.0154^{***}	0.0097^{***}	-0.0610^{***}	0.0131^{***}
	(0.0031)	(0.0020)	(0.0018)	(0.0026)	(0.0021)
Drafted	0.0154^{***}	0.0144^{***}	0.0067^{***}	-0.0093^{***}	0.0110^{***}
	(0.0043)	(0.0027)	(0.0025)	(0.0035)	(0.0029)
Substitute	-0.0230^{***}	-0.0156^{***}	-0.0042	0.0032	-0.0012
	(0.0058)	(0.0035)	(0.0032)	(0.0047)	(0.0038)
Type of regiment					
Cavalry	-0.0146^{***}	-0.0002	-0.0187^{***}	-0.0033^{***}	-0.0327^{***}
	(0.0016)	(0.0010)	(0.0009)	(0.0013)	(0.0010)
Heavy Artillery	0.0079^{***}	0.0084^{***}	-0.0049^{***}	0.0058^{***}	-0.0094^{***}
	(0.0027)	(0.0016)	(0.0015)	(0.0022)	(0.0018)
Light Artillery	-0.0467^{***}	-0.0030^{*}	-0.0339^{***}	-0.0082^{***}	-0.0454^{***}
	(0.0030)	(0.0018)	(0.0017)	(0.0024)	(0.0020)
Special	-0.0521^{***}	-0.0121^{***}	-0.0194^{***}	0.0081^{**}	-0.0396^{***}
	(0.0044)	(0.0028)	(0.0025)	(0.0036)	(0.0029)
State F.E.	✓	✓	✓	✓	✓
Observations	491,506	536,943	536,943	491,506	491,506
Adj. R^2	.0226	.0221	.0157	.0499	.028

Note: Summary statistics for the 2.7 millions Union Army Military Records. The total number of records remains larger to the total number of soldiers in the Union Army because of re-enlistments. IHS wealth is the inverse hyperbolic sine (IHS) transform of a soldier's house value in the 1860 Census. Since many soldiers did not own their home, the IHS can deal with zeroes but preserves the properties of a log transformation. Skill definitions follow the 1950 occupational definition by the U.S. Census Bureau. Professional soldiers refer to those in the regular Army and Navy or who are commissioned officers. Substitutes are soldiers who replaced other drafted men in return for payment. Special units refer mainly to sharp shooters. Significance levels are denoted by * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 2.4: Information on the 10 Major Land Battles of the U.S. Civil War

Battle	Dates	Union strength	Killed	Wounded	Missing or POW
Shiloh	April 6-7, 1862	63,000	1,754	8,408	2,885
Seven Days	June 25 - July 1, 1862	114,691	1,734	8,062	6,053
2nd Bull Run	Aug. 28-30, 1862	77,000	1,747	8,452	4,263
Stones River	Dec. 31, 1862 - Jan. 2, 1863	43,400	1,677	7,543	3,686
Chancellorsville	April 30 - May 6, 1863	133,868	1,606	9,762	5,919
Gettysburg	July 1-3, 1863	104,256	3,155	14,529	5,365
Chickamauga	Sept. 18-20, 1863	60,000	1,657	9,756	4,757
Wilderness	May 5-7, 1864	124,232	3,469	16,000	3,383
Spotsylvania	May 8-21, 1864	110,000	2,725	13,416	2,258
Cold Harbor	May 31 - June 12, 1864	117,000	1,845	9,077	1,816

Note: The 20 most important land battles of the Civil War from Selcer (2006). Casualty information from various sources and authors' own computations. * marks joint casualty numbers for deaths, wounded, and missing combined. Some battles were known by different names in the Union and Confederacy. These are Shiloh (Battle of Pittsburg Landing) and the 2nd Bull Run (Second Battle of Manassas). The Seven Days Battle was a series of seven different battles fought over the course of seven days: Oak Grove, Beaver Dam Creek, Gaines' Mill, Garnett's and Golding's Farm, Savage's Station, Glendale and White Oak Swamp, Malvern Hill).

Table 2.5: First Stage of Fathers' Death Probability and the Top 10 Deadliest Battles

	Outcome: probability of dying						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Top 10 Battle	0.088***	0.089***	0.072***	0.070***	0.068***	0.068***	0.068***
	(0.005)	(0.005)	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)
Observations	41,828	41,536	41,370	41,364	41,363	41,253	41,093
F-stat	292.09	294.42	129.61	123.79	115.47	115.81	119.43
Reg. type F.E.		✓	✓	✓	✓	✓	✓
Enl. date poly			✓	✓	✓	✓	✓
Enl. rank F.E.				✓	✓	✓	✓
State F.E.					✓	✓	
Father controls						✓	✓
County F.E.							✓

Note: Regression of father's death probability on participation in one of the top 10 bloodiest battles (top 10 battle) of the Civil War. Regiment type fixed effects are for infantry, cavalry, artillery, and specialized units such as sharpshooters. Geographic fixed effects are for the residence of fathers. The enlistment date polynomial controls for time of exposure to the war, and enlistment rank fixed effects include dummies for the initial rank (85% of soldiers start as privates). Father controls include information from the 1860 Census on father's age, nationality, literacy, wealth (house value), and occupational skill group. Standard errors clustered by regiment in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 2.6: Effect of Dead or Disabled Fathers on Children's 1880 Socioeconomic Outcomes

	occupational score	high-skilled	semi-skilled	low-skilled	farmer	migrant	ever married
Father died	-0.049***	-0.007*	-0.018***	0.011	0.013**	-0.003	0.015**
	(0.014)	(0.004)	(0.006)	(0.007)	(0.006)	(0.008)	(0.006)
Father disabled	-0.021	-0.002	-0.012*	0.015**	-0.000	-0.000	-0.003
	(0.015)	(0.004)	(0.007)	(0.007)	(0.006)	(0.008)	(0.007)
Observations	40,968	40,968	40,968	40,968	40,968	41,021	39,969
Reg. type F.E.	✓	✓	✓	✓	✓	✓	✓
Enl. rank F.E.	✓	✓	✓	✓	✓	✓	✓
State F.E.	✓	✓	✓	✓	✓	✓	✓
Father controls	✓	✓	✓	✓	✓	✓	✓
Own controls	✓	✓	✓	✓	✓	✓	✓

Note: Regiment type fixed effects are for infantry, cavalry, artillery, and specialized units such as sharpshooters. Geographic fixed effects are for the residence of fathers. Enlistment rank fixed effects include dummies for the initial rank (85% of soldiers start as privates). Father controls include information from the 1860 Census on father's age, nationality, literacy, wealth (house value), and occupational skill group. A son's own controls include his age, age squared, and literacy in 1860. Standard errors clustered by family in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 2.7: OLS Effect of Dead, Disabled, or Injured Fathers

	occupational score	high-skilled	semi-skilled	low-skilled	farmer	migrant	ever married
Father died	-0.048***	-0.007*	-0.017***	0.011	0.012**	-0.003	0.014**
	(0.014)	(0.004)	(0.007)	(0.007)	(0.006)	(0.008)	(0.006)
Father disabled	-0.023	-0.002	-0.013*	0.014*	0.001	-0.001	-0.003
	(0.015)	(0.004)	(0.007)	(0.007)	(0.006)	(0.008)	(0.007)
Father injured	-0.019	0.005	-0.012	-0.007	0.011	-0.002	0.007
	(0.020)	(0.006)	(0.009)	(0.009)	(0.008)	(0.010)	(0.008)
Observations	40,939	40,939	40,939	40,939	40,939	40,992	39,940
Reg. type F.E.	✓	✓	✓	✓	✓	✓	✓
Enl. rank F.E.	✓	✓	✓	✓	✓	✓	✓
State F.E.	✓	✓	✓	✓	✓	✓	✓
Father controls	✓	✓	✓	✓	✓	✓	✓
Own controls	✓	✓	✓	✓	✓	✓	✓

Note: Regiment type fixed effects are for infantry, cavalry, artillery, and specialized units such as sharpshooters. Geographic fixed effects are for the residence of fathers. Enlistment rank fixed effects include dummies for the initial rank (85% of soldiers start as privates). Father controls include information from the 1860 Census on father's age, nationality, literacy, wealth (house value), and occupational skill group. A son's own controls include his age, age squared, and literacy in 1860. Standard errors clustered by family in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 2.8: Effect of Dead Fathers on Children's 1880 Socioeconomic Outcomes: Battle IV Results

	occupational score	high-skilled	semi-skilled	low-skilled	farmer	migrant	ever married
Father died	-0.285*	-0.036	-0.181**	0.089	0.078	0.363***	-0.004
	(0.172)	(0.050)	(0.084)	(0.085)	(0.067)	(0.101)	(0.071)
Observations	41,204	41,204	41,204	41,204	41,204	41,257	40,201
K-P F-stat	114.03	114.03	114.03	114.03	114.03	114.13	112.01
Reg. type F.E.	✓	✓	✓	✓	✓	✓	✓
Enl. date poly	✓	✓	✓	✓	✓	✓	✓
Enl. rank F.E.	✓	✓	✓	✓	✓	✓	✓
State F.E.	✓	✓	✓	✓	✓	✓	✓
Father controls	✓	✓	✓	✓	✓	✓	✓
Own controls	✓	✓	✓	✓	✓	✓	✓

Note: Regiment type fixed effects are for infantry, cavalry, artillery, and specialized units such as sharpshooters. Geographic fixed effects are for the residence of fathers. The enlistment date polynomial controls for the time of exposure to the war, and enlistment rank fixed effects include dummies for the initial rank (85% of soldiers start as privates). Father controls include information from the 1860 Census on father's age, nationality, literacy, wealth (house value), and occupational skill group. A son's own controls include his age, age squared, and literacy in 1860. Standard errors clustered by regiment in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 2.9: Death and Disability Effects by Father's Occupation

	Child's 1880 Occupation			
	high-skilled	semi_skilled	low-skilled	farmer
Fth died × fth high-skilled	-0.016	0.000	-0.010	0.018
	(0.025)	(0.032)	(0.025)	(0.022)
Fth died × fth semi-skilled	-0.007	-0.054***	0.040***	0.003
	(0.009)	(0.014)	(0.013)	(0.009)
Fth died × fth low-skilled	0.006	-0.016	-0.014	0.017
	(0.008)	(0.016)	(0.017)	(0.013)
Fth died × fth farmer	-0.008	-0.004	0.003	0.016
	(0.005)	(0.009)	(0.011)	(0.011)
Fth disabled × fth high-skilled	-0.025	-0.035	0.063**	-0.015
	(0.025)	(0.030)	(0.027)	(0.020)
Fth disabled × fth semi-skilled	-0.014*	-0.016	0.027**	0.007
	(0.008)	(0.013)	(0.012)	(0.009)
Fth disabled × fth low-skilled	-0.002	-0.011	-0.014	0.010
	(0.009)	(0.018)	(0.018)	(0.014)
Fth disabled × fth farmer	0.009	-0.017	0.016	-0.004
	(0.007)	(0.011)	(0.012)	(0.012)
Observations	40,963	40,963	40,963	40,963
Reg. type F.E.	✓	✓	✓	✓
Enl. rank F.E.	✓	✓	✓	✓
State F.E.	✓	✓	✓	✓
Father controls	✓	✓	✓	✓
Own controls	✓	✓	✓	✓

Note: Regiment type fixed effects are for infantry, cavalry, artillery, and specialized units such as sharpshooters. Geographic fixed effects are for the residence of fathers. Enlistment rank fixed effects include dummies for the initial rank (85% of soldiers start as privates). Father controls include information from the 1860 Census on father's age, nationality, literacy, wealth (house value), and occupational skill group. A son's own controls include his age, age squared, and literacy in 1860. Standard errors clustered by family in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 2.10: Heterogeneity by Father's 1860 Wealth

	occupational score	high-skilled	semi-skilled	low-skilled	farmer	migrant	ever married
Father died ×	-0.072***	-0.009*	-0.029***	0.014	0.014*	-0.003	0.017**
bottom half	(0.018)	(0.005)	(0.009)	(0.010)	(0.008)	(0.010)	(0.008)
Father died ×	-0.024	-0.005	-0.007	0.007	0.012	-0.003	0.014
top half	(0.020)	(0.006)	(0.009)	(0.009)	(0.009)	(0.011)	(0.009)
Father disabled ×	-0.039*	-0.006	-0.011	0.007	0.002	-0.014	0.001
bottom half	(0.020)	(0.006)	(0.010)	(0.010)	(0.008)	(0.011)	(0.009)
Father disabled ×	-0.001	0.001	-0.014	0.023**	-0.002	0.015	-0.008
top half	(0.023)	(0.007)	(0.010)	(0.010)	(0.010)	(0.012)	(0.009)
Observations	40,968	40,968	40,968	40,968	40,968	41,021	39,969
Reg. type F.E.	✓	✓	✓	✓	✓	✓	✓
Enl. rank F.E.	✓	✓	✓	✓	✓	✓	✓
State F.E.	✓	✓	✓	✓	✓	✓	✓
Father controls	✓	✓	✓	✓	✓	✓	✓
Own controls	✓	✓	✓	✓	✓	✓	✓

Note: Regiment type fixed effects are for infantry, cavalry, artillery, and specialized units such as sharpshooters. Geographic fixed effects are for the residence of fathers. Enlistment rank fixed effects include dummies for the initial rank (85% of soldiers start as privates). Father controls include information from the 1860 Census on father's age, nationality, literacy, wealth (house value), and occupational skill group. A son's own controls include his age, age squared, and literacy in 1860. Standard errors clustered by family in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

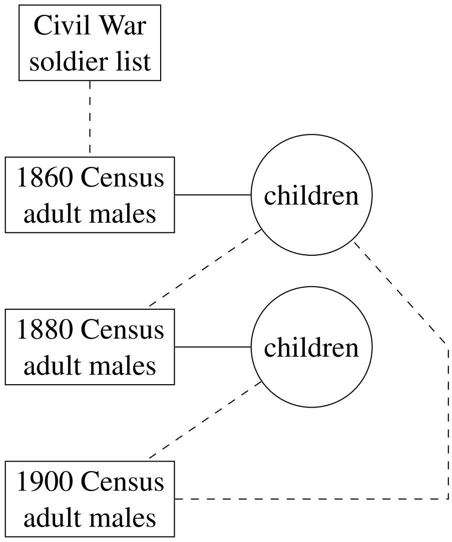
Table 2.11: Effect on the Grandchildren

	occupational score	high-skilled	semi-skilled	low-skilled	farmer	migrant	ever married
Granfather died	-0.083	-0.005	-0.042*	0.018	0.007	0.019	0.017
	(0.053)	(0.017)	(0.024)	(0.026)	(0.018)	(0.028)	(0.023)
Grandfather disabled	-0.106*	-0.025	-0.014	0.024	-0.024	0.013	0.015
	(0.059)	(0.019)	(0.027)	(0.028)	(0.017)	(0.030)	(0.025)
Observations	3,029	3,029	3,029	3,029	3,029	3,063	3,056
Reg. type F.E.	✓	✓	✓	✓	✓	✓	✓
Enl. rank F.E.	✓	✓	✓	✓	✓	✓	✓
State F.E.	✓	✓	✓	✓	✓	✓	✓
Father controls	✓	✓	✓	✓	✓	✓	✓
Own controls	✓	✓	✓	✓	✓	✓	✓

Note: Regiment type fixed effects are for infantry, cavalry, artillery, and specialized units such as sharpshooters. Geographic fixed effects are for the residence of fathers. Enlistment rank fixed effects include dummies for the initial rank (85% of soldiers start as privates). Father controls include information from the 1860 Census on father's age, nationality, literacy, wealth (house value), and occupational skill group. A son's own controls include his age, age squared, and literacy in 1900. Standard errors clustered by family in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

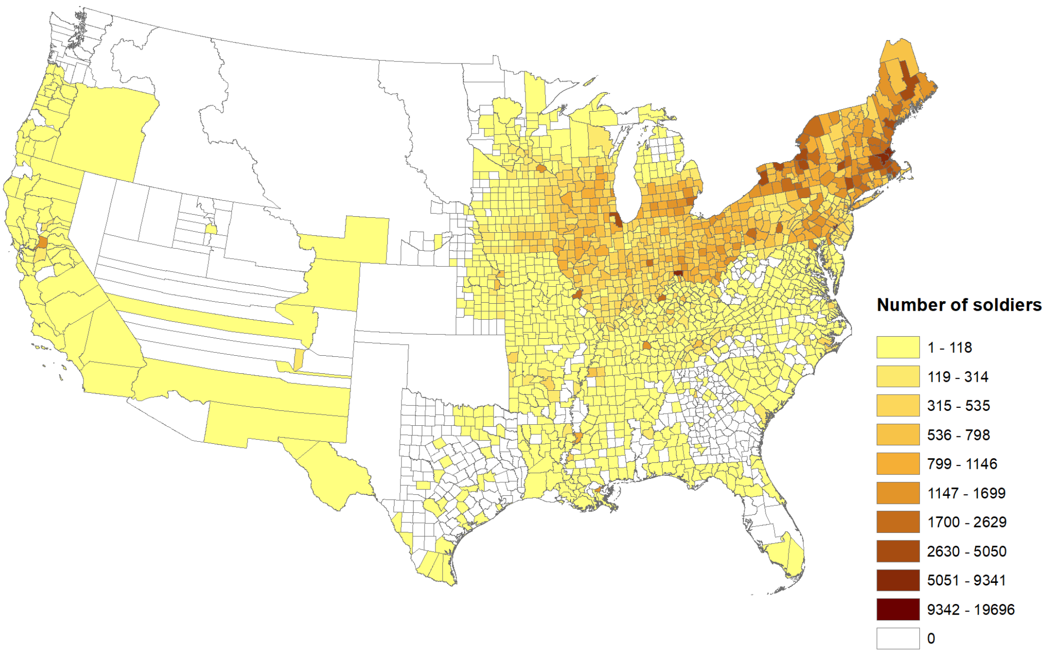
2.9 Figures

Figure 2.1: Record Linkage Schematic



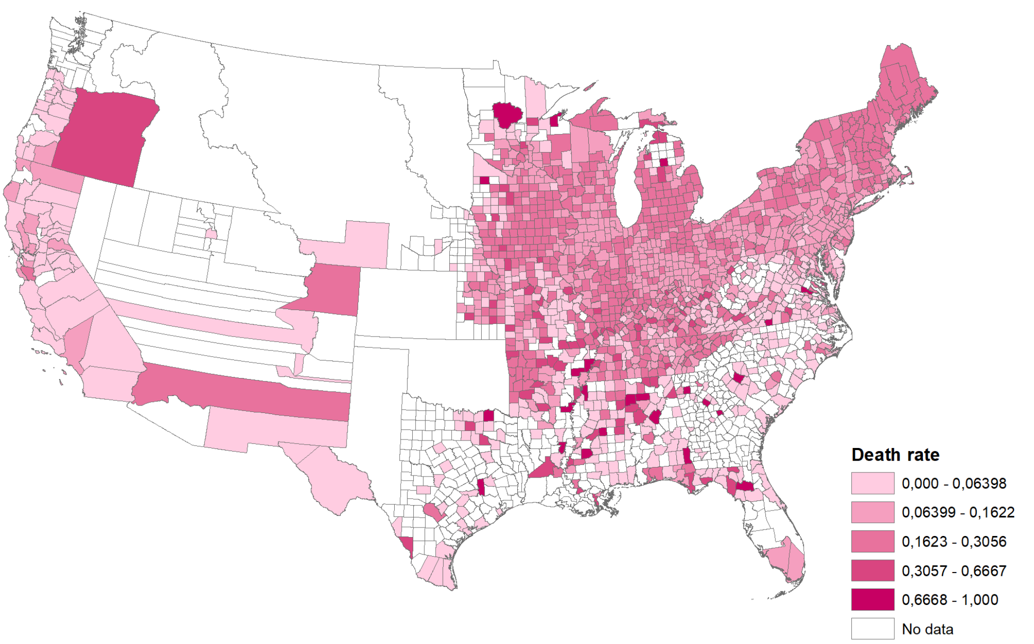
Note: Record linkage schematic showing the three-generation links created between military and Census records. Dashed lines are generated links, solid lines are available links inside a given data set.

Figure 2.2: Geographic Distribution of Soldiers Linked to the 1860 Census



Note: County-level distribution of soldiers enlisted in the Union Army in the county boundaries of 1860. These are all enlistments until the end of the war. Note that also a minority of soldiers hailed from the border states as well as the South.

Figure 2.3: Death Rates by County



Note: County-level mortality rates among soldiers who enlisted for the Union Army. All types of deaths are included. Note that also a minority of soldiers hailed from the border states as well as the South.

Figure 2.4: 22nd MA Volunteer Infantry Regiment Records Example

Twenty-Second Regiment Infantry, M. V.—(Three Years.)—Continued.

NAME AND RANK.	Age.	Bounty.	Residence or Place credited to.	Date of Muster.	Termination of Service and cause thereof.
Company E—Con.					
Murphy, Charles,	25	–	Roxbury,	Sept. 9, '61,	Killed June 27, 1862, Gaines' Mills, Va.
Nayson, William E.,	30	–	Roxbury,	13, '61,	Dec. 7, 1863, disability.
Nickerson, James,	30	–	Roxbury,	9, '61,	Killed July 1, 1862, Malvern, Hill, Va.
Nolan, Henry J.,	22	–	Boston,	9, '61,	Died Oct. 27, 1862, New York Harbor.
Norton, James,	30	–	Roxbury,	9, '61,	Dropped from rolls, Oct. 1, 1861.
Noyes, Joseph P.,	40	–	Lynn,	9, '61,	Oct. 21, 1862, disability.
Pearl, George W.,	18	–	Boston,	28, '61,	4, 1864, expiration of service.
Petterson, Leonard,	29	–	Roxbury,	18, '61,	Killed May 8, 1864, Laurel Hill, Va.
Pierce, Philip R. W.,	39	–	Roxbury,	9, '61,	Nov. 5, 1862, disability.
Quinn, William,	30	–	Roxbury,	13, '61,	Sept. 24, 1862, disability.
Ray, John, J.,	19	–	Boston,	9, '61,	Feb. 1, 1864, to re-enlist.
Raymond, William T.,	19	–	Roxbury,	9, '61,	Sept. 24, 1862, disability.
Richardson, James,	34	–	Roxbury,	9, '61,	Oct. 20, 1864, expiration of service.
Robinson, John,	43	–	Boston,	Aug. 28, '62,	Dec. 15, 1862, disability.

Note: Example entry from the Adjutant General Report for the state of Massachusetts showing the roster of the 22nd Massachusetts Volunteer Infantry Regiment for part of company E. Information on soldiers includes their full names, age, enlistment bounty (if any), place of residence, date of muster, and the date and cause of their exit from the unit.

2.10 Appendix

A Data Sources

Table 2.A.1: List of Sources for the Union Soldier Data

- **California**: Orton, R.H. (1890) "Records of California Men in the War of the Rebellion 1861 to 1867", State Office, J. D. Young, Supt. State Printing, Sacramento, CA
- **Connecticut**: Barbour, L.A., Camp, F.E., Smith, S.R., and White, G.M. (1889) "Record of Service of Connecticut Men in the Army and Navy of the United States During the War of the Rebellion", Case, Lockwood, & Brainard Company, Hartford, CT
- **Illinois**: Reece, J.N. (1900) "Report of the Adjutant General of the State of Illinois", Vols. 1-9, Philips Bros. State Printers, Springfield, IL
- **Indiana**: Terrell, W.H.H. (1866) "Report of the Adjutant General of the State of Indiana", Vols. 1-5, Samuel M. Douglass State Printers, Indianapolis, IN
- **Iowa**: Thrift, W.H. (1908) "Roster and Record of Iowa Soldiers in the War of Rebellion", Vol. 1-6, Emory H. English State Printers, Des Moines, IA
- **Kansas**: Fox, S.M. (1896) "Report of the Adjutant General of the State of Kansas", The Kansas State Printing Company, Topeka, KS
- **Maine**: Adjutant General (1861-66) "Supplement to the Annual Reports of the Adjutant General of the State of Maine", Stevens & Sayward State Printers, Augusta, ME
- **Massachusetts**: Schouler, W. (1866) "Report of the Adjutant General of the Commonwealth of Massachusetts", Wright & Potter State Printers, Boston, MA
- **Michigan**: Crapo, H.H. (1862-66) "Report of the Adjutant General of the State of Michigan", John A. Kerr & Co. State Printers, Lansing, MI
- **Minnesota**: Marshall, W.R. (1861-66) "Report of the Adjutant General of the State of Minnesota", Pioneer Printing Company, Saint Paul, MN
- **Nebraska**: Dudley, E.S. (1888) "Rosters of Nebraska Volunteers from 1861 to 1869", Wigton & Evans State Printers, Hastings, NB
- **New Hampshire**: Head, N. (1865) "Report of the Adjutant General of the State of New Hampshire", Vols. 1& 2, Amos Hadley State Printers, Concord, NH
- **New Jersey**: Stryker, W.S. (1874) "Report of the Adjutant General of the State of New Jersey", Wm. S. Sharp Steam Power Book and Job Printers, Trenton, NJ
- **New York**: Sprague, J.T. (1864-68) "A Record of the Commissioned Officers, Non-Commissioned Officers and Privates of the Regiments which were Organized in the State of New York into the Service of the United States to Assist in Suppressing the Rebellion", Vols. 1-8, Comstock & Cassidy Printers, Albany, NY
- **Ohio**: Howe, J.C., McKinley, W., and Taylor, S.M. (1893) "Official Rosters of the Soldiers of the State of Ohio in the War of the Rebellion 1861-65", Vols. 1-12, The Werner Company, Akron, OH
- **Pennsylvania**: Russell, A.L. (1866) "Report of the Adjutant General of Pennsylvania", Singerly & Myers State Printers, Harrisburg, PA
- **Vermont**: Peck, T.S. (1892) "Revised Roster of Vermont Volunteers and Lists of Vermonters who Served in the Army and Navy of the United States during the War of the Rebellion 1861-66", Press of the Watchman Publishing Co., Montpelier, VT
- **Wisconsin**: Rusk, J.M. and Chapman, C.P. (1886) "Roster of Wisconsin Volunteers, War of the Rebellion 1861-65", Democrat Printing Company, Madison, WI

Note: The table lists the main data sources from which the soldier and regiment data for this project originate. All of these can be accessed and downloaded from https://www.hathitrust.org and https://archive.org.

B Front Line Service and Socioeconomic Regiment Composition

A potential threat to our identification strategy is a correlation between military strategy and the socioeconomic composition of regiments. Suppose leaders place regiments from the poorest areas in the front lines where they have a higher probability of dying. Regression analyses might then attribute too much of the change in children's later-life outcomes to losing a father which absorbs the effect of the lower socioeconomic status. However, the opposite argument is also plausible when leaders want to occupy the front rows with the most able-bodied soldiers. In this case, we would underestimate the effect of losing a father when children come from the upper classes of society which have the means to alleviate such a loss with more wealth and household resources.

To test for such potential selection, we collected and digitized 128 battle maps from the Civil War Preservation Trust.[20] The idea is to compute the distance of Union regiments to the nearest enemy regiment in order to then regress these distances on the economic composition of Union units and their military characteristics. The maps provide information on the location of Union and Confederate regiments and maintain the same color codes and symbols throughout. Regiments are represented by rectangles and artillery units are marked with a canon symbol. Using pattern recognition techniques, we digitized the location of these symbols on each map. The color schemes were used to differentiate between Union and Confederate units, as well as different battle stages.[21]

For each Union unit, the distance to the nearest Confederate unit was computed for a given battle and battle stage as the point-to-point distance on the Cartesian plane. The distance measure therefore does not have an interpretation in geographic units. Generating a geographic distance variable is complicated by the fact that maps are on different scales. For this reason regressions will use log distances and battle fixed effects. Figure 2.B.1 provides an example.

This resulted in 4,147 unit-battle-stage locations for a total of 128 battles and 799 unique Union units. Battles tend to be large with an average number of 20.5 Union units where a typical infantry regiment consists of 1,000 men. To compute the economic composition of each regiment, we used the individual-level soldier data to link soldiers' residence county to economic and population data from the 1860 county-level Census. A given Census variable x_c

[20] The maps were retrieved from: https://www.battlefields.org/learn/maps on May 27th, 2018.

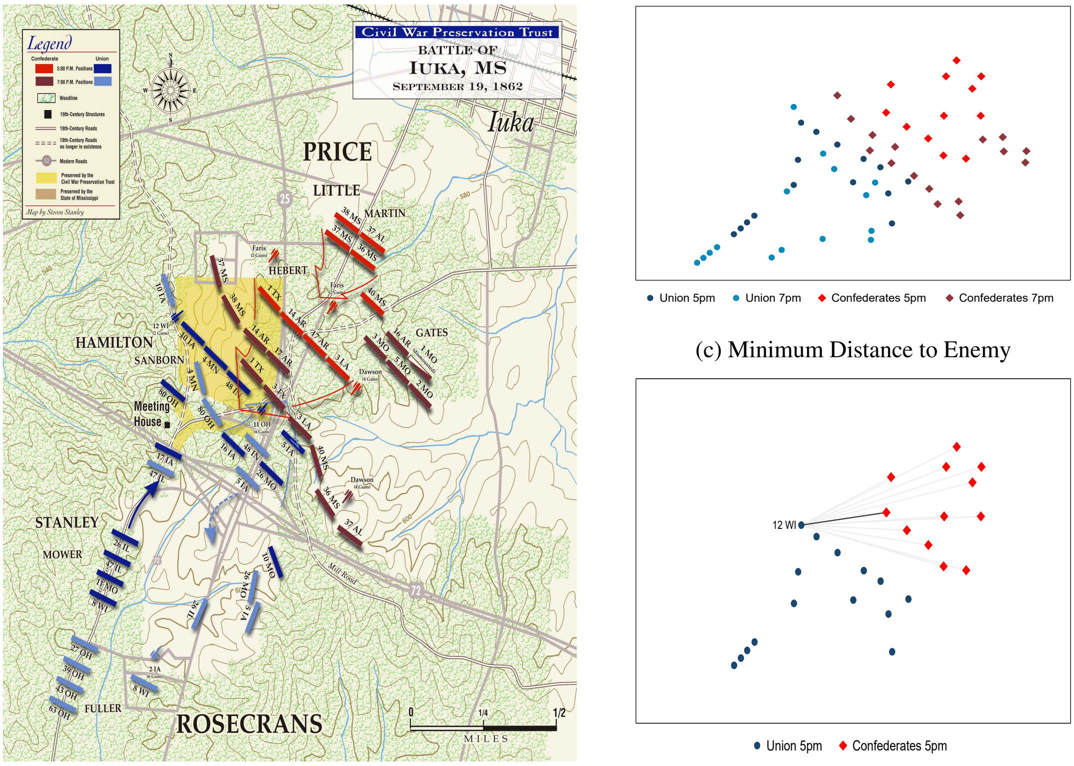
[21] 88 of the 128 maps show unit positions for different stages of a battle. This means that there is within-battle variation in the location of regiments. The average battle has 1.45 stages with a maximum of 5.

Figure 2.B.1: Digitizing Civil War Battle Maps

(a) Raw Battle Map

(b) Digitized Battle Map

(c) Minimum Distance to Enemy



Note: Panel a) shows the raw battle map for the Battle of Iuka, Mississippi on September 19, 1862. Union and Confederate regiment positions are shown for two phases of the battle. These are at 5pm (dark blue Union, light red Confederacy) and at 7pm (light blue Union, dark red Confederacy). Panel b) shows the digitized version of the map. Panel c) plots Union and Confederate regiments in their 5pm location, computes the distances to the closes enemy units from the 12th Wisconsin, and marks the minimum distance with a black rather than a gray line. The digitized maps look different due to the way in which they are displayed here, however, relative positions of the regiments to each other are not affected. Battle maps were obtained from the Civil War Preservation Trust (https://www.battlefields.org/learn/maps) and digitized by the authors via pattern recognition algorithms in Python.

for county $c = 1, 2, ..., C$ was then averaged to the regiment level,

$$\overline{x}_r = \frac{\sum_{c=1}^{C} x_c n_{rc}}{\sum_{c=1}^{C} n_{rc}}$$

where the weights $n_{rc} = \sum_{i=1}^{I} n_{irc}$ are the total number of soldiers in regiment r from county c. Variables taken from the 1860 Census are the average cash value, machinery, and livestock value per farm, the share of men aged 14 to 29, the share of employment in manufacturing, the average value of capital, and output per manufacturing establishment, the value of personal real estate per family, the number of churches per 1,000 inhabitants, the average value of church property, and the ratio of foreign- to native-born men.

The military regiment characteristics are the regiment type (infantry, cavalry, artillery), indicators for whether a unit belongs to the regular Army, the average enlistment age of soldiers in the unit, the share of fighting soldiers (to distinguish support units on the field), and measures for unit cohesion such as the total number of counties from which soldiers in the unit joined, and the shares of voluntarily enlisted, soldiers transferred into the unit, and the share of deserted soldiers. Note that most of these measures are only available at the end of the war. This means they should be thought of as totals. For instance, the number of counties in a regiment looks surprisingly large with an average of 30.5. This is mainly due to re-enlistments where soldiers stated a different county and transfers. Hence the average Union regiment had soldiers from about 31 different counties during the entire duration of the war. Summary statistics are reported in table 2.B.1.

Table 2.B.1: Battle Distance Summary Statistics

	Observations = 4,147			
	Mean	St. Dev.	Min.	Max.
Military Information				
Distance	254.240	278.327	5.099	2, 206.181
ln(Distance)	5.152	0.867	1.629	7.699
Number of Union units per battle	20.514	18.318	1	94
Number of battle stages	1.450	0.720	1	5
Infantry	0.948	0.221	0	1
Cavalry	0.030	0.170	0	1
Artillery	0.022	0.146	0	1
Regular Army	0.038	0.192	0	1
Mean enlistment age	25.267	2.426	16	39
Share fighting soldiers	98.544	4.062	70.461	100
Share enlisted	90.456	12.070	17.670	100
Share deserted	6.645	6.911	0	40.970
Counties present in unit	30.572	24.467	1	161
County Information				
Share men aged 14-29	69.225	3.166	52.285	77.579
Ratio of foreign to native men	0.317	0.230	0.004	1.474
Mean farm value	10,630.411	17, 488.969	803.022	80, 026.117
Mean machinery value per farm	148.403	83.505	50.444	425.238
Mean value of livestock per farm	472.014	132.702	173.590	1, 639.027
% employed in manufacturing	4.523	3.457	0.241	20.084
Mean capital value per firm	8,064.809	4, 530.886	1, 512.564	46, 688.063
Mean value of output per firm	15,764.820	9, 320.380	3, 229.907	65, 403.676
Value of real estate per family	935.332	527.008	360.179	13, 141.862
No. churches per 1,000 population	1.569	0.675	0	5.120
Mean value of church property	9,641.684	11, 427.625	0	45, 486.945

Note: Summary statistics for the 4,147 unit-battle observations for 799 Union regiments in 128 Civil War battles. Distance to the nearest enemy unit is measured as point-to-point distance on the Cartesian plane. County characteristics are weighted averages at the regiment level. These were computed as the mean characteristic from all counties represented in a regiment weighted by the number of soldiers in the regiment from each county.

The test for selection into front line service amounts to regressing,

$$\ln(\text{distance})_{rbs} = \delta_b + \phi_s + \overline{x}_r'\gamma + m_r'\beta + \eta_{rbs} \tag{2.6}$$

where the outcome is the natural logarithm of a Union unit's distance to the nearest enemy unit in a given battle b and battle stage s. The vectors $\overline{x}_r$ and m_r contain the economic composition information and military characteristics of the unit, respectively. Battle fixed effects δ_b account for the different geographic scaling of maps while phase fixed effect ϕ_p absorb systematic location differences between earlier and later stages of a battle. Standard errors are clustered at the battle level.

Results are reported in table 2.B.2. Columns 1 and 2 show the fixed effects only regressions for battles with more than one stage. When adding regiment fixed effects, the adjusted R^2 increase from 47.2 to 49.5 which implies that unobserved time-invariant regiment characteristics are not a major determinant of their distance to the nearest enemy unit. Columns 3 and 4 add military and economic characteristics separately, and jointly in column 5. Again, the adjusted R^2 barely changes and none of the coefficients is a significant correlate with the distance measure in any regression specification.

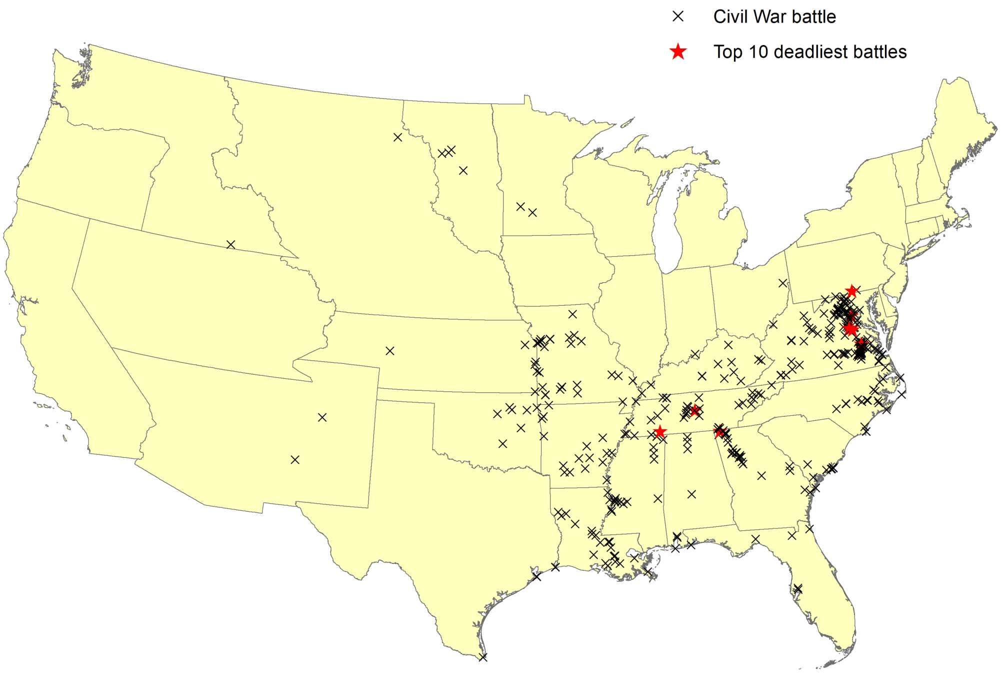
For the majority of variables these coefficients are tightly estimated zeroes and are not just insignificant due to measurement error in the outcome. The only coefficients with an economically sizable magnitude are those for the artillery dummy, however, this is imprecisely estimated. It should also be noted that there are only 16 black regiments among our 799 units because there were very few black combat units. Overall there seems to be little evidence for military, economic, and time-invariant regiment specific characteristics to play an important role in the determination of units' front line proximity.

Table 2.B.2: Determinants of Distance to Nearest Enemy on the Battlefield

	Outcome: log distance to nearest enemy unit				
	(1)	(2)	(3)	(4)	(5)
Cavalry			0.002		0.005
			(0.060)		(0.061)
Artillery			-0.090		-0.087
			(0.060)		(0.062)
Regular Army			0.034		0.082
			(0.085)		(0.091)
Enlistment age			-0.004		-0.004
			(0.004)		(0.004)
% combat soldiers			0.001		0.001
			(0.003)		(0.004)
% enlisted			0.001		0.002
			(0.001)		(0.002)
County diversity			0.000		0.000
			(0.001)		(0.001)
% deserted			-0.002		0.000
			(0.002)		(0.003)
Mean farm value				0.000	0.000
				(0.000)	(0.000)
Mean farm machinery value				-0.001	-0.001
				(0.000)	(0.000)
Mean value of livestock				-0.000	-0.000
				(0.000)	(0.000)
% employed in manufact.				0.001	0.003
				(0.007)	(0.007)
Manufact. capital value				0.000	0.000
				(0.000)	(0.000)
Manufact. output value				-0.000	-0.000
				(0.000)	(0.000)
Mean real estate value				-0.000	-0.000
				(0.000)	(0.000)
Churches per 1k pop.				0.023	0.027
				(0.028)	(0.030)
Value of church property				0.000	0.000
				(0.000)	(0.000)
Ratio foreign to native men				0.065	0.072
				(0.079)	(0.080)
Share men aged 14-29				0.000	0.000
				(0.007)	(0.006)
Observations	3,065	3,065	4,147	4,147	4,147
Battles	88	88	128	128	128
Adj. R^2	0.472	0.495	0.499	0.499	0.498
Regiment FE		Yes			

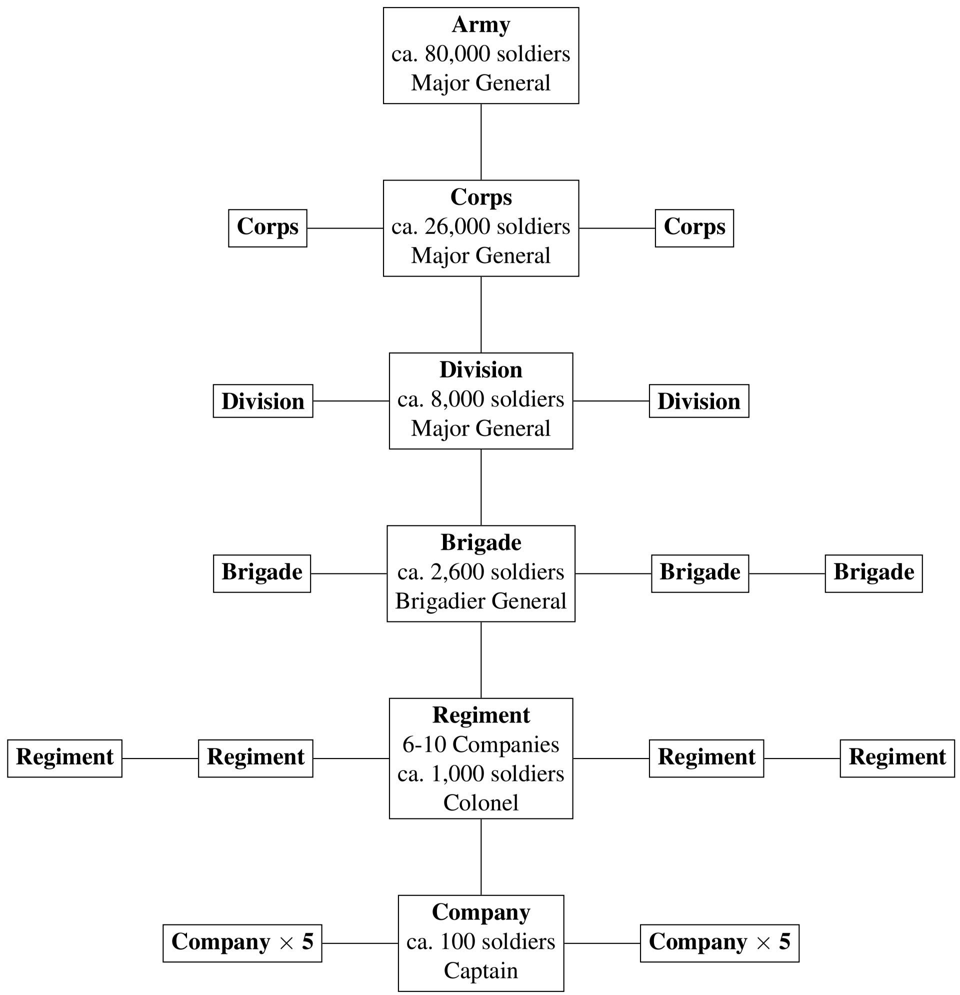
Note: Regressions of the log point-to-point distance of Union regiments to the nearest Confederate unit on military characteristics and measures of the socioeconomic composition of Union units. Columns (1) and (2) report fixed effects regressions for battles with multiple stages only (88 out of 128 battles). County characteristics are weighted averages at the regiment level. These were computed as the mean characteristic from all counties represented in a regiment weighted by the number of soldiers in the regiment from each county. All regressions include battle and battle stage fixed effects. Standard errors clustered at the battle level. Significance levels are denoted by * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Figure 2.B.2: Map of Civil War battles



Note: Map of battles during the U.S. Civil War with major battles marked by a red star for the 10 battles with the highest number of deaths. These are: Gettysburg,Wilderness, Spotsylvania, Chancellorsville, Chickamauga, Seven Days, 2nd Bull Run, Shiloh, Stones River, and Cold Harbor.

Figure 2.B.3: Union Army Organizational Chart



Note: Typical structure of an Infantry Army of the Union. In total there were 16 Armies. One Army typically contained 2-3 Corps, a Corps had 2-3 Divisions, a Division had 2-4 Brigades, a Brigade had 2-5 Regiments, and a Regiment had 6-10 companies of a 100 men each. Armies, Corps, and Divisions were commanded by Major Generals, Brigades by Brigade Generals, Regiments by Colonels, and Companies by Captains.

C Union Army pensions

The Union Army pension program was America's first social insurance program (Skocpol, 1992; Costa, 1995, 1997; Salisbury, 2017). Though it slowly became a universal disability and old-age pension program for veterans of the Union Army, it began as a restricted program compensating soldiers disabled by the war as well as the relatives of deceased soldiers. Assessing the generosity of these pensions, as well as take-up, is important for us to understand the mechanisms behind the effect of father loss on sons' socioeconomic outcomes, as monetary compensations can partly shut down the income channel.

Compensation amounts were determined by a series of laws passed between 1862 and 1873. Pensions for disability depended on rank and type of injury. Originally a private totally disabled for manual labor received $ 8 per month, and widows of deceased soldiers received the same amount. This represented less than 1/2 of the monthly income of a farm laborer (Salisbury, 2017). If a widow remarried, the pension was given to the minor children (younger than 16) of the deceased soldier (Salisbury, 2017).

Pensions were gradually increased in the decades following the war. In 1870, the average monthly pension received by invalid veterans was $ 8.7 (about 20% of the unskilled wage) and the average monthly pension received by widows and dependents was $ 14 (about a third of the unskilled wage) (Glasson, 1918; Long, 1960).[22] Take-up, however, was extremely low: Skocpol (1992) estimates that only 25% of the survivors of Union soldiers killed during the war received dependent pensions in 1875, and that only 43% of wounded men claimed a pension.

[22] Total amounts and number of pensioner from Glasson (1918), average daily unskilled wage from Long (1960). We assume workers work 26 days in a month.

D The Bias of OLS and IV Resulting from Linkage Errors

The linking of Census or other historical records without individual identifiers has become a very active research area. Since the first rare-name matching algorithm introduced by Ferrie (1996), more recent papers have introduced supervised (Feigenbaum, 2016) and unsupervised (Abramitzky, Mill and Perez, 2018b) machine learning techniques for automated record linkage, as well as evaluations of the performance of such algorithms (Bailey et al., 2017). While a lot of effort is currently devoted to producing more accurate and faster linkage techniques and best practice guides to establish a unified approach (Abramitzky, Boustan, Eriksson, Feigenbaum and Perez, 2018a), we know relatively little about what happens to our OLS and IV estimates when we get those links wrong. Abramitzky et al. (2018b) state that a promising direction for future research, "is how to adjust regression coefficients when dealing with imperfectly linked data." (p. 11).

Thinking about the impact of record linkage errors on different types of estimators is conceptually challenging because this depends on the nature of the right-hand side variable of interest, whether linkage errors are systematically related to individuals' characteristics,[23] and on the number of data sets that need to be linked, e.g. if an instrument comes from an additional data set.

In the following, we provide a first attempt at quantifying a highly simplified worst-case scenario. Assume that we linked two data sets such as the 1860 and 1880 U.S. Census. In the case of this paper, let the true share of orphans be denoted by $T^* = \Pr(x^* = 1)$, where a child with $x^* = 1$ is truly an orphan. Variables with a superscript asterisk denote true values, individual subscripts i are omitted for clarity. In the linked sample, we observe a share of $\tilde{T} = \frac{1}{N}\sum x$ individuals marked as orphans, and a share of $\tilde{C} = (1 - \tilde{T})$ individuals who are marked as non-orphans.[24] Among the children marked as orphans, ν are actually non-orphans and among the children marked as non-orphans, η have lost a father but this error is not observed by the econometrician.

Assume the extreme case wherein every linkage error also results in a reversion in treatment status. The mis-measured orphan status can be thought of as measurement error and this error is non-classical. Whenever a child is wrongly marked as orphan, the only other value that the true orphan status can take is the exact opposite ($x = 1, x^* = 0$). This induces a negative correlation between the true and observed treatment status. This is the framework considered

[23] For instance, individuals with longer names can be linked more accurately because they contain more information and are usually rarer than shorter names. However, longer names have been shown to correlate with higher incomes and levels of education (Bailey et al., 2017).

[24] T and C denote the treatment and control group, respectively.

by Aigner (1973) who shows that measurement error in a binary treatment attenuates OLS estimates.

The true share of orphans relates to the observed quantities as,

$$T^* = (1-\nu)\tilde{T} + \eta\tilde{C} \tag{2.7}$$

and the mis-measured orphan status can be expressed as

$$x = x^* + u \tag{2.8}$$

where u is the error induced by wrong record linkages, and $x^* \sim \text{Ber}(T)$ and $x \sim \text{Ber}(\tilde{T})$. To derive the bias of the OLS estimator, Aigner (1973) states the following quantities:

$$\begin{aligned}
\mathbb{E}(u) &= \nu(\tilde{T}) - \eta\tilde{C} \\
Var(u) &= \nu\tilde{T} + \eta\tilde{C} - (\nu\tilde{T} - \eta\tilde{C})^2 \\
Cov(x,u) &= (\nu+\eta)\tilde{T}\tilde{C}.
\end{aligned}$$

Then for the model $y = \alpha + \beta x^* + \epsilon$, the OLS estimator is,

$$\begin{aligned}
\widehat{\beta}_{\text{OLS}} &= \frac{Cov(\alpha + \beta x^* + \epsilon, x^* + u)}{Var(x)} \\
&= \beta\left[\frac{Var(x^*) + Cov(x^*,u)}{Var(x)}\right] \\
&= \beta\left[\frac{T(1-T) + Cov(x,u) - Var(u)}{\tilde{T}(1-\tilde{T})}\right]
\end{aligned} \tag{2.9}$$

Now substitute the following quantities into (2.9),

$$\begin{aligned}
Var(x^*) &= T(1-T) \\
&= \left[(1-\nu)\tilde{T} + \eta\tilde{C}\right]\left[1 - (1-\nu)\tilde{T} - \eta\tilde{C}\right] \\
&= (1-\nu)\tilde{T} - \left[(1-\nu)\tilde{T}\right]^2 - 2\eta\tilde{T}\tilde{C}(1-\nu) + \eta\tilde{C} - \left[\eta\tilde{C}\right]^2 \\
Cov(x,u) &= \nu\tilde{T}\tilde{C} + \eta\tilde{T}\tilde{C} \\
Var(u) &= -\nu\tilde{T} - \eta\tilde{C} + \left[\nu\tilde{T}\right]^2 - 2\eta\nu\tilde{T}\tilde{C} + \left[\eta\tilde{C}\right]^2
\end{aligned}$$

to derive the OLS bias as,

$$
\begin{aligned}
\widehat{\beta}_{\text{OLS}} &= \beta \left[\frac{T(1-T) + Cov(x,u) - Var(u)}{\tilde{T}(1-\tilde{T})} \right] \\
&= \beta \left[\frac{\left[(1-\nu)\tilde{T} + \eta\tilde{C}\right]\left[1-(1-\nu)\tilde{T} - \eta\tilde{C}\right] + (\nu\tilde{T}\tilde{C} + \eta\tilde{T}\tilde{C})}{\tilde{T}(1-\tilde{T})} \right] \\
&+ \beta \left[\frac{-\nu\tilde{T} - \eta\tilde{C} + \left[\nu\tilde{T}\right]^2 - 2\eta\nu\tilde{T}\tilde{C} + \left[\eta\tilde{C}\right]^2}{\tilde{T}(1-\tilde{T})} \right] \\
&= \beta \left[\frac{\tilde{T} - \nu\tilde{T} - \tilde{T}^2 + 2\nu\tilde{T} - \left[\nu\tilde{T}\right]^2 + 2\eta\nu\tilde{T}\tilde{C} - 2\eta\tilde{T}\tilde{C} + \eta\tilde{C} - \left[\eta\tilde{C}\right]^2 + \nu\tilde{T}\tilde{C} + \eta\tilde{T}\tilde{C}}{\tilde{T}(1-\tilde{T})} \right] \\
&+ \beta \left[\frac{-\nu\tilde{T} - \eta\tilde{C} + \left[\nu\tilde{T}\right]^2 - 2\nu\eta\tilde{T}\tilde{C} + \left[\eta\tilde{C}\right]^2}{\tilde{T}(1-\tilde{T})} \right] \\
&= \beta \left[\frac{\tilde{T} - \tilde{T}^2 - 2\nu\tilde{T} + 2\nu\tilde{T}^2 - \eta\tilde{T}\tilde{C} + \nu\tilde{T}\tilde{C}}{\tilde{T}(1-\tilde{T})} \right] \\
&= \beta \left[\frac{\tilde{T} - \tilde{T}^2 - 2\nu\tilde{T} + 2\nu\tilde{T}^2 - \eta\tilde{T}(1-\tilde{T}) + \nu\tilde{T}(1-\tilde{T})}{\tilde{T}(1-\tilde{T})} \right] \\
&= \beta \left[\frac{\tilde{T}(1-\tilde{T}) - \nu\tilde{T}(1-\tilde{T}) - \eta\tilde{T}(1-\tilde{T})}{\tilde{T}(1-\tilde{T})} \right] \\
&= \beta \left[1 - \nu - \eta\right] \qquad (2.10)
\end{aligned}
$$

It follows from (2.10) that OLS is biased towards zero for a type I error rate of $\nu + \eta < 1$. For very high error rates that are $\nu + \eta > 1$, the OLS estimate will reverse in sign. Note that if all true orphans are wrongly classified as non-orphans ($\eta = 1$) and if all true non-orphans are classified as orphans ($\nu = 1$), then OLS will recover the true coefficient but with the opposite sign.

For the IV estimator, assume that we have an instrumental variable z which relates to the true orphan status via the first stage regression,

$$x^* = \pi_0 + \pi_{x^*z}z + \xi \qquad (2.11)$$

and that satisfies the exclusion restriction. Let $\delta_{yz} = \frac{Cov(y,z)}{Var(z)}$ denote the reduced form coefficient from the regression of y on z. An IV estimate can then be constructed as,

$$\widehat{\beta}_{\text{IV}} = \frac{\delta_{yz}}{\pi_{x^*z}} \qquad (2.12)$$

however, while the reduced form is unbiased, the first stage is not. This is because instead of x^* we observe the mis-measured x. Meyer and Mittag (2017) show that the OLS estimate of the first stage with the mis-measured binary dependent variable will be

$$\pi_{xz} = (1 - \nu - \eta)\pi_{x^*z}$$

and therefore the bias of the IV estimator is,

$$\begin{aligned}\widehat{\beta}_{\text{IV}} &= \frac{\delta_{yz}}{\pi_{xz}} \\ &= \frac{\delta_{yz}}{(1 - \nu - \eta)\pi_{x^*z}} \\ &= \frac{1}{1 - \nu - \eta}\beta_{\text{IV}}\end{aligned} \tag{2.13}$$

The IV bias is the inverse of the OLS bias. For the case where $\nu + \eta = 1$ exactly, the IV estimator does not exist. And again, if treatment and control group are switched around with $\nu + \eta = 2$, also the IV estimator recovers the true parameter with the opposite sign.

How does this result relate to practice? The typical type I error rate of automated linkage methods in Bailey et al. (2017) ranges between 0.22 and 0.69. For the lowest error rate, OLS will be attenuated to 78% and IV will be inflated to 128% of the true coefficient value. For the highest error rate instead, OLS will only be 31% and IV will be 323% of the true coefficient. Even though the scenario described here is highly simplified and a worst-case situation in which each wrong link leads to a treatment status change, the example shows how linkage errors can potentially lead to large differences between OLS and IV estimates which cannot be motivated with the typical LATE explanation.

Also note that, in the absence of other endogeneity problems, OLS and IV will set identify the true parameter value by providing lower and upper bounds $\widehat{\beta}_{\text{OLS}} < \beta < \widehat{\beta}_{\text{IV}}$. Without further assumptions, these bounds are sharp. This means that even in the presence of linkage errors the OLS and IV estimates can be informative.

Future work could potentially extend this framework to more realistic scenarios which allow for

1. different distributional assumptions on the outcome and treatment (binary, discrete, continuous)

2. changes in the outcome, the treatment, or both due to linkage errors where wrong links change these quantities only with a certain probability (e.g. we might have wrongly

linked to individuals but both of them are true orphans, hence $x^* = x$ despite the linkage error)

3. multiple links across different data sets

4. linkage error when the instrument comes from another data set for binary, discrete, or continuous instruments, where linkage errors in the instrument may correlate with linkage errors in the treatment (which would violate the exclusion restriction)

5. differential linkage errors by observed and unobserved individual characteristics such as name length, the complexity of a name (e.g. the number of rare letters per name), rarity of the name, among others

E Evidence from a Simulation Exercise

To test the theoretical framework above, we simulate a data set of 10,000 individuals, half of whom are in the treatment and control group respectively, $T = C = 0.5$. For 10% of individuals on both groups we then assume a linkage error that reverses their treatment status, such that $x = 1 - x^*$, implying a total error rate of $\nu + \eta = 0.1 + 0.1 = 0.2$, which is roughly the type I error rate found for the Ferrie (1996) rare-name linkage algorithm in Bailey et al. (2017). The observed treatment status x is then generated as described above with $x = x^* + u$.

The true estimating equation is,

$$y_i = 1x_i^* + \epsilon_i \tag{2.14}$$

where $\epsilon_i \sim N(0, 1)$ is an *iid* error term, and the coefficient of the true treatment effect is $\beta = 1$. Suppose we have a valid instrument z which relates to x^* via the first stage regression,

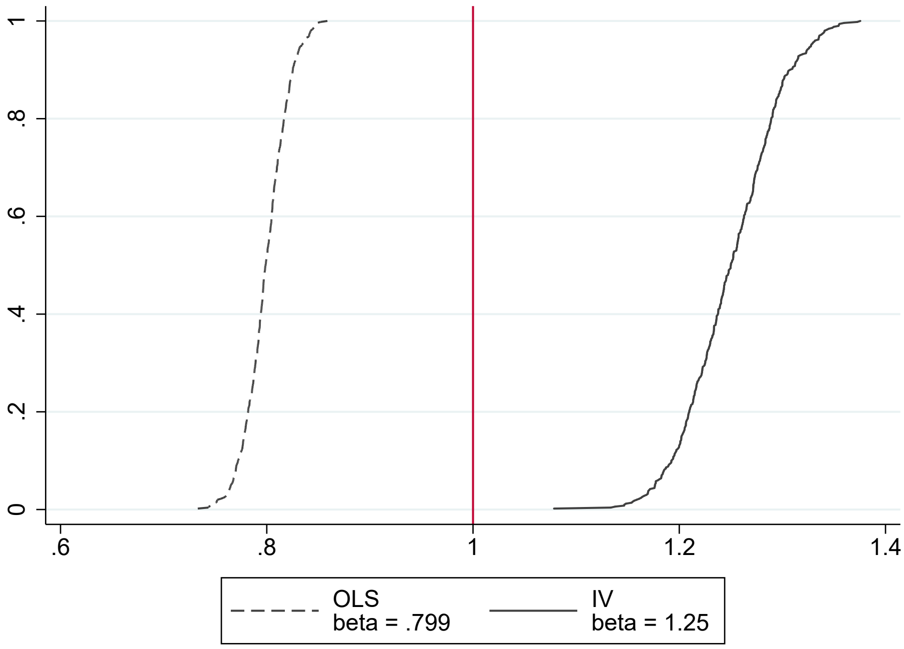
$$x_i^* = \frac{2}{3}z_i + \xi_i \tag{2.15}$$

with $\xi_i \sim N(0, 1)$ *iid* errors, a first stage coefficient $\pi = \frac{2}{3}$, and $Corr(\epsilon, \xi) = 0$.[25] We simulate (2.14) by substituting x^* with x and we do this 500 times to observe the behavior of the OLS and IV estimates. The CDFs of the OLS and IV estimates obtained from these 500 simulations are graphically reported in figure 2.E.1 and numerically in table 2.E.1.

As predicted by the theory outlined in the previous section, OLS recovers 80% of the true parameter value while IV is inflated to 125% of the true coefficient. Note that IV has more than twice the dispersion of OLS, yet none of the two estimators includes the true value in their 95% confidence interval. In practice, however, this will depend on the strength of the first stage and whether any other endogeneity concerns are present. The true first stage coefficient is estimated when using the treatment variable without linkage error which yields $\widehat{\pi}_{x^*z} = 0.6669$, while the first stage with the mis-measured treatment produces the predicted coefficient of $(1 - \nu - \eta)\pi_{x^*z} = (1 - 0.2)\frac{2}{3} = 0.5338$. Also the simulation confirms that $\widehat{\beta}_{\text{OLS}} < \beta < \widehat{\beta}_{\text{IV}}$, given that no other endogeneity problem was simulated.

[25] The distinction of whether z is binary or continuous does not matter in this context.

Figure 2.E.1: Simulated OLS and IV Bias with Mis-Measured Binary Treatment due to Linkage Errors



Note: OLS and IV CDFs from 500 simulations of a data set with 10,000 individuals, half of whom are in the treatment group. Misclassification rates for both treatment and control are set to 0.1 each (i.e. a total misclassification error of 20%) and a true treatment effect of 1 which is marked by the red line. The figure reports the median bias of OLS and IV below the graph.

Table 2.E.1: Summary Statistics for Simulated OLS and IV Estimations with a Mis-Measured Binary Treatment due to Linkage Errors

	obs.	mean	st. dev.	min	max
$\widehat{\beta}_{\text{OLS}}$	500	0.7994	0.0207	0.7331	0.8588
$\widehat{\beta}_{\text{IV}}$	500	1.2504	0.0458	1.0785	1.3756
$\widehat{\pi}_{x^*z}$	500	0.6669	0.0031	0.6556	0.6751
$\widehat{\pi}_{xz}$	500	0.5338	0.0072	0.5081	0.5554

Note: Summary statistics for OLS, IV and first stage estimates from 500 simulations of a data set with 10,000 individuals, half of whom are in the treatment group. Misclassification rates for both treatment and control are set to 0.1 each (i.e. a total misclassification error of 20%). Rows from top to bottom are for the OLS estimator $\widehat{\beta}_{\text{OLS}}$, the IV estimator $\widehat{\beta}_{\text{IV}}$, the first stage using the true treatment variable as outcome $\widehat{\pi}_{x^*z}$, and the first stage using the mis-measured treatment as outcome $\widehat{\pi}_{xz}$.

F Additional results

Table 2.F.1: Effect of Losing a Father and One or More Brothers

	occupational score	high-skilled	semi-skilled	low-skilled	farmer	migrant	ever married
Father died	-0.155***	-0.005	-0.074***	0.052**	0.026	-0.042	-0.008
	(0.046)	(0.015)	(0.021)	(0.023)	(0.022)	(0.027)	(0.021)
brother(s) died	-0.006	-0.028**	0.017	-0.023	0.039*	0.019	-0.005
	(0.046)	(0.012)	(0.022)	(0.022)	(0.024)	(0.025)	(0.020)
Father disabled	0.044	0.012	-0.007	-0.002	-0.003	-0.021	-0.007
	(0.051)	(0.015)	(0.024)	(0.023)	(0.024)	(0.027)	(0.022)
brother(s) disabled	0.053	0.022	-0.015	0.007	-0.032	-0.008	0.029
	(0.085)	(0.024)	(0.032)	(0.031)	(0.029)	(0.037)	(0.030)
Observations	3,483	3,483	3,483	3,483	3,483	3,489	3,415
Reg. type F.E.	✓	✓	✓	✓	✓	✓	✓
Enl. rank F.E.	✓	✓	✓	✓	✓	✓	✓
State F.E.	✓	✓	✓	✓	✓	✓	✓
Father controls	✓	✓	✓	✓	✓	✓	✓
Own controls	✓	✓	✓	✓	✓	✓	✓

Note: Regiment type fixed effects are for infantry, cavalry, artillery, and specialized units such as sharpshooters. Geographic fixed effects are for the residence of fathers. Enlistment rank fixed effects include dummies for the initial rank (85% of soldiers start as privates). Father controls include information from the 1860 Census on father's age, nationality, literacy, wealth (house value), and occupational skill group. A son's own controls include his age, age squared, and literacy in 1860. Standard errors clustered by family in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 2.F.2: IV Results Controlling for Disabled Fathers

	occupational score	high-skilled	semi-skilled	low-skilled	farmer	migrant	ever married
Father died	-0.232	-0.008	-0.171**	0.115	0.040	0.330***	-0.075
	(0.166)	(0.047)	(0.080)	(0.081)	(0.065)	(0.095)	(0.069)
Father disabled	-0.057	-0.002	-0.043**	0.037**	0.005	0.071***	-0.029*
	(0.038)	(0.011)	(0.018)	(0.018)	(0.015)	(0.022)	(0.016)
Observations	34,056	34,056	34,056	34,056	34,056	34,098	33,239
K-P F-stat	123.12	123.12	123.12	123.12	123.12	123.14	119.49
Reg. type F.E.	✓	✓	✓	✓	✓	✓	✓
Enl. date poly	✓	✓	✓	✓	✓	✓	✓
Enl. rank F.E.	✓	✓	✓	✓	✓	✓	✓
State F.E.	✓	✓	✓	✓	✓	✓	✓
Father controls	✓	✓	✓	✓	✓	✓	✓
Own controls	✓	✓	✓	✓	✓	✓	✓

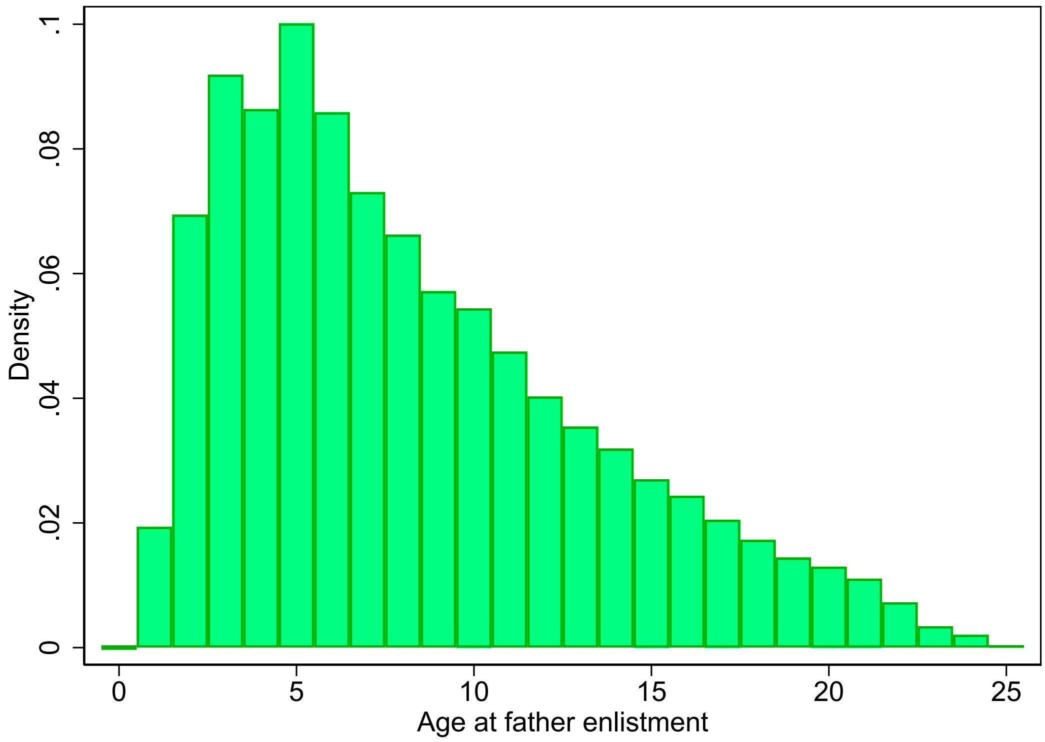
Note: Regiment type fixed effects are for infantry, cavalry, artillery, and specialized units such as sharpshooters. Geographic fixed effects are for the residence of fathers. Enlistment rank fixed effects include dummies for the initial rank (85% of soldiers start as privates). Father controls include information from the 1860 Census on father's age, nationality, literacy, wealth (house value), and occupational skill group. A son's own controls include his age, age squared, and literacy in 1900. Standard errors clustered by regiment in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 2.F.3: Effect of Dead or Disabled Fathers, restriction to uncommon names

	occupational score	high-skilled	semi-skilled	low-skilled	farmer	migrant	ever married
Father died	-0.044	0.013	-0.043*	0.026	-0.001	0.020	0.035*
	(0.047)	(0.014)	(0.023)	(0.022)	(0.020)	(0.025)	(0.021)
Father disabled	-0.009	0.000	-0.035	0.047**	-0.020	0.040	-0.012
	(0.049)	(0.015)	(0.025)	(0.023)	(0.018)	(0.027)	(0.023)
Observations	7,418	7,418	7,418	7,418	7,418	7,430	7,228
Reg. type F.E.	✓	✓	✓	✓	✓	✓	✓
Enl. rank F.E.	✓	✓	✓	✓	✓	✓	✓
State F.E.	✓	✓	✓	✓	✓	✓	✓
Father controls	✓	✓	✓	✓	✓	✓	✓
Own controls	✓	✓	✓	✓	✓	✓	✓

Note: Regiment type fixed effects are for infantry, cavalry, artillery, and specialized units such as sharpshooters. Geographic fixed effects are for the residence of fathers. Enlistment rank fixed effects include dummies for the initial rank (85% of soldiers start as privates). Father controls include information from the 1860 Census on father's age, nationality, literacy, wealth (house value), and occupational skill group. A son's own controls include his age, age squared, and literacy in 1900. Standard errors clustered by family in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Figure 2.F.1: Distribution of child's age when father enlisted



Note: Age distribution of children in the 1860 Census at the time of their father's enlistment in the Union Army.

Table 2.F.4: Heterogeneity by child's age when father enlisted

	occupational score	high-skilled	semi-skilled	low-skilled	farmer	migrant	ever married
Father died	-0.041**	-0.007	-0.006	-0.000	0.009	0.003	0.014*
	(0.018)	(0.005)	(0.009)	(0.010)	(0.008)	(0.010)	(0.008)
Father died ×	0.003	0.004	-0.018	0.016	-0.001	-0.022	-0.000
age at enlistment≥8	(0.028)	(0.008)	(0.013)	(0.013)	(0.012)	(0.014)	(0.012)
Observations	40,667	40,667	40,667	40,667	40,667	40,721	39,676
Reg. type F.E.	✓	✓	✓	✓	✓	✓	✓
Enl. rank F.E.	✓	✓	✓	✓	✓	✓	✓
State F.E.	✓	✓	✓	✓	✓	✓	✓
Father controls	✓	✓	✓	✓	✓	✓	✓
Own controls	✓	✓	✓	✓	✓	✓	✓

Note: Regiment type fixed effects are for infantry, cavalry, artillery, and specialized units such as sharpshooters. Geographic fixed effects are for the residence of fathers. Enlistment rank fixed effects include dummies for the initial rank (85% of soldiers start as privates). Father controls include information from the 1860 Census on father's age, nationality, literacy, wealth (house value), and occupational skill group. A son's own controls include his age, age squared, and literacy in 1900. Standard errors clustered by family in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Chapter 3

WWI Anti-German Sentiment and Economic Growth in U.S. Counties

3.1 Introduction

Discrimination against a certain group can have significant effects on economic growth through reduced innovation (Cook, 2014) or forgone wages due to lost productivity (Cavalcanti and Tavares, 2016) or by pying a cost to avoid members of the other group (Heedegaard and Tyran, 2018). While previous work has considered systemic factors that determine these types of discrimination, such as racial discrimination in the U.S. which is deeply rooted in cultural norms, these same factors can potentially affect economic growth in the long-run directly. What is less well known is whether temporary changes in taste-based discrimination can impact long-term economic outcomes. To study this question, we consider the case of Germans and German-Americans in the United States during and after World War I. Germans, a large, economically successful, and socially respected immigrant group in the U.S. in the early 20th century, were subjected to strong anti-group sentiment during the war years.

We show that anti-German discrimination had negative effects on economic growth in discriminating counties in the post-war decades until the 1940s. The particular channel we identify is the outmigration of Germans from counties with the highest discriminatory pressure. The latter is measured by exploiting variation in the World War I casualty rate across counties in a reduced form difference-in-differences setting. With newspapers reporting daily from the front lines, communities that experienced losses were significantly more likely to retaliate against members of the nation that was responsible for these deaths. We digitize new military data for the Great War for the number of drafted, enlisted, and fallen soldiers and combine these with county level data from several Census years. Our results show that counties in the top quintile of the casualty rate distribution lost on average 20% of their German-born population while the bottom quintile received Germans by approximately the same amount. Migration responses to the casualty shock were particularly pronounced in areas where Germans used to be a salient immigrant group, mainly in the Midwest. Preferred destinations for German cross-state migrants included the South and the New England area.

Using digitized newspapers from Chronicling America together with linked individual Census data from the 1910 and 1920 full-count Census files, we provide further evidence for the relation between discrimination and the casualty rate shock. We measure anti-German sentiment as the share of newspaper articles on Germans which refer to them as *Huns*, i.e. the derogatory term used for Germans during the war. The use of this term only began to spike after the U.S. ultimately entered the war, which is when anti-German sentiment reached its peak. The rationals for the instrument is that deaths abroad in the war against the German Empire should increase anti-German sentiment in U.S. counties without having a direct effect on the migration decision of Germans in the U.S.

Instrumenting our newspaper-based discrimination measure with the World War I casualty rate, we find a strong positive effect on the probability of migrating to another county from 1910 to 1920. These effects are strongest for German-born individuals and those who were born to German parents. Second-generation Germans on the other hand were less likely to migrate. These results suggest a link between the cost of assimilation and migration similar to the work by Fouka (2018a,b). Observable characteristics matter for the outmigration decision. Those with common German surnames and those who state German as their mother-tongue in the 1910 Census, which reflects in a German accent when speaking English, were significantly more likely to move in response to increased discriminatory pressure.

Why would the in- or outflow of Germans across U.S. counties have any effect on the local economies? According to the 1910 Census, only 15.7% of the labor force worked in manufacturing, a sector which accounted for 42% of national output (Kendrick, 1961). The Census also shows that 37.1% of male German-born labor force participants worked in operatives and craftsmen occupations, and 23.6% of them worked in manufacturing industries. This compares to 26.5% of non-Germans in operatives and craftsmen jobs, and 15.4% in manufacturing. Losing these workers can therefore have a substantial negative effect on the local economies, especially those relying on income generated by manufacturing.

Using the county-level data from the first part, difference-in-differences regressions show that counties that experienced a net-outflow of Germans from 1910-20 had lower manufacturing wages, production, fewer firms, and slightly smaller firms in manufacturing in the following decades. We condition on pre-war economic and population characteristics to compare similar counties. To relax the assumption that parallel trends and pre-war county characteristics suffice for identification, we further instrument the German outflow with an indicator for being in the top quintile of the casualty rate distribution. Losses in a community due to the war against the Germans abroad significantly affected anti-German sentiment but should not have directly impacted economic outcomes at the county level after shutting down labor market and population effects with the appropriate controls. Our 2SLS estimates confirm the OLS findings.

This paper is the first to quantify the direct effects of the war on anti-German sentiment and its effect on the relocation decisions of Germans and German Americans in the U.S. We contribute to a literature on the relationship between discrimination and economic growth (Cook, 2014; Cavalcanti and Tavares, 2016), and to work relating to the consequences of anti-German sentiment in the U.S. (Moser, 2012; Moser and Voena, 2012; Baten, Bianchi and Moser, 2017) and the unintended effects of such discrimination (Fouka, 2018a,b). Another related strand of the literature is that on the economics of forced migration which is surveyed in Becker and Ferrara (2019).[1] Relative to work on receiving economies, the effects of forced migration on sending economies is less well explored. The case of Germans in the U.S. is a special setting given that we can observe displaced individuals in the sending and receiving counties within the same country. Also we can exploit localized measures of anti-group sentiment leading to forced migratory responses from the affected group instead of relying on government mandated expulsion policies.

The paper is structured as follows. Section 3.2 reviews the historical context of Germans in the U.S. and the forms of discrimination that was targeted at them during World War I by the state and the wider public. We also review the previous scholarly work on this topic. Section 3.3 explores the relationship between the World War I casualty rate and German migratory movements within the U.S. In particular, it shows the dynamics and determinants of sending and receiving counties across the casualty rate distribution via difference-in-differences regressions. To show that the channel of forced migration is indeed the hypothesized discrimination argument, section 3.4 combines linked individual Census data for Germans and German-Americans with newspaper data that seeks to capture anti-German sentiment at the local level via OLS and 2SLS regressions where discrimination is instrumented with the WWI casualty rate. Section 3.5 returns to the county-level analysis and studies the effect of WWI-induced German outflows on counties' long-term economic performance with a focus on the

[1] Even though forced migration is typically associated with state-mandated expulsions or the direct effects of wars, relocation in response to discrimination is also a type of forced migration.

manufacturing sector, i.e. a small but highly productive sector in which Germans tended to be concentrated. The final section concludes.

3.2 Historical Background and Related Literature

Since the end of the colonial period until the middle of the 20th century, Germans have been among the largest immigrant groups in the United States. According to the 1900 Census, the share of German-born and second generation Germans among the total population in the U.S. was over 10%. Aside from the size of their group, Germans were also known for their economic successes. Higham (1998) cites a survey of businessmen from 1908, who ranked immigrant nationalities by traits. Respondents named Germans before the English and even attributed them with more positive traits than Americans in some respects. Abramitzky et al. (2014) compute earnings penalties for different nationalities based on occupations and show that Germans had the lowest penalty relative to Americans. The penalty disappeared entirely for those who had stayed in the country for more than 30 years.

Table 3.1 reports average socioeconomic characteristics of Americans and different immigrant groups from the Census of 1910.[2] While Germans were similar to Swedes and English with respect to literacy, urbanization, or naturalization, they stood out in terms of their occupational earnings and education scores which were closer to those of Americans'. When it comes to home or business ownership, they even outperformed Americans. Germans tended to speak English at a lower rate than Swedish and English immigrants, and only Italians had a much lower propensity to learn the language.

Germans were known for being hard working and economically successful, but also for tending to their language and customs. Cities with larger German populations even offered bilingual education (Fouka, 2018a). Other examples include the gymnastics (*Turnvereine*) and shooting (*Schützenvereine*) societies, newspapers publishing in the German language, or German churches (Lübke, 1974). German culture and their local communities left a permanent mark on the landscape of the U.S. Especially in the Midwest, a region with high rates of historic German settlement, this influence is still visible today in town names such as Berlin, Wisconsin, or Bremen, Indiana. Despite their social preferences, Germans and German-Americans were well respected as hard-working, rapidly assimilating, and patriotic members of society (Higham, 1998). For decades they managed a balance between their old and new home, which is reflected in the saying: "*Germania my mother, Columbia my bride*" (Lübke, 1974, p. 48)

This changed dramatically with the onset of World War I (1914-18). Germans and

[2] Americans here means U.S.-born individuals whose parents were also born in the United States.

German-Americans had seen animosities rising against them since the beginning of the war. The fact that some German churches and societies tried to raise funds for the German war effort, or the lobbying for the U.S. to remain neutral during the conflict, did not help these trends (Lübke, 1999). However, the peak of Anti-Germanism in the U.S. was reached after the country eventually entered the war in 1917. Figure 3.1 shows the share of newspaper articles including the words *enemy* or *Huns* among articles mentioning *Germans*. Especially the term Huns was meant to be derogatory. Usage of the word *enemy* in relation to Germans saw a surge after 1914 and then doubled again with the entry of the U.S. into the war. This is also when the share of articles about Germans using the word *Huns* spiked.[3]

Figure 3.2 displays different depictions of Germans in the American press and literature. Panel (a) shows the "German Hun" as rapist who is stopped by an American soldier as advertisement for war bonds, and panel (b) depicts a German soldier as goosestepping child murderer. This type of discrimination was not only targeted at Germans but also against naturalized German-Americans. Panel (c) displays a spy who, under the cover of citizenship, seeks to sabotage the U.S. war economy. Germans in the U.S. were frequently accused of spying for the Empire,[4] and were under constant surveillance even by para-official organizations such as the American Protective League (APL).[5] Germans were forced to buy war bonds, to kiss the U.S. flag, and to denounce the German Emperor (Lübke, 1974).

The level and extend of Anti-Germanism reached into all parts of life. Moser (2012) shows that the share of operas by German composers fell from 43% to less than 7%, that the use of Otto or Wilhelm as first names for newborn children declined dramatically, and that applicants to the NYSE with a German sounding surname saw a doubling in their rejection rates during the war years. Sauerkraut consumption fell by 75% from 1914 to 1918, and Hamburgers were renamed to *liberty steaks* (Fouka, 2018a). Aside from the economic and social discrimination, Germans also had to fear physical harm. Robert Prager was lynched on April 5, 1918, in Collinsville, Illinois, and beatings or taring and feathering were other more common forms of assault (Lübke, 1974). Several men involved in the Prager lynching subsequently faced trial in a court of law but none of them were convicted.

The state not only turned a blind eye but actively benefited from expropriating Germans. The Office of the Alien Property Custodian was established in October 1917 and tasked with expropriating German assets. Miller (1922) details the corporations, firms, trade-marks,

[3]Data for this graph were taken from Chronicling America, a source which is described in more detail in the later part of this paper.

[4]See also panel (b) of figure 3.7.

[5]The APL was founded in 1917 and at its peak had 250,000 members in over 600 cities who were looking out under cover for enemy activities or possible spies (Higham, 1998).

copyrights, and patents seized and sold under the office in relation to the Trading With the Enemy Act of October 6, 1917. The total value of these exceeded \$ 500 million in 1919, which corresponds to \$ 7.5bn in 2018 dollars using the CPI from the BLS inflation calculator.

The fate of Germans in the U.S. during World War I has recently attracted the interest of economists. Moser and Voena (2012) and Baten et al. (2017) study the effect of the almost 6,340 patents disowned from Germans and German companies by the Alien Property Custodian on domestic patenting and patenting behavior of German firms. Moser and Voena (2012) show that expropriation and reselling German chemical patents to domestic inventors raised domestic patenting by 20% relative to the average patenting activity between 1919 and 1939. Most of the gains happen through learning-by-doing effects which occur with a lag of eight to nine years. Interestingly, German firms responded by patenting more in affected fields after the war (Baten et al., 2017).

Fouka (2018a) studies how Germans reacted to the forced assimilation policies via the prohibition of teaching the German language in schools. She shows that such policies backfired and led affected children of German immigrants to volunteer less for military service during World War II, to marry within their own group, and to more frequently give typical German names to their children. However, if German parents had been in the U.S. for longer, they tried to increase assimilation efforts and gave their children English names (Fouka, 2018b). The assimilation response therefore seems to have depended significantly on the costs of assimilation.

According to Fouka (2018b), many Germans chose to anglicize their names and to petition for naturalization, especially in states with higher incidences of violence against Germans. Negative employment effects from having a German sounding surname have been found for job applicants at the NYSE during World War I (Moser, 2012). Biavaschi, Giulietti and Siddique (2017) provide evidence of general positive payoffs for name Americanization by migrants in the early 20th century. Aside from changing their names, some Germans sought to prove their loyalty to the U.S. by volunteering for military service after which they had an increased tendency to marry Americans and to naturalize, albeit at a lesser rate than other immigrant groups such as Italians or Eastern Europeans (Mazumder, 2018).

What did Germans and German-Americans do who refused to assimilate? This question has received less attention but visual inspection of patterns of German migration within the U.S. shows a clear trend. Panel (a) of figure 3.3 maps the change in the share of German population (net of total population changes) from 1910 to 1920 using county-level Census data. The largest outflows occurred in areas where Germans were a large immigrant and therefore salient group, mainly in the Midwest. Also the results found by Fouka (2018b) cannot be explained away by Germans leaving the country. The evidence points towards Germans relocating to areas with lower anti-German sentiment rather than leaving the U.S.

Moves due to discrimination or violence targeted at specific groups have received renewed attention in the literature on forced migration which has particularly strongly evolved in the last ten years and which is survey in Becker and Ferrara (2019). Several examples show how the loss or gain of forcefully removed groups can have lasting effects on the sending or receiving economies. The arrival of expelled Huguenots from France in Prussia in 1685 brought knowledge and technology with them which subsequently raised productivity in the long-run (Hornung, 2014). Conversely, the expulsion of 3 million Germans from the Czech borderlands after World War II negatively affected economic growth in these areas due to a lack of agglomeration economies and the erosion of property rights (Testa, 2018). Pascali (2016) shows that Italian municipalities that expelled their Jewish population during the 15th and 16th century have lower incomes today and a less developed banking system.

3.3 WWI Casualties and German Outmigration

What led to the change in location patterns for Germans and German-Americans from 1910 to 1920 as shown in panel (a) of figure 3.3? To explore this, we collected data on deaths during World War I to compute county-level mortality rates among soldiers.[6] Panel (b) of the same figure plots the spatial distribution of World War I casualty rates across counties. With the exception of the Western parts of the country, the share of deaths sustained during the conflict correlates with changes in the German population share. On average, counties with lower casualty rates saw an inflow while those with higher casualty rates saw an outflow of Germans.

Newspapers reported daily on those who had given their lives on the battlefields of Europe. Back at the home front, people were particularly aware of losses in their own community. Local newspapers would highlight such casualties as shown in panel (a) of figure 3.7. Losses in a community would make the war even more salient and fuel the already high sentiment against Germans and German-Americans, i.e. the people belonging to the country that was responsible for these deaths. The two potential responses of the local German population was then to either increase their assimilation efforts as argued by Fouka (2018b), or to relocate to other counties where discriminatory pressure was lower.

Figure 3.4 plots the evolution of the share of German population in counties with a below or above median WWI casualty rate in panel (a). While evolving in a parallel fashion

[6] The data sources are described in more detail in the data appendix.

before the war, the share of Germans markedly dropped in counties with above median casualty rates from 1910 to 1920. Panel (b) below plots the relationship between the change in the share of Germans from 1910-20 and the WWI county-level casualty rate. The fitted regression line shows a negative relation between the two quantities.

To formalize the analysis, we combine our casualty rate measure with county-level data from the U.S. decennial Census from 1870 to 1940 and estimate the following regression:

$$\begin{aligned} \text{\% German pop.}_{ct} = & \sum_{q=1, q\neq 3}^{5} \tau_q Q_q(\text{WWI Casualty rate})_c \times \text{Post-WWI}_t \\ & + \alpha_c + \lambda_t + X'_{ct}\gamma + \epsilon_{ct} \end{aligned} \tag{3.1}$$

where the outcome is the share of the German-born population of a county's total population. $Q_q(\cdot)$ the q^{th} quintile of the World War I casualty rate distribution which is interacted with a post-war indicator that equals one from 1910 and is zero otherwise. Quintile three is omitted and acts as the baseline for comparison.

The rationale for letting the casualty treatment effect vary by treatment intensity is motivated by panel (b) in figure 3.4. Lower casualty rate values are associated with an inflow of Germans whereas the opposite holds for higher casualty rates. Regressing eq. (3.1) using the casualty rate interacted with a post-war indicator would therefore average over a positive and a negative effect which would cancel each other. If local casualties led to anti-German sentiment, and if Germans sought to evade such animosity, then those counties with lower casualty rates should receive Germans and those with higher losses should lose them, i.e. τ_q should be decreasing in q.

Time invariant county characteristics are captured by county fixed effects α_c and aggregate shocks common to all counties are absorbed by time fixed effects λ_t. Controls in X'_{ct} include pre-war county characteristics interacted with the post-war indicator: the average share of German population, the World War I draft rate, population, male-to-female ratio, share of manufacturing employment, and the share of urban population. In additional specifications we also include linear county-specific time trends $\alpha_c t$, or quadratic time trends $\alpha_c t^2$. Inclusion of such trend terms relaxes the parallel trends assumption and probes for robustness of our findings with respect to underlying differential secular trends in the outcome across high and low casualty rate counties that might be driving the results. All unexplained variation remains in the error term ϵ_{ct}. To account for heteroscedasticity and autocorrelation we cluster standard errors at the county level.

The identifying assumptions are that the share of German-born population evolved in a parallel way across low- and high-casualty rate counties and that there are no unobserved time-varying factors that confound the relationship between WWI casualties and the German population share. The unconditional parallel trends plots for above and below median casualty counties are shown in 3.4 panel (a). Figure 3.5 shows the coefficient plots of eq. (3.1) with the treatment quintile indicators being interacted with decade fixed effects rather than the post-war dummy.[7] Error bars are 95% confidence intervals. Both plots do not show any significant differences in pre-trends relative to the third quintile. After the war there is a divergence in the share of German-born population with an increase in counties with the lowest casualty rates and a decrease for counties in the top quintile of the casualty distribution.

The results from estimating eq. (3.1) are reported in table 3.2. The τ_q coefficients are decreasing in q with counties in the bottom quintile experiencing an inflow of Germans which reduces as we move up the quintiles and eventually turns negative from quintile four on. Being in the top quintile of the casualty rate distribution is associated with an outflow of Germans of 0.32 to 0.35 percentage points. Relative to the average share of Germans of 1.681% in 1910, this is a decrease of 19 to 21% in the German-born county population. In levels this corresponds to an average loss of 166 to 183 Germans. Compared to the total population this is a small number. However, as stated earlier, Germans tended to be particularly overrepresented in manufacturing. With an average of 1,324 manufacturing workers per county in 1910, the outflow of a group specialized in this industry can have the potential to impact production and the latter part of this paper will provide evidence for this.

The largest and most significant effects are found in the tails of the casualty rate distribution where the estimated τ_1 and τ_5 coefficients are always at least significant at the 5% level. The top and bottom quintile effects are robust to the inclusion of county-level controls as well as to different types of county-specific time trends. Even though the predicted inflow to counties with the lowest casualty rates is slightly larger than the predicted outflow from high-casualty counties, the difference between τ_1 and τ_5 is not statistically significant. Conditional on the pre-war German share, this implies that outflows in one part of the country were approximately equal to inflows in another. Germans therefore appear to have relocated within the U.S. rather than having left the country entirely. This argument is consistent with robustness checks by Fouka (2018b), who finds that her results are also not driven by Germans exiting the U.S.

[7] The figure plots coefficients for quintiles 1 and 5 only for better visibility. Plots with all quintiles are reported in the appendix in figure 3.A.1.

Where did the Germans go and where did they leave from? To answer this question, we modify the previous regression and estimate,

$$\begin{aligned}\%\text{ German pop.}_{ct} = \beta_s \left[(\text{WWI Casualty rate})_c \times \text{Post-WWI}_t \times I(\text{State} = s)\right] \\ + \alpha_c + \lambda_t + X'_{ct}\gamma + \epsilon_{ct} \qquad (3.2)\end{aligned}$$

where we interact the WWI casualty rate with a post-war indicator and with state dummies. $I(\cdot)$ denotes the indicator function. By letting the casualty rate effect vary by states, this gives a geographic approximation as to where differing WWI mortality rates were associated with an increase or decrease in the local German population.[8] We plot the β_s as map in figure 3.6, which provides a convenient way to visualize the results. Light yellow shades mark outflows of Germans and dark orange and red shades mark inflows.

The outflows are strongest in areas with larger pre-war shares of Germans and higher WWI casualty rates (cf. figure 3.3, panel b). German outflows mainly occur in the Midwest and to a lesser extend in the Great Plains regions where they are a larger and therefore more salient group. Interestingly, Southern states are the largest receivers of Germans as well as the New England region.

3.4 Anti-German Sentiment as Channel of Forced Migration

The previous section established a relationship between war casualties and the relocation of Germans across the country. To test whether the mechanism behind this relation was indeed an increase in anti-German sentiment, we now turn to individual level data from the 1910 and 1920 full-count Census files and digitized newspaper data. In particular, we analyze 1.9 million newspaper articles published across the country between 1914-18 to obtain a measure of anti-German sentiment. Our main measure of discrimination at the county level is computed as the share of articles about Germans that mention the derogatory term *Huns* during the war years, which is plotted over time as country aggregate in figure 3.1. The discrimination measure is computed as

$$\text{Discrimination}_c = \frac{\sum_t \sum_a I(\text{Huns=1})_{act}}{\sum_t \sum_a I(\text{German=1})_{act}} \times 100 \qquad (3.3)$$

where a indexes newspaper articles, t years, and c counties where the article was published. If a county did not have its own newspaper outlet, we assigned the value of the nearest county with such an outlet and applied a linear spatial distance weight as first-order approximation to newspaper circulation.

[8] It is possible to do the same exercise at the county-level which involves estimation of more parameters, hence the state-level was more convenient both for visualization and estimation purposes.

The spatial distribution of our anti-German sentiment measure is plotted in panel (c) of figure 3.7. The discrimination measure is then combined with individual Census records which we use to track Germans and German-Americans and their location decisions in the U.S. from 1910 to 1920. To link individuals across Census years, we follow the methodology introduced by Ferrie (1996) with the exception of not restricting names to those that occur less than 10 times in the entire country. Instead, we match male individuals on their full name, place of birth, as well as mother's and father's place of birth. The 1910 sample was restricted to individuals who were younger than 60 such that we could observe them in 1920 and limit the possibility that they might have died in between Census years. The restriction also allows us to observe labor market variables in 1920 which would not be available for older individuals who would have retired in the meantime.

We also match on year of birth which tolerates a deviation of plus or minus three years.[9] The fuzzy matching on year of birth is due to age heaping in the Census which is introduced by individuals' tendency to round numbers. All other matching variables are required to match strictly between Census records. In case of multiple matches, the pairing with the smallest birth year difference was kept. If a tie could not be uniquely resolved in this way, we dropped the corresponding observations. The final sample consists of almost 586,000 linked individuals.

An individual was marked as mover if he resided in different counties in 1910 and 1920. An individual might be observed in a different county in 1920 because he actually moved, or because the record linkage algorithm picked up a different person. In case of such linkage errors, our estimates would be biased towards zero due to misclassification in a binary outcome variable (Meyer and Mittag, 2017). We later provide robustness checks to rule out that our results could be affected by such errors.

Also note that linking German individuals this way will only capture those who chose to not fully assimilate. Those under the highest pressure or with the lowest cost of assimilating would anglicize their names (Lübke, 1974). Consequently our linking procedure will miss such individuals. However, our focus here is on studying the response to discrimination by those who did not want to fully assimilate by Americanizing their names. To provide plausibly exogenous variation in the level of discrimination faced by a given worker, we instrument the discrimination measure with the WWI casualty rate from their county county of residence in 1910. The idea is that deaths abroad in Europe affect anti-German sentiment, but otherwise would not directly impact the relocation decision of Germans and German-Americans in the U.S.

[9] The focus on men is because women tend to change their surnames upon marriage and therefore are much harder to accurately link across Census years without additional information.

To estimate the effect of anti-German sentiment, as captured by our newspaper-based discrimination measure, we regress

$$\begin{aligned}\Pr(\text{mover}=1)_{ict} = &\ \theta_1 \text{Discrimination}_c \times \text{Post-WWI}_t \\ &+ \theta_2 \text{Discrimination}_c \times I(\text{Both parents German})_i \times \text{Post-WWI}_t \\ &+ \theta_3 \text{Discrimination}_c \times I(\text{German-born})_i \times \text{Post-WWI}_t \\ &+ \phi_i + \pi_t + X'_{ic,1910}\psi + \nu_{ict} \end{aligned} \tag{3.4}$$

where individuals are indexed by i, counties by c, and the two Census years 1910 and 1920 as t. The coefficients of interest are θ_2 and θ_3, which measure the effect of an increase in anti-German sentiment on the location decisions of German-Americans whose parents were both born in Germany, and German-born individuals.The comparison is relative to individuals of German descent who only have one German-born parent or German-born grandparents and for whom the anti-German sentiment effect would be captured by θ_1.[10] If German-born individuals faced higher costs to evade discrimination compared to those who had German parents but were born in the U.S., then we would expect that $\theta_3 > \theta_2$ in absolute terms.

To alleviate potential endogeneity concerns, we include a broad range of pre-treatment individual characteristics in $X'_{ic,1910}$ which are interacted with the post-war indicator. These are measured in 1910 and include indicators for an individual's urban, farm, employment, and marital status, their skill group, literacy, family size, school attendance, labor force participation, years in the United States, and weeks of unemployment. Skill groups are assigned to the 9 groups available in the 1950 occupational definition of the U.S. Census Bureau. This is to exclude the possibility that individuals moved for other reasons that spuriously correlate with levels of discrimination. For instance, semi-skilled workers who moved to industrial centers to satisfy the labor demands of the war economy might have subsequently faced lower levels of discrimination in these urban areas due to higher tolerance as compared to rural areas.

In additional specifications we include state-specific time trends by interacting state fixed effects with the post-war indicator, as well as group-specific trends which interact the post-war indicator with the dummies for whether both parents were German (if the individual was born in the U.S.) and for whether the individual was born in Germany.[11] Such trend terms are supposed to pick up unobserved secular trends in discrimination which might have existed

[10]In terms of migration responses to discrimination, individuals with only one German-born parent were statistically indistinguishable from those with German-born grandparents only. Hence both groups were included in the comparison group.

[11]Unlike the county-level regressions, we cannot include county-specific linear time trends as this would require a minimum of three time periods.

in areas with higher shares of German populations, or secular trends in the propensity to migrate among German- versus native-born individuals. Time-invariant individual characteristics are captured by ϕ_i and unobserved aggregate time effects by π_t. Standard errors are clustered at the county-level in 1910.

Table 3.4 reports the results for the OLS estimation of eq. (3.4) in panel (a), and the reduced form regressions which replace the discrimination measure with the WWI casualty rate in panel (b). A one percentage points increase in the newspaper-based discrimination measure is associated with a 1 to 2 percentage points increase in the probability of German-born individuals to relocate between 1910 and 1920.[12] The effect size gets reduced by almost one half after including controls and state or group-specific time trends relative to the baseline specification in column 1. Neither controls nor trend terms explain away the finding of an increased propensity to change county from individuals born in German in response to an increase in the discriminatory pressure faced during the war years. This effect is significant at the five percent level in all specifications.

The reduced form coefficients appear to be much larger, however, this is mainly due a difference in scaling. As shown in the summary statistics in table 3.3, the average WWI casualty rate was 0.18%.[13] Hence a one unit increase would correspond to more than five times the mean. At average casualty rates, the coefficient in column (1) of panel (b) for German-born individuals would be 18.5 percentage points. The reduced form coefficients are less sensitive to the controls and time trends in the other specifications.

Unlike in the OLS regressions using the newspaper-based discrimination measure, the effect is negative for individuals with German grandparents. At the average casualty rate, these individuals would be 7.5 to 12.3 percentage points less likely to migrate. Given that these individuals were born in the U.S. with at least one U.S.-born parent, they likely increased their integration efforts instead of opting into relocation as in Fouka (2018b). Again, effects survive the inclusion of baseline controls and time trends, and remain significant at the one percent level in all specifications.

What characteristics drove the cross-county migration responses by Germans and those born to German parents? To answer this, we interact the discrimination and casualty measures with observable characteristics. This includes whether an individual reported German to be their mother tongue in the Census, meaning that they were more likely to have a distinct German accent when speaking English, or whether their surname was among the top 30 most

[12]The average share of anti-German articles was 1.42% with a standard deviation of 4.3.

[13]This is slightly higher than the country-wide casualty rate of 0.13% Roberts and Burda (2018).

common German surnames.[14] Observable characteristics, such as physical traits or surnames, have been shown to have significant impacts on labor market outcomes (Hamermesh and Biddle, 1994; Biavaschi et al., 2017).

Table 3.5 reports the results from this heterogeneity check. The baseline OLS and reduced form results are again shown in columns 1 and 4, respectively. Columns 2 and 5 add the interactions of the discrimination variable with an indicator for German native speakers. This additional term soaks up the treatment effect in the OLS and lowers the reduced form results by almost one half. A substantial part of the migration response to discrimination or to increases in the WWI casualty rate is therefore driven by observable factors that those who discriminated against Germans could use to target their victims.

An even larger effect comes from the interaction with the indicator for one of the 30 most common German surnames, for which the coefficient is approximately twice as large as for the mother tongue interaction. Without knowledge of what signifies a German accent, surnames might have been a more reliable measure of whether a person belonged to the group towards which the discrimination was targeted. Even though names are oftentimes private information, the labor market consequences of such names which are known to employers have been shown to be quite substantial as in the paper by Moser (2012).

3.4.1 Instrumental Variables Results

What if unobserved factors were driving both the increase in discrimination against Germans as well as the relocation decisions of Germans and German-Americans? For example, an unobserved local economic downturn due to the war economy could have raised anti-German sentiment in areas with more Germans, while fewer employment opportunities would drive those Germans away, therefore creating a spurious relation between our discrimination measure and the outmigration of these groups. In this case, the coefficients θ_2 and θ_3 would be overestimated in eq. (3.4).

To deal with such concerns, we instrument the three discrimination $\times$ post-war interactions in eq. (3.4) with the county-level WWI casualty rate interacted with the post-war indicator (using the 1910 county of residence), and the corresponding group indicators for German-born individuals and for those who were born in the U.S. but whose parents were Germans. All other terms are as before. The first stage results are reported in table 3.6 using the same specifications as before by including controls, state-, and group-specific trends one by one. We

[14] The 30 most frequently occurring German surnames in the 1910 Census are: Schmidt, Meyer, Schultz, Wagner, Weber, Hoffmann, Schneider, Becker, Schröder, Müller, Wolf, Peters, Bauer, Fischer, Koch, Klein, Zimmermann, Krueger, Keller, Beck, Kramer, Mayer, Krause, Schwartz, Hahn, Schmitt, Hartmann, Lange, Schäfer, and Kaiser.

report the Kleibergen-Paap first stage F-statistic for the excluded instruments. The two-stage least squares regression results are reported in table 3.7. The identifying assumption now is that the WWI casualty rate is a significant shifter of local anti-German sentiment, but that it does not directly affect the relocation decision of Germans.

For a casualty-induced one percentage point increase in our newspaper-based discrimination measure, German-born individuals are between 23 and 27 percentage points more likely to migrate. This is more than 10 times larger than the corresponding OLS results. A potential difference in the magnitude between OLS and 2SLS results is that 2SLS estimates a local average treatment effect, i.e. the migration response of individuals in places that saw a rise in anti-group sentiment because of local war casualties. Since tarring and feathering and other violent acts were potential repercussions, as discussed earlier, this local average treatment effect can be larger than the OLS result.

3.4.2 Sensitivity to Linkage Errors

A concern with relying on linked Census data is that potential linkage errors generate arbitrary moves across counties. We would then consider an individual to be a mover when in fact they are a stayer but simply were linked to a wrong person in 1920 who lives in a different county. This problem is prevalent in studies using linked Census data (Bailey et al., 2017). It is not sufficient to argue that such linkage errors would have to be uncorrelated with our discrimination measure and the casualty instrument. Since the migration indicator might be mis-classified, this would lead to an attenuation bias for OLS (Meyer and Mittag, 2017). For the instrumental variables regression, the case is less clear. The reduced form will be attenuated for the same reasons as the OLS regression, but if we incorrectly linked certain individuals across time, they will also potentially have received an incorrect discrimination value from the newspaper data. The overall bias then depends on whether the reduced form or the first stage are affected more from such linkage errors.

To probe for the sensitivity of our results with respect to such errors, we repeat the previous OLS, reduced form, and 2SLS regressions and include measures of linkage quality one by one and interact them with the post-war indicator. This includes the absolute birth year difference between an individual in 1910 and their linked counterpart in 1920,[15] as well as an indicator for whether a link was unique. This means that a person in 1910 was linked to a person in 1920 without any other competing potential match in 1920.

The results from this sensitivity check are reported in tables 3.A.2 and 3.A.3 in the appendix. Larger absolute age differences in linked individuals are positively and significantly

[15]See figure 3.A.2 for the distribution of birth year differences.

associated with being in a different county in 1920. This might reflect the type of linkage error discussed in this section. Likewise, being uniquely matched significantly reduces the probability of being in a different county in 1920. These measures of linkage quality are strong predictors of the outcome and drop the estimated discrimination and casualty rate coefficients by almost half. Nonetheless, the main results are not explained away in this exercise and particularly the 2SLS results are less affected compared to the OLS.

3.4.3 Determinants of Migration

Another question of interest is who the migrating Germans are. Their characteristics, skills, and knowledge will shed light on the mechanisms through which their out- or inflow can affect the local economies. We use the OLS and 2SLS estimates of equation (3.4) to predict the mover status of a given individual and generate a binary relocation prediction as

$$\text{Pr}(\text{predicted mover} = 1)_i = I(\text{Pr}(\widehat{\text{mover}} = 1)_{ict} > 0.5) \tag{3.5}$$

which we then use to compare the baseline 1910 characteristics of predicted movers and stayers via pairwise t-tests. These results are shown in figure 3.8 where panel (a) uses the OLS prediction and panel (b) uses the 2SLS prediction. Inference is adjusted for different group sizes. A positive coefficient means a higher average value of a given variable for movers.[16]

Individuals who are predicted to move are on average more likely to be employed in semi-skilled occupations, especially in the manufacturing sector, living in urban areas, married, and labor force participants. They are less likely to be high-skilled, farmers, home owners, proficient in the English language, and business owners. High costs to moving, such as owning a farm or a house, therefore reduce the probability of relocation. Most striking is the significantly higher probability of German movers to be employed in semi-skilled manufacturing jobs. With manufacturing accounting for 41% of natural output, but only 16% of employment in 1910 according to the Census, an outflow of workers in this particular sector can have potentially damaging effects to the local economy.

3.5 Economic Consequences of German Outflows

The outflow of Germans and German Americans from counties with higher discriminatory pressure during the war years might have had a negative influence on the growth of sectors in which these workers were overrepresented. Such an industry is the manufacturing sector for

[16] The full t-test tables are reported in table 3.A.1. The plots omit the variables age and year of immigration due to their large influence which would visually obscure differences in the other variables.

which we have detailed information on output and wages at the county level data. To test our hypothesis regarding the effect of German and German-American outflows on such economic outcomes, we use the county-level Census data from the first part of the paper. We now restrict this to the period from 1900 to 1940 to have a more stable sample and to reduce the issue of missing values in the economic variables before 1900.[17]

To test whether the German outflow from 1910 to 1920 had an effect on the economic outcomes under consideration, we regress

$$y_{ct} = \alpha_c + \lambda_t + \xi \left(\text{German Outflow from 1910-20}\right)_c \times \text{Post-WWI}_t + X'_{ct}\gamma + \epsilon_{ct} \qquad (3.6)$$

where α_c and λ_t are county and time fixed effects, as before, and the treatment variable German Outflow from 1910-20 measures the percentage points reduction in the population share of Germans in a county from before to after the war while controlling for total population changes. It is coded as zero if there was no negative change. The outcomes we consider are the log wage per worker in manufacturing, the log real value of manufacturing output as measure of productivity, the log value of material input, the number of manufacturing establishments, and the ratio of manufacturing workers to manufacturing establishments as measure of the average firm size. Monetary values are deflated to 1910 U.S. dollar values using the CPI from the Handbook of Labor Statistics and estimates published by the Federal Reserve Bank of Minneapolis.[18]

Controls are again the pre-war characteristics interacted with a post-war dummy including the average share of German population, the World War I draft rate, population, male-to-female ratio, share of manufacturing employment, and the share of urban population. Standard errors are clustered at the county level.

The identifying assumption is that $Cov(\text{German Outflow from 1910-20}, \epsilon_{ct}) = 0$, however, the outflow of German-born individuals from certain areas may be driven by unobserved characteristics as argued before, such as economic changes due to the war economy. To tackle such concerns, we instrument the German outflow variable with a dummy for whether the county was in the top quintile of the WWI casualty rate distribution. This choice is motivated by the results from table 3.2 and panel (b) of figure 3.5 which visualizes the first stage in event study form.

For the casualty rate to be a valid instrument, the population controls take an important

[17] To give an example regarding the stability of the sample, from 1870 and 1900 there were 516 new counties whereas from 1900 to 1940 the increase was only 148 new counties.

[18] The CPI series can be accessed online on the website of the Minneapolis Fed: https://www.minneapolisfed.org/community/financial-and-economic-education/cpi-calculator-information/consumer-price-index-1800

role of shutting down direct effects of the casualty rate on economic outcomes through their effect on the size of the labor force. Even though the WWI casualty shock was significantly smaller than those experienced during WWII or the Civil War, we still want to close this channel to ensure that the effect of the casualties impacted the economic outcomes through the outflow of Germans from the affected counties only. A challenge is that post-WWI values of the population controls might themselves be outcomes of the WWI casualty rate which would make them bad controls. In the main specifications, we therefore only control for pre-war characteristics but provide additional evidence in the appendix that also controlling for time varying measures of population size, male and female population, and workforce do not alter the results.

The OLS results from estimating eq. (3.6) are reported in table 3.8. Having a higher pre-war share of Germans is positively associated with most outcomes while firms tend to be slightly smaller. For an average of 2.424 percent of Germans before the war, this is not an economically significant difference. A one percentage points reduction in the share of German population in a county from 1910 to 1920 is associated with with a drop in the average manufacturing wage of approximately 1.5 percentage points. The average decrease in German population from before to after the war was 1 percentage point (which is approximately 392 Germans in absolute numbers) across counties that experience any decrease. While this is not a large drop in absolute terms, it is sizable for a relatively small outflow of population.[19] To give a rough approximation for the economic impact, consider the following back-of-the-envelope calculation. For the average annual earnings of \$ 6,957 in 1910 and an average employment of 1,317 manufacturing workers, the total wage loss amounts to \$ 137,435 per year. This is \$ 3,695,529 each year in 2018 values.

There also seems to be a small negative effect on log manufacturing output (productivity) but this is not statistically significant. Part of this can be explained by lower general production as proxied by the log of material inputs which drop by approximately 6 percentage points at the average German outflows. Affected counties also saw a reduction in the number of firms with an exit of 12 manufacturing establishments and a reduction in average firm size by 2 workers. The firm exists might partially be due to closure of smaller family-owned businesses by the relocating German population, although the individual analysis in the previous part suggests that firm owners are actually less likely to move.

The corresponding 2SLS results are reported in table 3.9 in which the German outflow variable is instrumented with the WWI casualty rate. The instrumental variables regression

[19]The average county in 1910 has 31,160 inhabitants, which compares to 392 exiting Germans on average between 1910 and 1920.

mirrors the OLS results in terms of the direction of the effects resulting from an outflow of German population between 1910 to 1920. In most cases, the 2SLS results tend to be larger in absolute value but also more imprecisely estimated despite a relatively strong first stage with a Kleibergen Paap F-statistic of 33.1. For instance, the negative wage effect in the 2SLS regression is 5.73 percentage points as compared to the approximately 1.5 percentage points drop found in the OLS. However, even nowadays people are willing to forgo 8 percent of their earnings to avoid working with individuals from another racial group Heedegaard and Tyran (2018), even the 2SLS effect size is not unreasonable. If OLS is biased, then the bias appears to be an attenuation bias.

3.6 Conclusion

With the onset of World War I, Germans in the U.S. faced tremendous discriminatory pressure. Unlike in many other cases, such as discrimination against blacks in the U.S. or Jews in medieval Italy, the discrimination was short-lived but led to significant movements of Germans across counties. In this paper we first argue that discriminatory pressure against Germans and German-Americans varied with the geographic distribution of war casualties sustained by the different counties. The resulting outflow tended to be strongest in areas with high casualty rates and where Germans had been a salient group to begin with, mainly the Midwestern states. The individual level analysis using a newspaper-based measured of anti-German sentiment with linked Census data has provided further evidence that the channel through which war casualties drove German migratory patterns was indeed an increase in discrimination. Observable characteristics that identified individuals as members of the German community were particularly important. The county-level data revealed that the outflow of Germans, a group that used to be concentrated in the manufacturing sector, i.e. a highly productive part of the economy, led to worse economic outcomes in the post-war decades. This translated into lower wages, productivity, firms, and firm sizes in the manufacturing sector. The forgone wage loss as approximately $ 3.7 million in 2018 values per year.

The empirical evidence supports the idea that counties which satisfied a short-run anti-group sentiment paid, probably unintentionally, a substantial price in terms of economic growth in the long-run. The negative economic effects persisted even after the initial reason for discriminating against Germans was long gone. This finding contributes to the literature on the unintended consequences of anti-German policies during World War I (Fouka, 2018a,b) and the effects of the war on Germans in the U.S. (Moser, 2012; Moser and Voena, 2012; Baten et al., 2017). The broader implications of the paper relate to work on the economic consequences of discrimination on economic growth (Cook, 2014; Cavalcanti and Tavares, 2016; Heedegaard and Tyran, 2018) through the channel of forced migration.

3.7 Tables

Table 3.1: Average Economic and Social Characteristics by Group in 1910

	Germans	Swedish	English	Italians	Americans
% urban	0.655	0.578	0.705	0.724	0.363
% farmers	0.209	0.232	0.102	0.027	0.377
% home owners	0.499	0.475	0.364	0.172	0.455
% naturalized	0.815	0.793	0.739	0.247	
% literate	0.957	0.976	0.987	0.677	0.911
% speak English	0.825	0.903	0.967	0.470	0.964
% business owner	0.147	0.116	0.082	0.038	0.144
Earnings score	111.404	94.190	108.083	87.319	115.557
Education score	77.154	59.783	68.500	49.871	92.013
Observations	1,198,372	347,935	404,200	820,743	17,474,027

Note: Average characteristics of immigrants and Americans (U.S.-born with both parents born in the United States) from the 1910 Census. Literacy refers to both reading and writing. Business owners refers to those whose employment status is assigned as *employer* in the Census. Occupational earnings scores were constructed from 1950 data to compute the median earnings of each occupation. The occupational education scores measures the percentage of individuals per occupation with one or more years of college education in 1950.

Table 3.2: County-Level Effect of WWI Casualties on German Migration, 1870-1940

	Outcome: Share of German Population (pre-war mean = 1.681)			
	(1)	(2)	(3)	(4)
1st Casualty Quintile × Post	0.416***	0.303**	0.371***	0.369***
	(0.129)	(0.121)	(0.077)	(0.077)
2nd Casualty Quintile × Post	0.020	0.036	0.190**	0.189**
	(0.150)	(0.137)	(0.084)	(0.084)
4th Casualty Quintile × Post	-0.175	-0.174	-0.063	-0.062
	(0.158)	(0.149)	(0.094)	(0.094)
5th Casualty Quintile × Post	-0.354**	-0.331**	-0.323***	-0.321***
	(0.164)	(0.157)	(0.116)	(0.116)
Observations	21,637	21,542	21,542	21,542
Counties	2,868	2,868	2,868	2,868
Adj. R^2	0.675	0.729	0.933	0.932
Controls		Yes	Yes	Yes
Linear county time trends			Yes	
Quadratic county time trends				Yes

Note: Difference-in-differences regressions of the share of Germans in county c in decade t, interacting quintiles of the WWI casualty rate with a post-WWI indicator. The comparison quintile is quintile three. The casualty rates in the first quintile are [0.001, 0.161), in the second quintile [0.161, 0.220), in the fourth quintile [0.278, 0.362), and in the top quintile [0.362, 2.911]. The pre-war outcome mean is measured in 1910. All regressions include county and decade fixed effects. Controls are pre-war county characteristics interacted with a post-war indicator and include the average share of German population before 1910, the World War I draft rate, population, male-to-female ratio, share of manufacturing employment, and the share of urban population. Standard errors are clustered at the county-level. Significance levels are denoted by * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 3.3: Summary Statistics for the Linked Census Sample of Germans and German-Americans

	Mean	St. Dev.	Min	Max
County-level variables				
Casualty Rate	0.183	0.116	0.005	1.396
Discrimination	1.421	4.302	0.000	49.057
Skill Group				
Professional/Technical	0.022	0.147	0.000	1.000
Farmer	0.163	0.369	0.000	1.000
Manager, Official, Proprietor	0.068	0.252	0.000	1.000
Clerical and Kindred	0.037	0.188	0.000	1.000
Sales worker	0.042	0.200	0.000	1.000
Craftsman	0.161	0.368	0.000	1.000
Operative	0.095	0.293	0.000	1.000
Service worker	0.029	0.167	0.000	1.000
Farm laborer	0.062	0.242	0.000	1.000
Laborer	0.078	0.268	0.000	1.000
Employment				
In the labor force	0.620	0.485	0.000	1.000
At work	0.729	0.444	0.000	1.000
Unemployed	0.026	0.160	0.000	1.000
Weeks unemployed	1.093	4.738	0.000	52.000
In school	0.169	0.375	0.000	1.000
Social				
Urban	0.597	0.490	0.000	1.000
Literate (reads and writes)	0.862	0.345	0.000	1.000
Married, spouse present	0.498	0.500	0.000	1.000
Married, spouse absent	0.012	0.110	0.000	1.000
Divorced	0.002	0.048	0.000	1.000
Widowed	0.017	0.128	0.000	1.000
Family size	5.116	2.618	1.000	55.000
In USA for 0 to 5 years	0.027	0.163	0.000	1.000
In USA for 6 to 10 years	0.021	0.144	0.000	1.000
In USA for 11 to 15 years	0.014	0.117	0.000	1.000
In USA for 16 to 20 years	0.042	0.201	0.000	1.000
Individuals	585,230			

Note: All individual level variables are measured in the pre-war Census year of 1910. Discrimination is measured as the share of newspaper articles mentioning Germans between 1914 and 1918 that include the word Hun or Huns. The WWI casualty rate is measured as the share of county-level deaths among the service eligible male population in 1910. Occupational classifications into skill groups follow the 1950 occupation definition of the U.S. Census Bureau. The sample consists of men in the 1910 and 1920 Censuses who were born in German or whose parents or grandparents were born in Germany.

Table 3.4: OLS and Reduced Form

	Outcome: Pr(moved county = 1) (mean=0.336)			
	Panel a: OLS			
	(1)	(2)	(3)	(4)
Discrimination × Post	-0.006**	-0.003	-0.003	0.005*
	(0.003)	(0.003)	(0.002)	(0.002)
Discrimination × Post × Both parents German	0.005**	0.004*	0.003*	0.007**
	(0.002)	(0.002)	(0.002)	(0.003)
Discrimination × Post × German-Born	0.019**	0.011**	0.009**	0.010**
	(0.008)	(0.005)	(0.004)	(0.005)
Controls		Yes	Yes	Yes
State-specific trends			Yes	Yes
Group-specific trends				Yes
Observations	1,152,620	1,152,620	1,152,620	1,152,620
Individuals	576,310	576,310	576,310	576,310
Adj. R^2	0.206	0.369	0.382	0.259
	Panel b: Reduced Form			
	(1)	(2)	(3)	(4)
Casualty Rate × Post	-0.528**	-0.415*	-0.518***	-0.685***
	(0.215)	(0.221)	(0.160)	(0.151)
Casualty Rate × Post × Both parents German	0.266***	0.236***	0.214***	0.507***
	(0.037)	(0.043)	(0.035)	(0.072)
Casualty Rate × Post × German-Born	1.025***	0.777***	0.748***	0.787***
	(0.171)	(0.168)	(0.163)	(0.164)
Controls		Yes	Yes	Yes
State-specific trends			Yes	Yes
Group-specific trends				Yes
Observations	1,170,314	1,170,314	1,170,314	1,170,314
Individuals	585,157	585,157	585,157	585,157
Adj. R^2	0.226	0.250	0.267	0.268

Note: The outcome is an indicator for whether an individual moved county between 1910 and 1920. Discrimination is the share of newspaper articles mentioning Germans and using the word Huns during the war years. Data are for individuals linked from the 1910 to 1920 full count Census files with German origin or ancestry, i.e. Germans, first-, and second-generation German-Americans. The category for both parents German refers American-born individuals only. Regressions include individual fixed effects and a 1920 indicator. Baseline controls are measured in 1910 and interacted with the 1920 indicator and include: urban status, skill group, farm ownership, employment status, literacy, marital status, years living in the U.S. (if German-born), family size, school attendance in 1910, labor force status, number of weeks unemployed. State-specific trends are state fixed effects interacted with the 1920 indicator. Group-specific time trends are nativity indicators for those with German parentage, or German-born interacted with the 1920 indicator. Standard errors are clustered at the county of residence in 1910. Significance levels are denoted by * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 3.5: Determinants of Anti-German Sentiment

	Outcome: Pr(moved county = 1) (mean=0.336)					
	(1)	(2)	(3)	(4)	(5)	(6)
Discrimination × Post	-0.003	-0.003	-0.004*			
	(0.002)	(0.002)	(0.002)			
Discrimination × Post × Both parents German	0.003*	0.003*	0.003*			
	(0.002)	(0.002)	(0.002)			
Discrimination × Post × German-Born	0.009**	0.003	0.002			
	(0.004)	(0.003)	(0.002)			
Discrimination × Post × German mother tongue		0.007**	0.007**			
		(0.003)	(0.003)			
Discrimination × Post × Common German surname			0.013***			
			(0.005)			
Casualty Rate × Post				-0.518***	-0.518***	-0.598***
				(0.160)	(0.160)	(0.160)
Casualty Rate × Post × Both parents German				0.214***	0.214***	0.207***
				(0.035)	(0.035)	(0.032)
Casualty Rate × Post × German-Born				0.748***	0.420***	0.431***
				(0.163)	(0.143)	(0.140)
Casualty Rate × Post × German mother tongue					0.332***	0.326***
					(0.113)	(0.110)
Casualty Rate × Post × Common German surname						0.843***
						(0.120)
Observations	1,152,620	1,152,620	1,152,620	1,170,314	1,170,314	1,170,314
Individuals	576,310	576,310	576,310	585,157	585,157	585,157
Adj. R^2	0.382	0.382	0.383	0.267	0.261	0.277

Note: The outcome is an indicator for whether an individual moved county between 1910 and 1920. Discrimination is the share of newspaper articles mentioning Germans and using the word Huns during the war years. Data are for individuals linked from the 1910 to 1920 full count Census files with German origin or ancestry, i.e. Germans, first-, and second-generation German-Americans. The category for both parents German refers American-born individuals only. Regressions include individual fixed effects and a 1920 indicator. The German mother tongue and common German surname variables are indicators for whether a respondent stated German to be their mother tongue and for those with one of the most common (top 30 most frequent) German surnames, respectively. Baseline controls are measured in 1910 and interacted with the 1920 indicator and include: urban status, skill group, farm ownership, employment status, literacy, marital status, years living in the U.S. (if German-born), family size, school attendance in 1910, labor force status, number of weeks unemployed. State-specific trends are state fixed effects interacted with the 1920 indicator. Group-specific time trends are nativity indicators for those with German parentage, or German-born interacted with the 1920 indicator. Standard errors are clustered at the county of residence in 1910. Significance levels are denoted by * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 3.6: Anti-German Sentiment and WWI Casualties First Stage Regressions

First stage one outcome: Discrimination$_c$ × Post-WWI$_t$				
	(1)	(2)	(3)	(4)
WWI Casualty rate$_c$ × Post-WWI$_t$	-4.334	-4.515	-6.511*	-6.464*
	(3.702)	(3.790)	(3.642)	(3.616)
WWI Casualty rate$_c$ × Post-WWI$_t$ ×	0.374	0.321	0.177	0.096
Both parents German	(0.032)	(0.363)	(0.249)	(0.267)
WWI Casualty rate$_c$ × Post-WWI$_t$ ×	0.604*	0.391	0.076	0.066
German-born	(0.350)	(0.603)	(0.309)	(0.309)
First stage two outcome: Discrimination$_c$ × I(Both parents German)$_i$ × Post-WWI$_t$				
	(1)	(2)	(3)	(4)
WWI Casualty rate$_c$ × Post-WWI$_t$	-4.551**	-5.112**	-6.178***	-1.291*
	(2.026)	(2.243)	(2.346)	(0.675)
WWI Casualty rate$_c$ × Post-WWI$_t$ ×	4.593***	4.702***	4.632***	-3.866***
Both parents German	(0.032)	(0.037)	(0.036)	(0.073)
WWI Casualty rate$_c$ × Post-WWI$_t$ ×	0.253***	1.381***	1.232***	0.113***
German-born	(0.023)	(0.040)	(0.040)	(0.032)
First stage three outcome: Discrimination$_c$ × I(German-born)$_i$ × Post-WWI$_t$				
	(1)	(2)	(3)	(4)
WWI Casualty rate$_c$ × Post-WWI$_t$	-2.899**	-2.132**	-2.806***	-5.862***
	(1.226)	(0.891)	(0.931)	(2.181)
WWI Casualty rate$_c$ × Post-WWI$_t$ ×	0.156**	-0.245***	-0.316***	4.999***
Both parents German	(0.013)	(0.012)	(0.015)	(0.067)
WWI Casualty rate$_c$ × Post-WWI$_t$ ×	4.723***	2.698***	2.560***	3.260***
German-born	(0.039)	(0.050)	(0.049)	(0.053)
Controls		Yes	Yes	Yes
State-specific trends			Yes	Yes
Group-specific trends				Yes
Observations	1,152,474	1,152,474	1,152,474	1,152,474
Individuals	576,237	576,237	576,237	576,237
K-P F-stat	16.51	15.11	12.84	9.86

Note: First stage regressions for the relation between anti-German sentiment and WWI casualties sustained in each county. Discrimination is the share of newspaper articles mentioning Germans and using the word Huns during the war years. The individual level data are for respondents linked from the 1910 to 1920 full count Census files with German origin or ancestry, i.e. Germans, first-, and second-generation German-Americans. The category for both parents German refers American-born individuals only. Regressions include individual fixed effects and a 1920 indicator. Baseline controls are measured in 1910 and interacted with the 1920 indicator and include: urban status, skill group, farm ownership, employment status, literacy, marital status, years living in the U.S. (if German-born), family size, school attendance in 1910, labor force status, number of weeks unemployed. State-specific trends are state fixed effects interacted with the 1920 indicator. Group-specific time trends are nativity indicators for those with German parentage, or German-born interacted with the 1920 indicator. Standard errors are clustered at the county of residence in 1910. Significance levels are denoted by * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 3.7: Anti-German Sentiment and Migration Decisions - 2SLS Results

	Outcome: Pr(moved county = 1) (mean=0.336)			
	(1)	(2)	(3)	(4)
Discrimination × Post	-0.090	-0.117	-0.072	-0.143*
	(0.065)	(0.078)	(0.074)	(0.080)
Discrimination × Post ×	0.057***	0.072***	0.067***	0.174***
Both parents German	(0.002)	(0.002)	(0.002)	(0.004)
Discrimination × Post ×	0.226***	0.269***	0.263***	0.238***
German-born	(0.002)	(0.005)	(0.005)	(0.004)
Controls		Yes	Yes	Yes
State-specific trends			Yes	Yes
Group-specific trends				Yes
Observations	1,152,474	1,152,474	1,152,474	1,152,474
Individuals	576,237	576,237	576,237	576,237
K-P F-stat	16.51	15.11	12.84	9.86

Note: The outcome is an indicator for whether an individual moved county between 1910 and 1920. Discrimination is the share of newspaper articles mentioning Germans and using the word Huns during the war years and is instrumented with the WWI casualty rate in a given county interacted with a post-war and the corresponding nativity indicators. Data are for individuals linked from the 1910 to 1920 full count Census files with German origin or ancestry, i.e. Germans, first-, and second-generation German-Americans. The category for both parents German refers American-born individuals only. Regressions include individual fixed effects and a 1920 indicator. Baseline controls are measured in 1910 and interacted with the 1920 indicator and include: urban status, skill group, farm ownership, employment status, literacy, marital status, years living in the U.S. (if German-born), family size, school attendance in 1910, labor force status, number of weeks unemployed. State-specific trends are state fixed effects interacted with the 1920 indicator. Group-specific time trends are nativity indicators for those with German parentage, or German-born interacted with the 1920 indicator. Standard errors are clustered at the county of residence in 1910. Significance levels are denoted by * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 3.8: Economic Effects of WWI-Induced Outflows of Germans - OLS Regression

	(1) ln(wages)	(2) ln(productivity)	(3) ln(input value)	(4) No. of firms	(5) firm size
Post-war German Outflow	-0.015***	-0.002	-0.064**	-12.740***	-2.138***
	(0.004)	(0.018)	(0.025)	(2.016)	(0.501)
Pre-war German share	0.009***	0.003	0.112***	7.263***	-0.253***
	(0.001)	(0.002)	(0.010)	(0.874)	(0.061)
Observations	10,474	10,474	10,474	10,474	10,474
Counties	2,258	2,258	2,258	2,258	2,258
Adj. R^2	0.876	0.892	0.971	0.605	0.645

Note: Regressions of economic county-level outcomes from 1900 to 1940 on the outflow of German population after WWI net of total county population changes. Log wages are per worker, productivity is the log real value of manufacturing output, input value is the log of value of materials used for production, the number of firms refers to the total number of manufacturing establishments, and firm size measures the average number of workers per manufacturing establishment. Controls are pre-war county characteristics interacted with a post-war indicator and include the average share of German population before 1910, the World War I draft rate, population, male-to-female ratio, share of manufacturing employment, and the share of urban population. Monetary values are deflated to 1910 U.S. Dollars. Standard errors are clustered at the county-level. Significance levels are denoted by * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

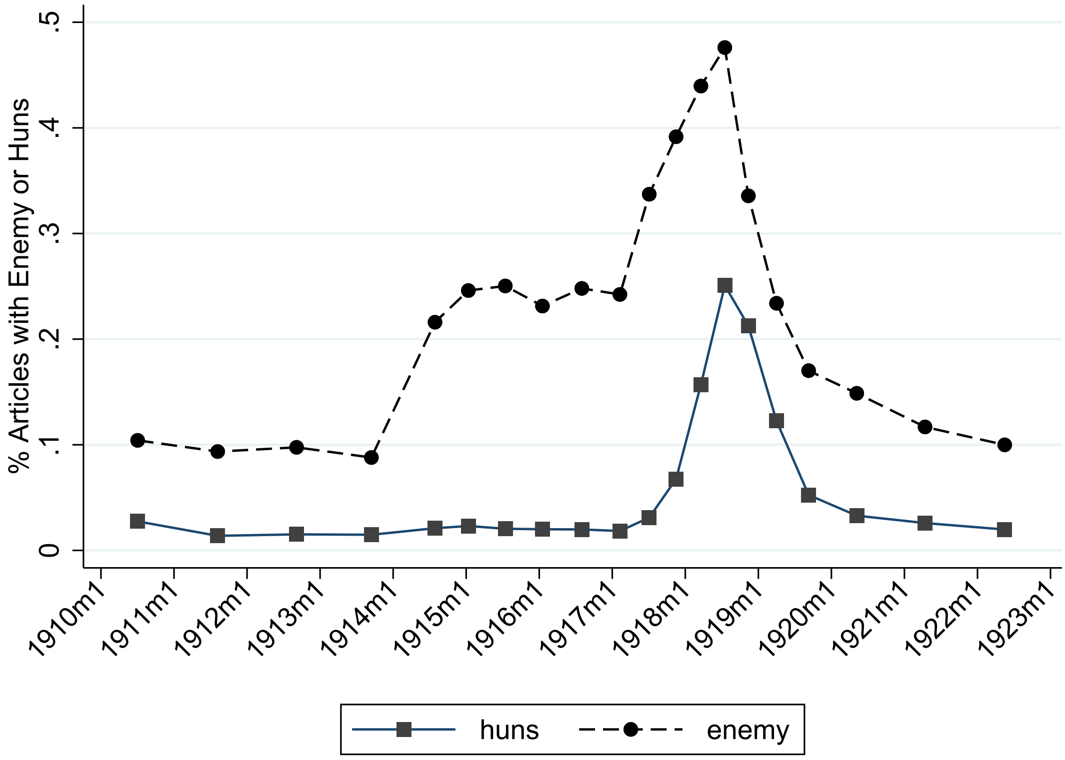
Table 3.9: Economic Effects of WWI-Induced Outflows of Germans - 2SLS Regression

	(1) ln(wages)	(2) ln(productivity)	(3) ln(input value)	(4) No. of firms	(5) firm size
Post-war German Outflow	-0.059*	-0.101	-0.225	-2.195	-2.260
	(0.034)	(0.164)	(0.225)	(8.750)	(4.484)
Pre-war German share	0.015***	0.018	0.134***	5.657***	-0.234
	(0.005)	(0.024)	(0.035)	(1.560)	(0.683)
Observations	10,367	10,367	10,367	10,367	10,367
Counties	2,230	2,230	2,230	2,230	2,230
K-P F-stat	33.060	33.060	33.060	33.060	33.060

Note: Regressions of economic county-level outcomes from 1900 to 1940 on the outflow of German population after WWI net of total county population changes. The outflow of German population after WWI is instrumented with the county-level WWI casualty rate. Log wages are per worker, productivity is the log real value of manufacturing output, input value is the log of value of materials used for production, the number of firms refers to the total number of manufacturing establishments, and firm size measures the average number of workers per manufacturing establishment. Controls are pre-war county characteristics interacted with a post-war indicator and include the average share of German population before 1910, the World War I draft rate, population, male-to-female ratio, share of manufacturing employment, and the share of urban population. Monetary values are deflated to 1910 U.S. Dollars. Standard errors are clustered at the county-level. Significance levels are denoted by * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

3.8 Figures

Figure 3.1: Share of Newspaper Articles on Germans Mentioning the Words *Enemy* or *Huns*

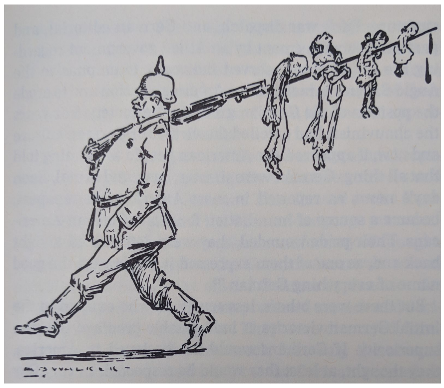


Note: Binned scatter lines of newspaper articles per month and year mentioning Germans together with the words *enemy* or *huns*. The density of dots represents the frequency of publishing in a given time interval. The figure shows how Germans are referred to as *enemy* (enemy, enemies, foe) from the start of the war, however, this spikes together with the use of the derogatory word *huns* once the U.S. enter the war in the first half of 1917.

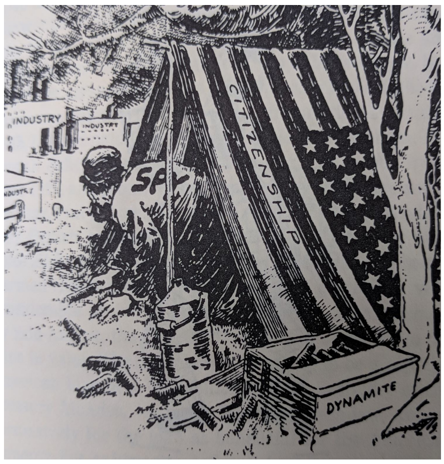
Figure 3.2: Anti-German Posters and Prints during WWI

(a) War Bond Advertisement

(b) German Soldier Shown as Child Murderer



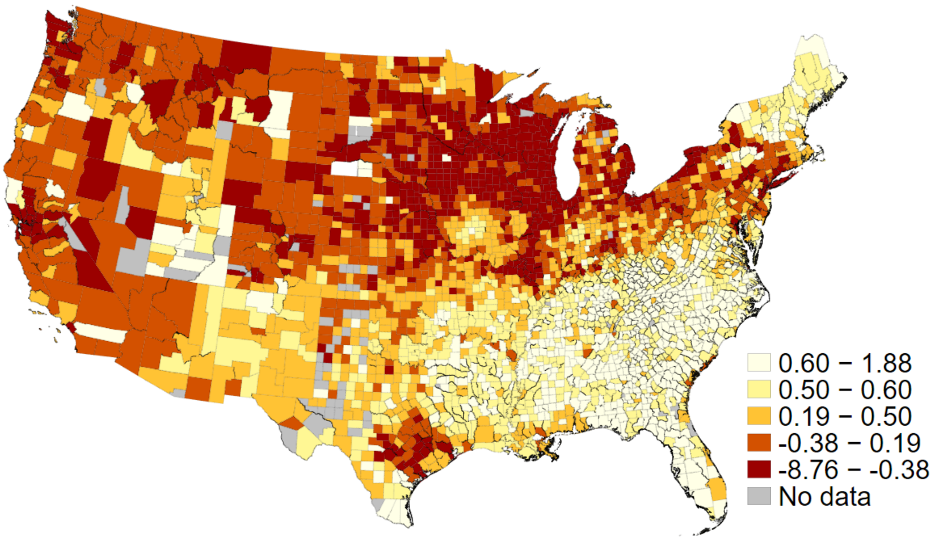
(c) Naturalized Germans Depicted as Spies



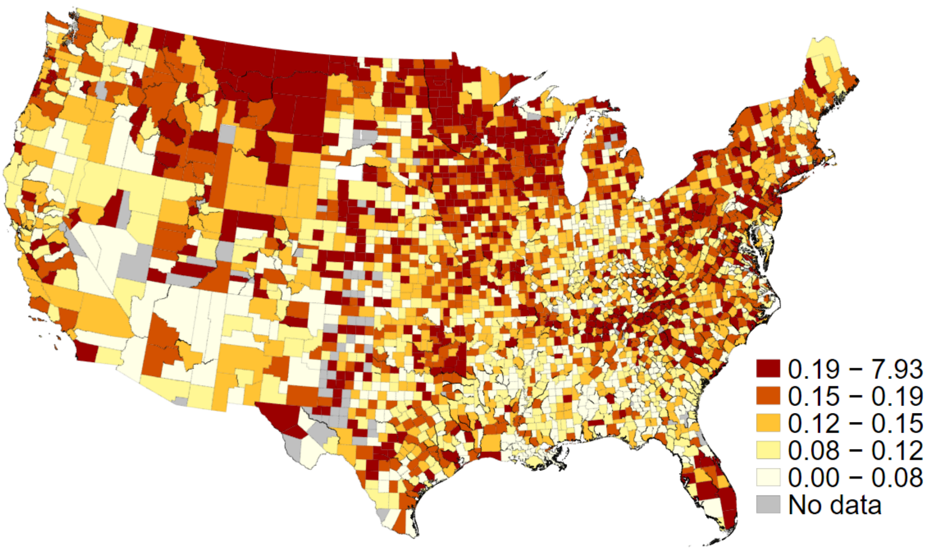
Note: Examples of Anti-German propaganda during WWI. Panel a) shows a German soldier as rapist who is stopped by an American soldier to promote the purchase of war bonds. Panel b) depicts a German soldier as child murderer. Panel c) discredits German-Americans who allegedly hide under the cover of U.S. citizenship to act as spies who target the U.S. industry shown in the background which they sabotage with the dynamite in the front of the image. Image source: Lübke (1974) pages 272 for image a), 87 for image b), and 147 for image c).

Figure 3.3: Spatial Distribution of German Population Flows and WWI Casualty Rate

(a) Change % German-born 1910-20

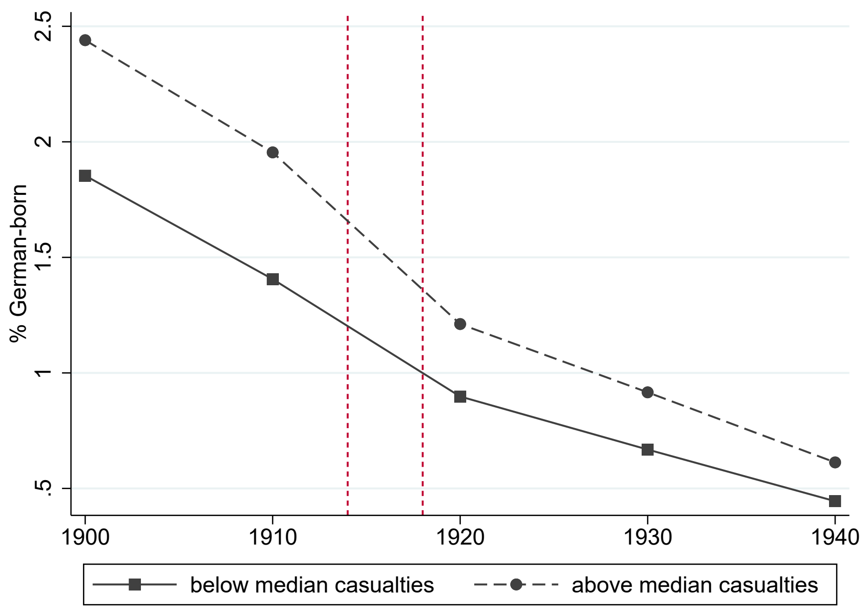


(b) WWI Casualty Rate

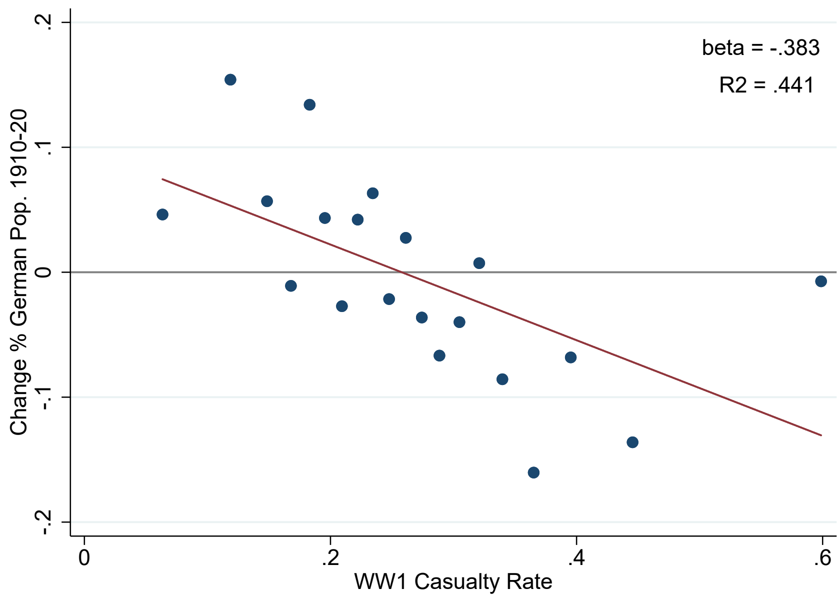


Note: Panel (a) maps the quintiles the change in the county-level share of the German-born population from 1910 to 1920. Total population changes have been partialled out to avoid confounding changes in the share by an influx of other immigrant groups, for instance. Panel (b) maps the county-level WWI casualty rate which is defined as the total number of WWI deaths over the male population of service eligible age in 1910, i.e. men aged 14-40, times one hundred.

Figure 3.4: Change in the Share of Germans over WWI Casualty Rates



(a) % German Population in Above/Below Median WWI Casualty Rate Counties



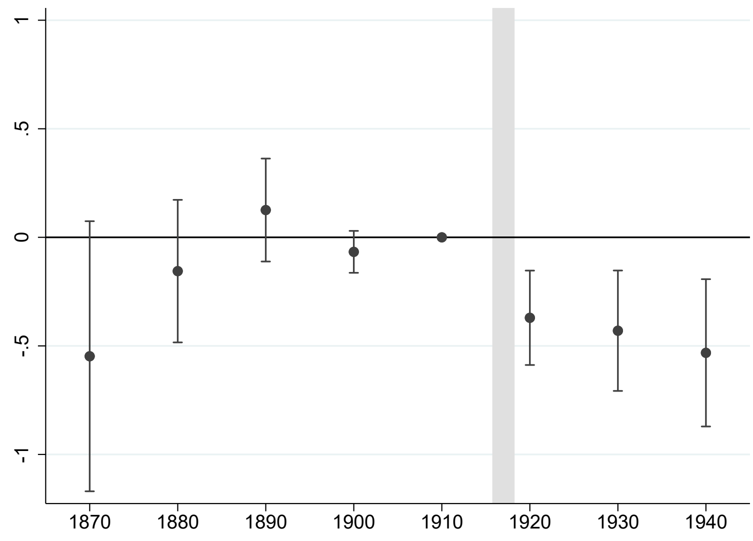
(b) Casualty Rates and Change in German Population Shares, 1910-20

Note: Panel (a) plots the evolution of the share of German population in above- and below-median WWI casualty rate counties over time. The war years are marked by the dashed vertical lines. Panel (b) shows the binned scatter plot in the change in the share of German population (net of total population changes) from 1910 to 1920 over the WWI casualty rate together with the corresponding regression line, as well as coefficient and R-squared values.

Figure 3.5: Effect of WWI Casualty Rates on the Share of Germans: Top and Bottom Casualty Rate Quintile

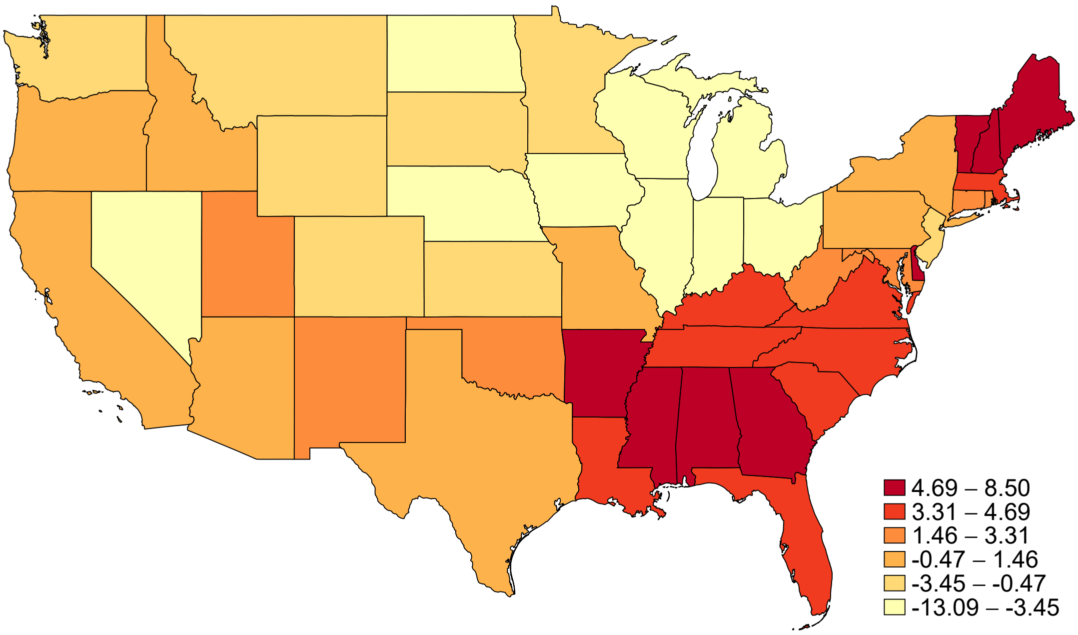
(a) Quintile 1

(b) Quintile 5



Note: Plotted coefficients from a regression of the county share of German population on WWI casualty rate quintiles interacted with time and county fixed effects. Panels (a) and (b) show the plots for the bottom and top quintiles of the casualty rate distribution, respectively. Error bars are 95% confidence intervals with standard errors being clustered at the county level. Controls include the WWI draft rate, as well as the average pre-war population, employment share in manufacturing, and male-to-female ratio interacted with time fixed effects.

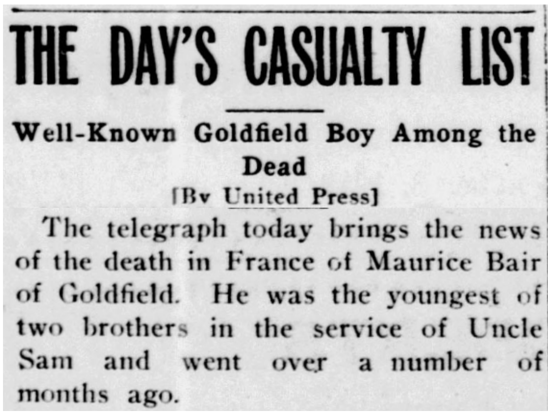
Figure 3.6: Treatment Effect Heterogeneity by State



Note: The map plots the coefficient values from a regression of the county population share of Germans on the WWI casualty rate interacted with state fixed effect and a post-war indicator for counties from 1870 to 1900. Each state corresponds to a given coefficient value for the given state-casualty term interaction. Coefficient values are divided into quintiles. Negative values (German outflows) are marked in yellow, and positive values (German inflows) are marked in red shading. The scale on the right provides the quintile values of the distribution of in- and outflows as estimated by the regression. This shows the migratory patterns of Germans across U.S. states following the WWI casualty shock.

Figure 3.7: Newspaper Articles Examples and Spatial Distribution of Anti-German Reporting

(a) Casualty Reports



THE DAY'S CASUALTY LIST

Well-Known Goldfield Boy Among the Dead

[By United Press]

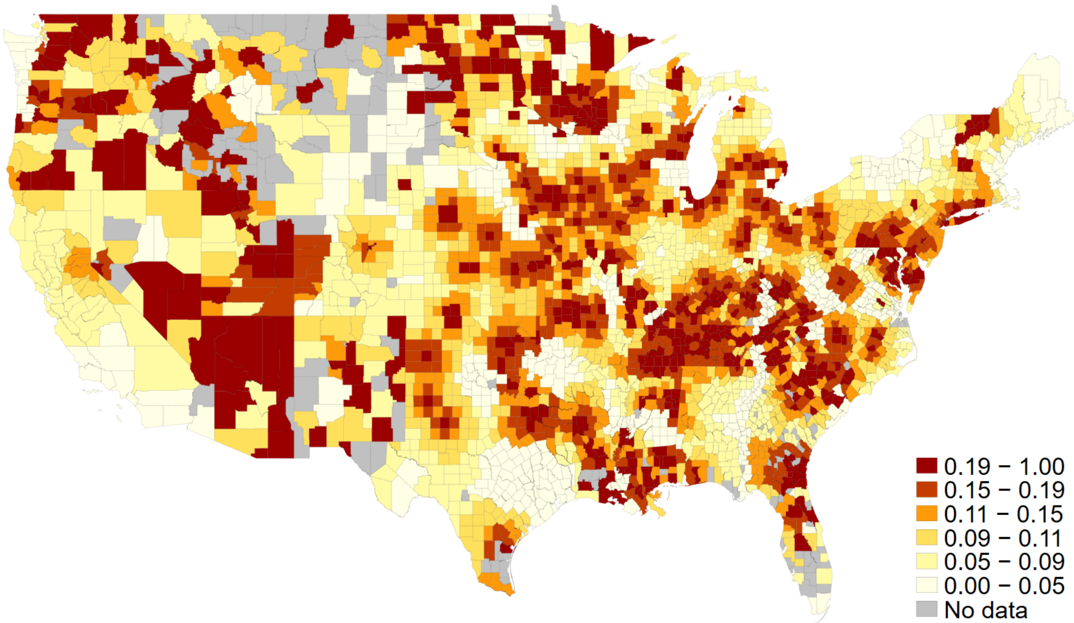
The telegraph today brings the news of the death in France of Maurice Bair of Goldfield. He was the youngest of two brothers in the service of Uncle Sam and went over a number of months ago.

(b) Anti-German Rumors in the News

Thinks Spies of Germans Are Working on the Inside in Departments in Capital

That German spies, working on the inside, in all probability, in one or more of the government departments, aided in conveying information to Germany regarding the movement of the American transports bearing the Pershing expedition to France, was the opinion expressed today by Senator Chamberlain of Oregon, chairman of the Senate military affairs committee.

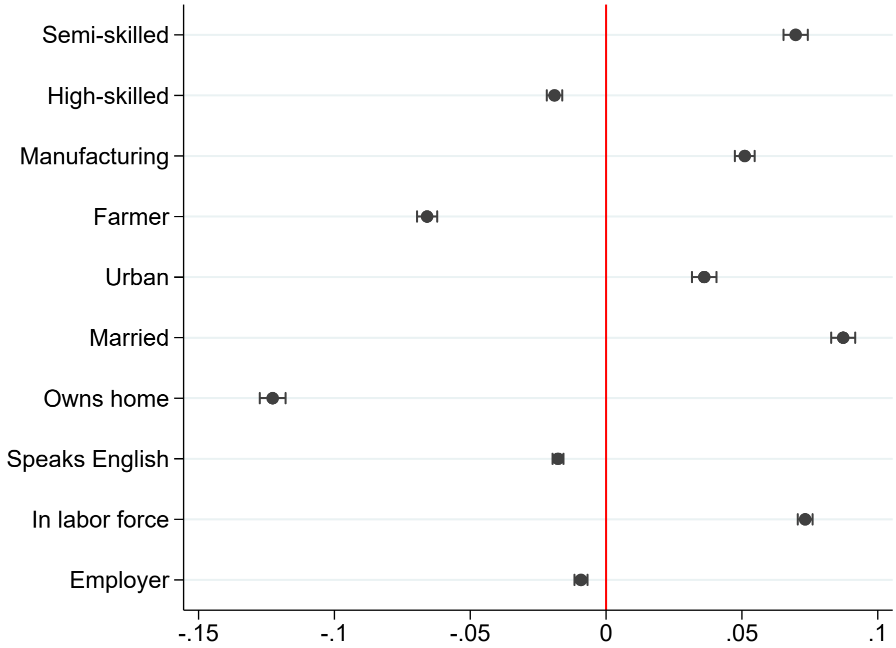
(c) Spatial Distribution of *Huns* Articles



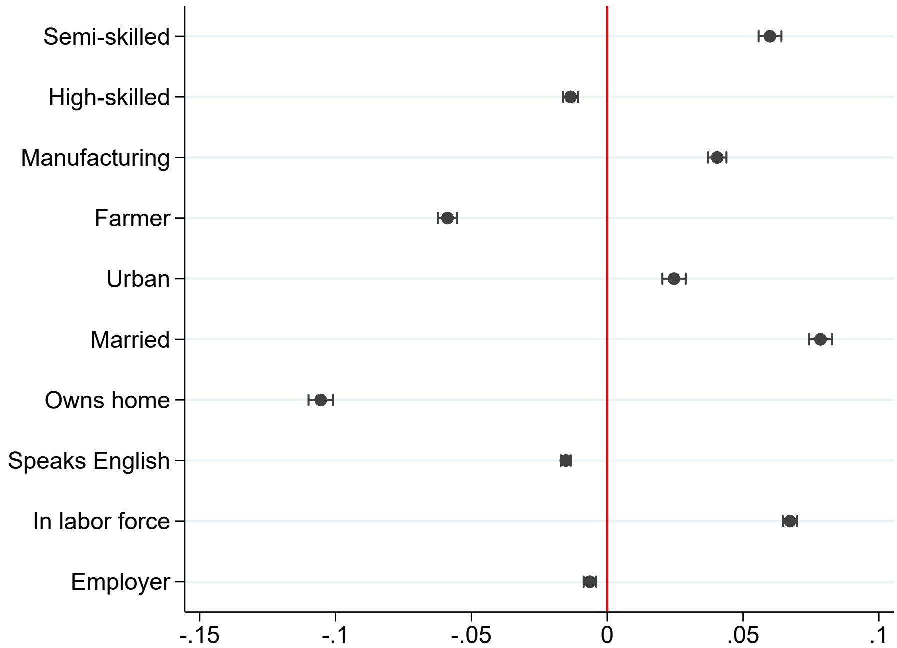
Note: Figure (a) local casualty report in the Carson City Daily from August 9, 1918. Figure (b) Evening Star Washington DC report of German spy activities in the government from July 4, 1917. Source: Chronicling America. Panel (c) maps the spatial distribution of the share of articles including *German(s)* which also include the word *Hun(s)* as measure of anti-German sentiment.

Figure 3.8: Characteristics of Predicted Movers and Stayers

(a) OLS Prediction



(b) 2SLS Prediction



Note: T-tests comparing observed characteristics of individuals who are predicted to move or stay in response to an increase in anti-German sentiment from OLS (figure a) and 2SLS (figure b) regressions. Anti-German sentiment is measured as the share of articles covering Germans mentioning the word *hun* or *huns* in a given county during the war years. If no newspaper outlet was available in a county, the share from the nearest county with an outlet was taken and weighted by the distance to that county. The 2SLS regression instruments the anti-German sentiment variable with the county-level WWI casualty rate. Error bars show 95% confidence intervals with variances being adjusted for different group sizes.

3.9 Appendix

A Additional Results

Table 3.A.1: Differences in Characteristics of Predicted Movers and Stayers

	Panel a: Movers Predicted by OLS					
	Stayer		Mover			
	Mean	St. Dev.	Mean	St. Dev.	Difference	t-stat
Semi-skilled	0.265	0.441	0.335	0.472	0.070^{***}	30.599
High-skilled	0.116	0.321	0.097	0.296	-0.019^{***}	-13.103
Manufacturing	0.132	0.339	0.183	0.387	0.051^{***}	27.431
Farmer	0.248	0.432	0.182	0.386	-0.066^{***}	-34.806
Urban	0.626	0.484	0.663	0.473	0.036^{***}	15.704
Married	0.604	0.489	0.691	0.462	0.087^{***}	38.707
Owns home	0.607	0.488	0.484	0.500	-0.123^{***}	-50.655
Speaks English	0.970	0.171	0.952	0.214	-0.018^{***}	-17.297
In labor force	0.841	0.365	0.915	0.279	0.073^{***}	52.580
Employer	0.077	0.266	0.067	0.251	-0.009^{***}	-7.546
Year of immigration	1883.679	11.119	1896.476	11.608	12.797^{***}	199.862
Age	40.020	16.257	41.816	11.837	1.796^{***}	30.216
Observations	539,911		46,014			
	Panel b: Movers Predicted by IV					
	Stayer		Mover			
	Mean	St. Dev.	Mean	St. Dev.	Difference	t-stat
Semi-skilled	0.265	0.441	0.325	0.468	0.060^{***}	27.983
High-skilled	0.116	0.320	0.102	0.303	-0.013^{***}	-9.625
Manufacturing	0.132	0.339	0.173	0.378	0.041^{***}	23.513
Farmer	0.248	0.432	0.189	0.392	-0.059^{***}	-32.346
Urban	0.627	0.484	0.652	0.476	0.025^{***}	11.219
Married	0.604	0.489	0.682	0.466	0.079^{***}	36.530
Owns home	0.606	0.489	0.501	0.500	-0.105^{***}	-45.992
Speaks English	0.970	0.171	0.954	0.208	-0.015^{***}	-16.135
In labor force	0.841	0.365	0.908	0.288	0.067^{***}	49.477
Employer	0.077	0.266	0.070	0.255	-0.006^{***}	-5.393
Year of immigration	1883.792	11.195	1896.197	11.725	12.406^{***}	191.818
Age	40.014	16.273	41.666	12.206	1.652^{***}	28.495
Observations	533,944		51,981			

Note: T-tests comparing observed characteristics of individuals who are predicted to move or stay in response to an increase in anti-German sentiment from OLS (panel a) and IV (panel b) regressions. Anti-German sentiment is measured as the share of articles covering Germans mentioning the word *hun* or *huns* in a given county. If no newspaper outlet was available in a county, the share from the nearest county with an outlet was taken and weighted by the distance to that county. The IV regression instruments anti-German sentiment with WWI casualties of a given county. Variances are adjusted for differences in group size. Significance levels are denoted by * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 3.A.2: OLS and Reduced Form Robustness to Census Linkage Quality

	Outcome: Pr(moved county = 1)					
	(1)	(2)	(3)	(4)	(5)	(6)
Discrimination × Post	-0.003	-0.002	0.000			
	(0.003)	(0.003)	(0.003)			
Discrimination × Post ×	-0.004*	0.003	0.002			
Both parents German	(0.002)	(0.002)	(0.001)			
Discrimination × Post ×	0.011**	0.009**	0.006**			
German-Born	(0.005)	(0.004)	(0.003)			
Abs. birth year diff = 1 × Post		0.021***	0.042***			
		(0.002)	(0.002)			
Abs. birth year diff = 2 × Post		0.208***	0.237***			
		(0.007)	(0.007)			
Abs. birth year diff = 3 × Post		0.357***	0.422***			
		(0.012)	(0.012)			
Unique match × Post			-0.338***			
			(0.012)			
Casualty Rate × Post				-0.415*	-0.372*	-0.266
				(0.221)	(0.217)	(0.202)
Casualty Rate × Post ×				0.236***	0.194***	0.098***
Both parents German				(0.043)	(0.039)	(0.034)
Casualty Rate × Post ×				0.777***	0.702***	0.450***
German-Born				(0.168)	(0.152)	(0.106)
Abs. birth year diff = 1 × Post					0.021***	0.041***
					(0.002)	(0.002)
Abs. birth year diff = 2 × Post					0.203***	0.234***
					(0.007)	(0.007)
Abs. birth year diff = 3 × Post					0.349***	0.416***
					(0.011)	(0.012)
Unique match × Post						-0.327***
						(0.011)
Observations	1,152,620	1,152,620	1,152,620	1,170,997	1,170,314	1,170,314
Individuals	576,310	576,310	576,310	585,157	585,157	585,157
Adj. R^2	0.242	0.273	0.431	0.246	0.280	0.320

Note: The outcome is an indicator for whether an individual moved county between 1910 and 1920. Discrimination is the share of newspaper articles mentioning Germans and using the word Huns during the war years. Data are for individuals linked from the 1910 to 1920 full count Census files with German origin or ancestry, i.e. Germans, first-, and second-generation German-Americans. The both parents German category refers American-born individuals only. Baseline controls are measured in 1910 and interacted with the 1920 indicator and include: urban status, skill group, farm ownership, employment status, literacy, marital status, years living in the U.S. (if German-born), family size, school attendance in 1910, labor force status, number of weeks unemployed. The absolute birth year difference is between an individual's observation in 1910 and the linked observation in 1920. Unique match is an indicator for whether an individual was uniquely linked from 1910 to 1920 without any other competing possible match. Standard errors are clustered at the county of residence in 1910. Significance levels are denoted by * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 3.A.3: 2SLS Regression Robustness to Census Linkage Quality

	Outcome: Pr(moved county = 1)				
	(1)	(2)	(3)	(4)	(5)
Discrimination × Post	-0.117	-0.102	-0.058	-0.043	-0.075
	(0.077)	(0.073)	(0.064)	(0.057)	(0.058)
Discrimination × Post ×	0.071^{**}	0.061^{**}	0.036	0.032	0.106^{**}
Both parents German	(0.034)	(0.031)	(0.024)	(0.023)	(0.045)
Discrimination × Post ×	0.269^{**}	0.246^{**}	0.168^{*}	0.163^{*}	0.143^{**}
German-born	(0.131)	(0.122)	(0.092)	(0.093)	(0.067)
Abs. birth year diff = 1 × Post		0.023^{***}	0.040^{***}	0.039^{***}	0.040^{***}
		(0.003)	(0.003)	(0.003)	(0.002)
Abs. birth year diff = 2 × Post		0.196^{***}	0.039^{***}	0.224^{***}	0.226^{***}
		(0.011)	(0.011)	(0.012)	(0.010)
Abs. birth year diff = 3 × Post		0.327^{***}	0.392^{***}	0.391^{***}	0.400^{***}
		(0.023)	(0.024)	(0.024)	(0.019)
Unique match × Post			-0.293^{***}	-0.289^{***}	-0.299^{***}
			(0.028)	(0.029)	(0.019)
State-specific trends				Yes	Yes
Group-specific trends					Yes
Observations	1,152,474	1,152,474	1,152,474	1,152,474	1,152,474
Individuals	576,237	576,237	576,237	576,237	576,237
K-P F-stat	15.11	15.07	14.81	12.78	9.80

Note: The outcome is an indicator for whether an individual moved county between 1910 and 1920. Discrimination is the share of newspaper articles mentioning Germans and using the word Huns during the war years and is instrumented with the WWI casualty rate in a given county interacted with a post-war and the corresponding nativity indicators. Data are for individuals linked from the 1910 to 1920 full count Census files with German origin or ancestry, i.e. Germans, first-, and second-generation German-Americans. The both parents German category refers American-born individuals only. Baseline controls are measured in 1910 and interacted with the 1920 indicator and include: urban status, skill group, farm ownership, employment status, literacy, marital status, years living in the U.S. (if German-born), family size, school attendance in 1910, labor force status, number of weeks unemployed. State-specific trends are state fixed effects interacted with the 1920 indicator. The absolute birth year difference is between an individual's observation in 1910 and the linked observation in 1920. Unique match is an indicator for whether an individual was uniquely linked from 1910 to 1920 without any other competing possible match. Group-specific time trends are nativity indicators for those with German parentage, or German-born interacted with the 1920 indicator. Standard errors are clustered at the county of residence in 1910. Significance levels are denoted by * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 3.A.4: Economic Effects of WWI-Induced Outflows of Germans - OLS Regression with Time-Varying Population and Labor Market Controls

	(1) ln(wages)	(2) ln(productivity)	(3) ln(input value)	(4) No. of firms	(5) firm size
Post-war German Outflow	-0.016***	-0.021	-0.061**	-11.676***	-2.962***
	(0.004)	(0.017)	(0.025)	(1.842)	(0.443)
Pre-war German share	0.009***	0.007***	0.109***	6.877***	-0.016
	(0.001)	(0.003)	(0.010)	(0.854)	(0.070)
Observations	10,474	10,474	10,474	10,474	10,474
Counties	2,258	2,258	2,258	2,258	2,258
Adj. R^2	0.878	0.901	0.972	0.624	0.707

Note: Regressions of economic county-level outcomes from 1900 to 1940 on the outflow of German population after WWI net of total county population changes. Log wages are per worker, productivity is the log real value of manufacturing output, input value is the log of value of materials used for production, the number of firms refers to the total number of manufacturing establishments, and firm size measures the average number of workers per manufacturing establishment. Controls are pre-war county characteristics interacted with a post-war indicator and include the average share of German population before 1910, the World War I draft rate, population, male-to-female ratio, share of manufacturing employment, and the share of urban population. Time varying controls include the total male and female population, and labor employed in manufacturing and agriculture. Monetary values are deflated to 1910 U.S. Dollars. Standard errors are clustered at the county-level. Significance levels are denoted by * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

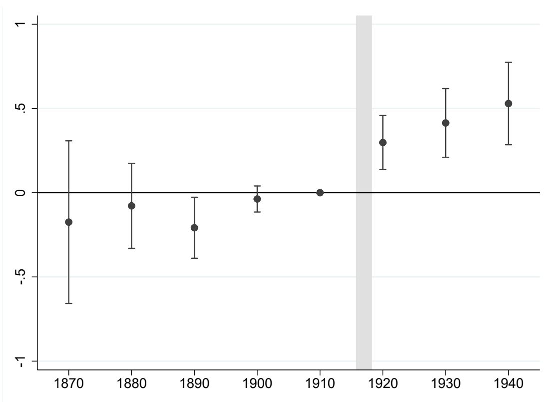
Table 3.A.5: Economic Effects of WWI-Induced Outflows of Germans - 2SLS Regression with Time-Varying Population and Labor Market Controls

	(1) ln(wages)	(2) ln(productivity)	(3) ln(input value)	(4) No. of firms	(5) firm size
Post-war German Outflow	-0.083**	-0.295*	-0.353	-6.917	-5.662
	(0.036)	(0.167)	(0.226)	(7.360)	(3.946)
Pre-war German share	0.019***	0.048*	0.151***	6.129***	0.385
	(0.005)	(0.025)	(0.035)	(1.389)	(0.601)
Observations	10,367	10,367	10,367	10,367	10,367
Counties	2,230	2,230	2,230	2,230	2,230
K-P F-stat	31.909	31.909	31.909	31.909	31.909

Note: Regressions of economic county-level outcomes from 1900 to 1940 on the outflow of German population after WWI net of total county population changes. The outflow of German population after WWI is instrumented with the county-level WWI casualty rate. Log wages are per worker, productivity is the log real value of manufacturing output, input value is the log of value of materials used for production, the number of firms refers to the total number of manufacturing establishments, and firm size measures the average number of workers per manufacturing establishment. Controls are pre-war county characteristics interacted with a post-war indicator and include the average share of German population before 1910, the World War I draft rate, population, male-to-female ratio, share of manufacturing employment, and the share of urban population. Time varying controls include the total male and female population, and labor employed in manufacturing and agriculture. Monetary values are deflated to 1910 U.S. Dollars. Standard errors are clustered at the county-level. Significance levels are denoted by * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

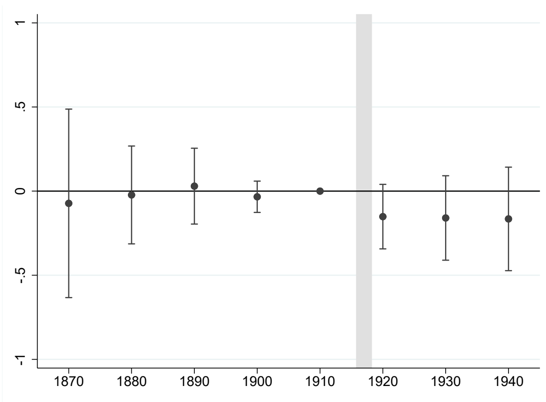
Figure 3.A.1: Effect of WWI Casualty Rates on the Share of Germans by Casualty Quintile

(a) Quintile 1

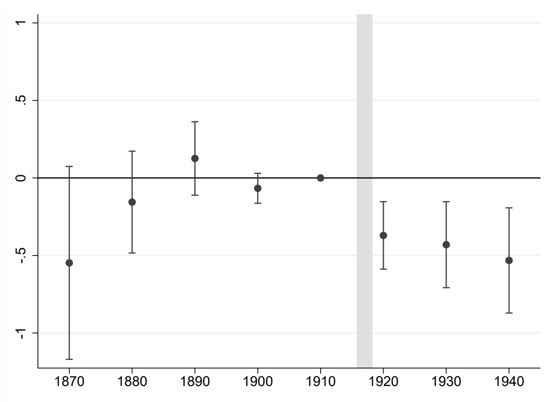


(b) Quintile 2

(c) Quintile 4

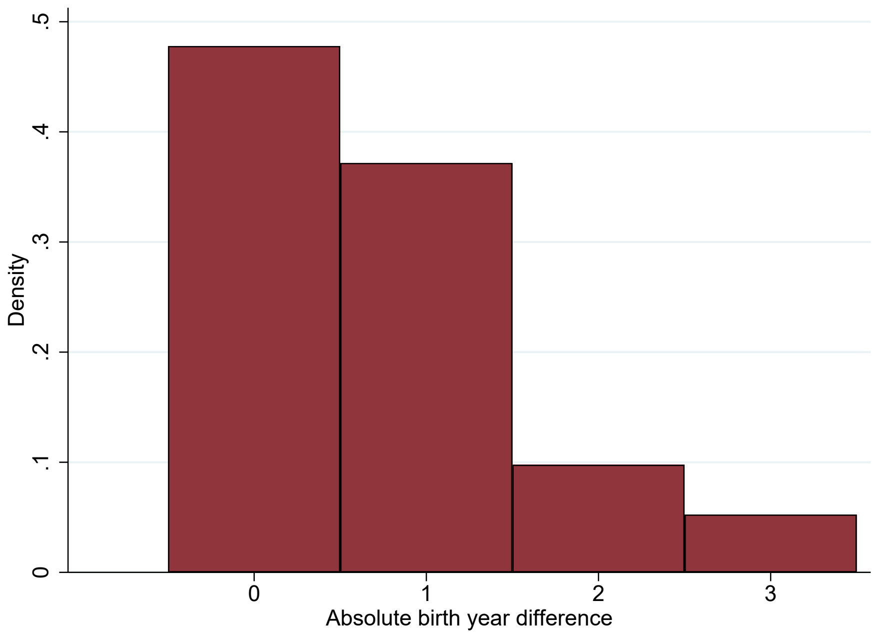


(d) Quintile 5



Note: Coefficient plots from a regression of the share of German population on quintiles of the WWI casualty rate distribution interacted with time fixed effects from 1870 to 1940. The baseline quintile is the third quintile. The regression includes county and time fixed effects. Controls include the WWI draft rate, as well as the average pre-war population, employment share in manufacturing, and male-to-female ratio interacted with time fixed effects. Error bars show 95% confidence intervals with standard errors being clustered at the county level.

Figure 3.A.2: Absolute Birth Year Difference of Linked Individuals



Note: Absolute birth year difference of linked Germans and German-Americans from the 1910 and 1920 Census files.

B Data Appendix

We digitized the WWI casualty data from Haulsee, Howe and Doyle (1920) for the Army and Washington (1920) for the Navy. Both sources cover almost 80,000 of the 110,000 total U.S. war deaths. The Army published residence information of the fallen soldiers together with their full name, rank, and cause of death. An example is provided in figure 3.B.1. The Navy published the residence information of a soldier's next of kin. In most cases this would be the spouse or their parents. Using this information, we geo-coded all residences to the county level using the 1910 county border definitions. The corresponding county FIPS codes then allowed us to link the casualty information with the aggregate county level information from the Census as well as with the individual-level Census data.

The most common cause of death were directly related to combat. 42.82% of soldiers were killed in action, 35.03% died of disease, another 17.11% of wounds or injuries, and 5.04% were killed in accidents. Roberts and Burda (2018) provide an analysis of the correlates of WWI casualties and socioeconomic characteristics at the county level. They show that Northern counties with higher war mortality rates were predominantly rural and had higher illiteracy rates, while in the South this relation was reversed where more urban counties saw higher casualty rates. The proportion of African American men had no effect on mortality rates.

We also digitized information on the number of enlisted men by county from table 20 in the Final Report of the Provost Marshal General (Crowder, 1920). The report lists the total number of soldiers called for service, those who were eventually inducted, accepted, and rejected, and those whose draft was ultimately canceled, for instance because they received an occupational deferment. The numbers are reported for each local draft board with most boards being responsible for a single county. For larger counties and cities, multiple boards were responsible for the drafting such as in Birmingham, Alabama, as shown in figure 3.B.2. Multiple boards for a single county were collapsed into one observation and given the FIPS code of the corresponding county. This covers the universe of the 2,960 counties in 1910.

Figure 3.B.1: WWI Casualty Lists

CALIFORNIA

KILLED IN ACTION

Lieutenant Colonels

CRAIG, J. M., San Francisco.
HOLLIDAY, William E., Santa Monica.

Majors

BEASLEY, Shadworth O., San Francisco.
MILLER, Oscar F., Los Angeles.
SMYTH, Roy Melvin, Alameda.

Captains

MacPHERSON, Harry H., San Francisco.
SMITH, Clarence F., Los Angeles.
VARNEY, Kit Roberts, San Francisco.

Lieutenants

BABCOCK, Robert C., San Francisco.
BARRY, David M., Santa Barbara.
BEACH, Egbert William, Piedmont.
BELL, Kenneth C., Pasadena.
BOYER, Herbert, San Francisco.
CROWELL, Fleming M., Los Angeles.
DAVIDSON, Gilford C., San Francisco.
ELAM, Edwin M., Berkeley.
FULTON, Hugh, Oakland.
GARD, Frank J., Glendora.
HAMMER, Earl M., San Francisco.
HANLY, William J., Oakland.
HARDING, Stacy Ludden, Antioch.
HARTER, Clifford C., Santa Barbara.
HITCHCOCK, Roger W., Los Angeles.
HOOPER, William J., Alameda.
KELLY, Charles J., San Francisco.
KIRK, Theodore T., Covina.
LANGENBACH, Paul J., Liveoak.
MADISON, Clinton R., Petaluma.
MARTIN, Leon, Berkeley.
McELROY, J. Willis, Berkeley.
NEVIUS, Ruliff, Los Angeles.
PARROTT, Edmund A., San Mateo.
PARTSCH, Herman D., Hayward.
ROBERTS, Edgar E., Chico.
ROBERTSON, David M., San Diego.
SHEPHERD, John S., Los Angeles.
SIMONDS, Albert C., Los Angeles.
SMITH, Paul D., Banning.
STEPHENSON, Wayne B., San Francisco.
UMSTED, Rolla P., Spring Valley.
WATERHOUSE, Hascall F., Oakland.
WEBSTER, Willard, San Diego.

Gunnery Sergeant

BELL, Jesse J., San Diego.

Sergeants

ANDERSON, Alfred E. L., Fresno.
BAILAR, Clarence W., Berkeley.
BARNES, Thomas, Los Angeles.
BARNES, Wilson B., San Francisco.
BENAPFL, Roscoe G., Los Angeles.
BERGES, Gaston J., Salinas.
BISBEE, Earl B., Los Angeles.
BRIMER, Frank M., Los Angeles.
BURROWS, Charles A., Ventura.
CAHILL, Thomas F., San Francisco.
CARTER, Alfred, Oakland.
CARTER, Carl C., Fresno.
CIRAVEGNA, Louis A., Soulsbyville.
COOPER, Robert S., San Francisco.
CROSSEN, Vernon J., San Francisco.
CURROTO, Virgilio, San Francisco.
DENNISON, John, Los Angeles.
DIVINE, Louis S., Vallejo.
EARL, Irwin, Long Beach.
EPPERSON, Uriah M., Modesto.
EVANS, Pryce N., Crescent City.
FOSTER, Alfred John, Orland.
GILLESPIE, Ralph, Lodi.
GRIFFIN, Norman E., Los Angeles.
GUSTAFSON, Carl R., Escalon.
HAINES, Richard B., Watsonville.
JACOBSEN, Arthur C., Fresno.
JONES, Carl Castlemann, Oakland.
JONES, Henry, San Diego.

Sergeants—Continued

JONES, James M., San Francisco.
LAKE, Thomas J., Los Angeles.
LARSEN, Peter W., San Miguel.
LUY, Richard L., San Gabriel.
MacPHERSON, William M., Madera.
MANDEVILLE, John L., San Diego.
McCAUSLAND, Clinton, Ripon.
McFALL, Hope, Manteca.
McKINNON, Elwyn Charles, Los Angeles.
McMILLAN, Laning R., Corona.
MESTROVITCH, James I., Fresno.
PATTERSON, Frederick H., Los Angeles.
PETERSON, Peter N., Neroman.
POWELL, Ballard B., Sacramento.
ROBBINS, George W., Los Angeles.
ROSS, George W., Oakland.
ROSS, Karl E., Stockton.
SHEEHY, Norman R., Los Angeles.
SIMMONS, Melvin K., Fairfield.
STEVENS, Edward J., San Francisco.
SULLIVAN, John Q., Lost Hills.
SWEETNAM, John M., Sebastopol.
THOMPSON, Charles H. ? ? ?
WALTERS, Charles, San Diego.
WHITE, Thomas R., Sacramento.
WHITNEY, William E., Oakland.
WILLIAMS, Charles V. G., Chino.

Corporals

ADAMS, Herbert H., Oakdale.
AGGELER, Jerrold J., Stockton.
ALTMAN, Henry, San Francisco.
BAHNEY, John W., Sacramento.
BALLARD, Blackburn W., Colura.
BATCHELOR, Louis W., San Francisco.
BERNARD, Harry F., North San Diego.
BEYER, Peter, Tassayara.
BLAU, Otto H., San Francisco.
BLEDSOE, Lawrence E., Los Angeles.
BLISS, William P., Hanford.
BROKAW, Charles S., Colton.
BROPHY, Anselm G., Los Angeles.
CAMP, George W., Fresno.
CARY, Harold E., San Francisco.
CASAJUS, John B., Ryde.
CATLIN, Samuel L., Kingsburg.
COLBURN, Elbert F., San Jose.
COOPER, Robert W., Oxnard.
COVILL, William Frank, San Jose.
DAVIS, Frank G., Santa Paula.
DAWSON, Harry J., San Francisco.
DE SANTI, Narciso, San Francisco.
DEWEY, Charles M., Los Angeles.
DUNN, James A., San Francisco.
FEELY, Aloysius E., Fresno.
FORSTNER, Joseph J., Legunitas.
FOX, Murray S., Venice.
FRANK, Chauncey R., San Francisco.
GORSKY, Anthony, Pasadena.
GRIFFIN, Lee R., Hayward.
GUIDO, Ernest E., East Oakland.
HAGEDORN, William, San Francisco.
HARRIS, Harry L., San Francisco.
HATCHER, William F., Oxnard.
HIGGINS, Hugh V., San Francisco.
HOLDZKOM, Paul R., Imperial.
HOLLYWOOD, Leonard B., Alameda.
HUGILL, Thomas W., Lodi.
INGALLS, Earle E., Sespe.
JEFFERS, Amzi H., Redlands.
KEELEY, Julius O., Lindsay.
KITT, Don H., Los Angeles.
LANCASTER, Elmer N., Alliance.
LECORNU, Herman G., San Francisco.
LEVERS, William H., San Francisco.
LISTER, John M., San Francisco.
LORENSON, Edward H., Watsonville.
LUNN, William, Jr., San Francisco.
McCOLLEY, Robert T., Huntington Park.
MacCONNELL, Charles F., Los Angeles.
MADSEN, John, Petaluma.
MASTERSON, Barton William, Oakland.

Corporals—Continued

MILLER, Harry A., Oakland.
MORRIS, Fred L., Los Angeles.
NEEDHAM, Clyde W., Lodi.
NUNES, Alfred, Centerville.
PALMERLEE, Chester C., Long Beach.
PASSERINI, Frank, San Francisco.
PEDRIOLI, Louis, Modesto.
PERRY, William S., Jr., Berkeley.
PETERSON, Arthur L., Long Beach.
POORE, Raymond, Pasadena.
RICHESON, Franklin Carter, Dinuba.
ROBERTS, Harold William, San Francisco.
ROBINSON, Glen H., Pescadero.
RUBIDOUX, Mack J., Riverside.
SAXEY, Harry, Willow Creek.
SCHMALZ, John W., San Francisco.
SCHNEIDER, Harry N., Morgan Hill.
SHANKLAND, Claude G., Bakersfield.
SIEVERS, Maxwell H., Salinas.
SOUZA, Manuel, Jr., Cambria.
SOWELL, Vernon L., Lemoore.
SPARGO, John, San Francisco.
STAPLES, Guy W., Linden.
SWEET, Ora A., Crockett.
TROMBLY, Charles H., Pasadena.
VINTHER, Claudius F., Berkeley.
WALL, Earnest W., Sacramento.
WEYLANDT, Lester L., Peters.
WIENS, Gary, Los Angeles.
WILKINS, James H., Jr., San Rafael.
WILSON, Robert H., Los Angeles.
WOODWARD, Earl, Lathrop.

Buglers

ADAMOLI, Matteo, Guadalupe.
CURRY, Charles R., San Diego.
GIAMBRUNO, Isidore, Oakland.
HALL, John T., Santa Barbara.
IRWIN, Bernard, Stockton.

Chauffeur

BARKER, Voltaine, Selma.

Cooks

CROSSLAND, Bert S., Los Angeles.
LUND, John W., San Francisco.
STUETTIG, Herman, Los Angeles.
ZVIJERKOVICH, John, Marysville.

Engineer

BALL, Joseph Barker, El Monte.

Mechanics

CARSON, Ben C., Oakland.
EUSTACE, Patrick, San Francisco.
GILLIGAN, Patrick, San Francisco.
GRISEDALE, Francis T., East Bakersfield.
RUSTING, Joseph F., Oakland.

Musician

PEDROTTI, Faust, Santa Rosa.

Wagoners

GATTO, Peter, El Centro.
LAWLOR, Reuben, Oakland.
LITTLE, Stanley Harrison, Taft.
McGANNEY, Edward J., Smartsville.
SHERLOCK, Phillip, Stockton.

Privates

ACUNA, John E., San Gabriel.
ADAMS, Robert H., Blythe.
ADELSBACH, Harry Ben, Fresno.
AITKEN, William H., Chico.
ALLEN, Thomas, San Francisco.
ALLMAN, Henry J., Lanare.
ALVES, Frank, Oxnard.
ANDELSTEDT, Raymond D., San Bernardino.
ANDERSON, Charley E., Santa Rosa.
ANDERSON, James B., Clements.
ANDERSON, Simeon M., San Ramon.
ANDRADE, Joseph F., Santa Clara.
ANDRIJASEVICH, Stephen, Los Angeles.
APPLING, Marvin C., Lewis.
ARATA, Joe, Stockton.

Note: Example from Haulsee et al. (1920) page 139 for the state of California. Casualties are ordered by cause of death, rank, and alphabet. Soldier-level information includes state, rank, first, middle, and surnames, as well as the city or county of residence. Causes of death are killed in action (42.82%), disease (35.03%), wounds and injuries (17.11%), and accidents (5.04%). Army casualties total almost 80,000 of the overall 110,000 war deaths sustained by the United States during World War I.

Figure 3.B.2: WWI Draft and Enlistment Data

TABLE 20.—*Calls, inductions, acceptances, and rejections, by local boards.*

ALABAMA.

Local board.	Total called.	Total inducted.	Total accepted.	Total rejected.	Total rejected, cancellation of draft.
Autauga	517	592	514	78	....
Baldwin	529	533	471	62	....
Barbour	666	790	708	76	6
Bibb	624	615	559	56	....
BIRMINGHAM, No. 1	795	892	815	77	....
BIRMINGHAM, No. 2	1,274	1,457	1,314	142	1
BIRMINGHAM, No. 3	1,426	1,648	1,507	141	....
BIRMINGHAM, No. 4	877	964	885	79	....
BIRMINGHAM, No. 5	1,241	1,388	1,275	113	....
BIRMINGHAM, No. 6	569	665	608	57	....
Blount	464	488	448	40	....
Bullock	529	531	502	29	....
Butler	765	841	757	84	....
Calhcun	1,257	1,410	1,280	124	6
Chambers	965	1,037	960	77	....
Cherokee	502	533	492	41	....
Chilton	596	607	536	70	1
Choctaw	581	539	498	41	....
Clarke	921	918	813	102	3
Clay	613	642	581	61	....
Cleburne	328	411	374	37	....
Coffee	647	663	571	92	....
Colbert	794	797	718	77	2
Conecuh	710	805	714	81	10
Coosa	402	412	364	48	....
Covington	1,075	1,221	1,082	139	....
Crenshaw	669	685	608	76	1
Cullman	799	836	742	90	4
Dale	465	487	423	64	....
Dallas	881	948	872	76	....
Dekalb	785	801	713	88	....
Elmore	708	786	711	75	....
Escambia	624	670	591	79	....
Etowah	1,303	1,398	1,236	158	4
Fayette	482	550	512	38	....
Franklin	544	598	540	58	....
Geneva	668	760	662	98	....
Greene	266	272	255	17	....
Hale	492	488	440	48	....
Henry	542	602	531	71	....
Houston	960	1,049	914	134	1
Jackson	998	1,144	1,037	106	1
Jefferson No. 1	2,045	2,219	2,074	145	....
Jefferson No. 2	683	800	726	74	....
Jefferson No. 3	923	1,075	984	90	1
Lamar	551	541	480	55	6
Lauderdale	803	959	847	103	9
Lawrence	509	545	508	37	....
Lee	867	836	758	78	....
Limestone	1,118	1,086	982	104	....
Lowndes	531	563	525	38	....
Macon	574	591	550	41	....
Madison	1,489	1,394	1,256	138	....
Marengo	828	734	659	75	....
Marion	619	664	593	71	....
Marshall	867	924	841	83	....
MOBILE No. 1	761	761	709	49	3
MOBILE No. 2	898	943	861	82	....
Mobile	1,128	1,212	1,085	126	1
Monroe	742	808	726	80	2
MONTGOMERY	1,305	1,425	1,310	109	6
Montgomery	599	642	584	58	....
Morgan	1,033	1,118	1,037	77	4
Perry	447	473	422	50	1
Pickens	635	708	646	62	....
Pike	746	782	706	76	....
Randolph	618	661	607	54	....
Russell	498	525	458	67	....
St. Clair	637	712	633	79	....
Shelby	604	719	636	82	1
Sumter	486	492	450	42	....
Talladega	884	991	894	97	....
Tallapoosa	855	959	863	88	8
Tuscaloosa	1,297	1,440	1,292	148	....
Walker	1,309	1,225	1,105	120	....
Washington	420	405	359	46	....
Wilcox	713	723	657	66	....
Winston	263	277	259	18	....
Total	60,138	64,405	58,215	6,108	82

Note: Example from Crowder (1920) page 55 for the state of Alabama showing the number of called, inducted, accepted, and rejected soldiers for each local draft board. Draft boards were typically associated with a single county with the exception of larger municipalities which were served by multiple boards such as Birmingham.

Bibliography

Aaronson, Daniel and Bhashkar Mazumder, "The Impact of Rosenwald Schools on Black Achievement," *Journal of Political Economy*, 2011, *119* (5), 821–888.

Abramitzky, Ran, Leah Boustan, and Katherine Eriksson, "Europe's Tired, Poor, and Huddled Masses: Self Selection and Economic Outcomes in the Age of Mass Migrations," *American Economic Review*, 2012, *102* (5), 1832–1856.

_ , _ , and _ , "A Nation of Immigrants: Assimilation and Economic Outcomes in the Age of Mass Migrations," *Journal of Political Economy*, 2014, *122* (3), 467–506.

_ , _ , _ , James Feigenbaum, and Santiago Perez, "Best Practices for Automated Linking Using Historical Data: A Progress Report," mimeo 2018.

_ , Roy Mill, and Santiago Perez, "Linking Individuals Across Historical Sources: A Fully Automated Approach," Working Paper 24324, National Bureau of Economic Research, Cambridge, MA 2018.

Acemoglu, Daron, David H. Autor, and David Lyle, "Women, War, and Wages: The Effect of Female Labor Supply on the Wage Structure at Midcentury," *Journal of Political Economy*, 2004, *112* (3), 497–551.

Aigner, Dennis J., "Regression with a Binary Independent Variable Subject to Errors of Observation," *Journal of Econometrics*, 1973, *1*, 49–60.

Ainsworth, Martha and Deon Filmer, "Inequalities in Children's Schooling: AIDS, Orphanhood, Poverty, and Gender," *World Development*, 2006, *34* (6), 1099–1128.

Alesina, Alberto, Reza Baqir, and William Easterly, "Public Goods and Ethnic Divisions," *Quarterly Journal of Economics*, 1999, *114* (4), 1243–1284.

Anderson, Karen Tucker, "Last Hired, First Fired: Black Women Workers during World War II," *Journal of American History*, 1982, *69* (1), 82–97.

Angrist, Joshua and Alan B. Krueger, "Why Do World War II Veterans Earn More Than Nonveterans?.," *Journal of Labor Economics*, 1994, *12* (1), 74 – 97.

Angrist, Joshua D., "Lifetime Earnings and the Vietnam Era Draft Lottery: Evidence from Social Security Administrative Records.," *American Economic Review*, 1990, *80* (3), 313 – 336.

_ , "Estimating the Labor Market Impact of Voluntary Military Service Using Social Security Data on Military Applicants.," *Econometrica*, 1998, *66* (2), 249 – 288.

Bailey, Martha, Connor Cole, Morgan Henderson, and Catherine Massey, "How well do automated methods perform in historical samples? Evidence from new ground truth," Working Paper 24019, National Bureau of Economic Research, Cambridge, MA 2017.

Bailey, Martha J. and William J. Collins, "The Wage Gains of African-American Women in the 1940s," *Journal of Economic History*, 2006, *66* (3), 737–777.

Baten, Joerg, Nicola Bianchi, and Petra Moser, "Compulsory licensing and innovation - Historical evidence from German patents after WWI," *Journal of Development Economics*, 2017, *126*, 231–242.

Bayer, Patrick and Kerwin K. Charles, "Divergent Paths: A New Perspective on Earnings Differences Between Black and White Men Since 1940," *Quarterly Journal of Economics*, 2018, *133* (3), 1459–1501.

Becker, Sascha O. and Andreas Ferrara, "Consequences of Forced Migration: A Survey of Recent Findings," *Labour Economics, forthcoming. DOI: https://doi.org/10.1016/j.labeco.2019.02.007*, 2019.

Beegle, Kathleen, Joachim De Weerdt, and Stefan Dercon, "The intergenerational impact of the African orphans crisis: a cohort study from an HIV/AIDS affected area," *International Journal of Epidemiology*, 2009, *38*, 561–568.

_ , _ , and _ , "Orphanhood and human capital destruction: Is there persistence into adulthood?," *Demography*, February 2010, *4* (1), 163–180.

Belloni, Alexandre, Victor Chernozhukov, and Christian Hansen, "High-Dimensional Methods and Inference on Structural and Treatment Effects," *Journal of Economics Perspectives*, 2014, *28* (2), 29–50.

Bhuller, Manudeep, Gordon B. Dahl, Katrine V. Loken, and Magne Mogstad, "Intergenerational Effects of Incarceration," *AEA Papers and Proceedings*, 2018, *108*, 234–240.

Biavaschi, Costanza, Corrado Giulietti, and Zahra Siddique, "The Economic Payoff of Name Americanization," *Journal of Labor Economics*, 2017, *35* (4), 1089–1116.

Bound, John and Richard B. Freeman, "What Went Wrong? The Erosion of Relative Earnings and Employment Among Young Black Men in the 1980s," *Quarterly Journal of Economics*, 1992, *107* (1), 201–232.

Boustan, Leah, *Competition in the Promised Land: Black Migrants in Northern Cities and Labor Markets*, Princeton University Press, Princeton, NJ, 1961.

_ , "Competition in the Promised Land: Black Migration and Racial Wage Convergence in the North, 1940-1970," *Journal of Economic History*, 2009, *69* (3), 755–782.

_ , "Was Postwar Suburbanization "White Flight"? Evidence from the Black Migration," *Quarterly Journal of Economics*, 2010, *125* (1), 417–443.

_ and Robert A. Margo, "A silver lining to white flight? White suburbanization and African-American Homeownership, 1940-1980," *Journal of Urban Economics*, 2013, *78*, 71–80.

Camargo, Braz, Ralph Stinebrickner, and Todd Stinebrickner, "Interracial Friendships in College," *Journal of Labor Economics*, 2010, *28* (4), 861–892.

Cameron, Colin A., Jonah B. Gelbach, and Douglas L. Miller, "Bootstrap-Based Improvements for Inference with Clustered Errors," *Review of Economics and Statistics*, 2008, *90* (3), 414–427.

Case, Anne, Christina Paxson, and Joseph Ableidinger, "Orphans in Africa," Working Paper 9213, National Bureau of Economic Research, Cambridge, MA 2002.

Cavalcanti, Tiago and José Tavares, "The Output Cost of Gender Discrimination: A Model-Based Macroeconomics Estimate," *Economic Journal*, 2016, *126*, 109–134.

Chamarbagwala, Rubiana and Hilcias E. Moran, "The Human Capital Consequences of Civil War: Evidence from Guatemala," *Journal of Development Economics*, 2011, *94* (1), 41 – 61.

Chambers, J.W., *To Raise an Army: The Draft Comes to Modern America*, Free Press, New York, NY, 1987.

Chetty, Raj, Nathaniel Hendren, Patrick Kline, and Emmanuel Saez, "Where is the Land of Opportunity? The Geography of Intergenerational Mobility in the United States," *Quarterly Journal of Economics*, 2014, *129* (4), 1553–1623.

Clubb, Jerome M., William H. Flanigan, and Nancy H. Zingale, "Electoral Data for Counties in the United States: Presidential and Congressional Races, 1840-1972," *ICPSR08611-v1. Ann Arbor, MI: Inter-university Consortium for Political and Social Research [distributor], 2006-11-13. DOI: https://doi.org/10.3886/ICPSR08611.v1*, 2006.

Coate, Stephen and Glenn C. Loury, "Will Affirmative-Action Policies Eliminate Negative Stereotypes?," *American Economic Review*, 1993, *83* (5), 1220–1240.

Collins, William J., "African-American Economic Mobility in the 1940s: A Portrait from the Palmer Survey," *Journal of Economic History*, 2000, *60* (3), 756–781.

_ , "Race, Roosevelt, and Wartime Production: Fair Employment in World War II Labor Markets," *American Economic Review*, 2001, *91* (1), 272–286.

_ **and Robert A. Margo**, *'Historical Perspectives on Racial Differences in Schooling in the United States', in Hanushek, E. and Welch, F. (eds.)*, Vol. 1 of *Handbook of the Economics of Education*, Elsevier, North Holland, NL, 2006.

_ **and** _ , "Race and Home Ownership from the End of the Civil War to the Present," *American Economic Review Papers and Proceedings*, 2011, *101* (3), 355–359.

Conley, Timothy G., "GMM estimation with cross sectional dependence," *Journal of Econometrics*, 1999, *92* (1), 1–45.

_ **, Christian B. Hansen, and Peter E. Rossi**, "Plausibly Exogenous," *Review of Economics and Statistics*, 2012, *94* (1), 260–272.

Cook, Lisa D., "Violence and Economic Activity: Evidence from African American Patents, 1870-1940," *Journal of Economic Growth*, 2014, *19* (2), 221–257.

Corak, Miles, "Death and Divorce: The Long-Term Consequences of Parental Loss on Adolescents," *Journal of Labor Economics*, 2001, *19* (3), 682–715.

Costa, Dora L., "Pensions and Retirement: Evidence from Union Army Veterans," *The Quarterly Journal of Economics*, 1995, *110* (2), 297–319.

_ , "Displacing the Family: Union Army Pensions and Elderly Living Arrangements," *Journal of Political Economy*, 1997, *105* (6), 1269–1292.

_ **and Matthew E. Kahn**, "Cowards and Heroes: Group Loyalty in the American Civil War," *Quarterly Journal of Economics*, 2003, *118* (2), 519–548.

_ **and** _ , "Forging a New identity: The Cost and Benefits of Diversity in Civil War Combat Units," *The Journal of Economic History*, 2006, *66* (4), 936–962.

_ **and** _ , "Surviving Andersonville: The Benefits of Social Networks in POW Camps," *American Economic Review*, 2007, *97* (4), 1467–1487.

_ **and** _ , *Heroes and Cowards: The Social Face of War*, Princeton University Press, Princeton, NJ, 2008.

_ **and** _ , "Health, Wartime Stress, and Unit Cohesion: Evidence from Union Army Veterans.," *Demography*, 2010, *47* (1), 45 – 66.

Crowder, Enoch H., *Final Report of the Provost Marshal General to the Secretary of War on the Operations of the Selective Service System to July 15, 1919*, Government Printing Office, Washington, DC, 1920.

Cunha, Flavio, James J. Heckman, and Lance Lochner, *'Interpreting the Evidence on Life Cycle Skill Formation', in Erik Hanushek and F. Welch (eds.)*, Vol. 1 of *Handbook of the Economics of Education*, Elsevier, North Holland, NL, 2006.

Currarini, Sergio, Matthew O. Jackson, and Paolo Pin, "An Economic Model of Friendship: Homophily, Minorities, and Segregation," *Econometrica*, 2009, *77* (4), 1003–1045.

Davidson, Russell and James G. MacKinnon, "Wild Bootstrap Tests for IV Regression," *Journal of Business and Economic Statistics*, 2010, *28* (1), 128–144.

Dippel, Christian, Robert Gold, Stephan Heblich, and Rodrigo Pinto, "Instrumental Variables and Causal Mechanisms: Unpacking the Effect of Trade on Workers and Voters," *NBER Working Paper 23209*, 2017, *Revised June 2018.*

Dobbie, Will, Hans Gronqvist, Susan Niknami, Marten Palme, and Mikael Priks, "The Intergenerational Effects of Parental Incarceration," Working Paper 24186, National Bureau of Economic Research, Cambridge, MA 2018.

Domingues, Patrick and Thomas Barre, "The Health Consequences of the Mozambican Civil War: An Anthropometric Approach.," *Economic Development and Cultural Change*, 2013, *61* (4), 755 – 788.

Donohue, John J. III and James Heckman, "Continuous Versus Episodic Change: The Impact of Civil Rights Policy on the Economic Status of Blacks," *Journal of Economic Literature*, 1991, *29* (4), 1603–1643.

Eriksson, Katherine, "Moving North and into jail? The great migration and black incarceration," *Journal of Economic Behavior and Organization*, 2018, *forthcoming. DOI: https://doi.org/10.1016/j.jebo.2018.04.024.*

_ **and Gregory T. Niemesh**, "Death in the Promised Land: The Great Migration and Black Infant Mortality," *mimeo*, 2016.

Evans, David K. and Edward Miguel, "Orphans and Schooling in Africa: A Longitudinal Analysis," *Demography*, 2007, *44* (1), 35–57.

Feigenbaum, James J., "A Machine Learning Approach to Census Record Linking," Working Paper 2016.

Ferrie, Joseph P., "A New Sample of Males Linked from the Public Use Microdata Sample of the 1850 U.S. Federal Census of Population to the 1860 U.S. Federal Census Manuscript Schedules," *Historical Methods: A Journal of Quantitative and Interdisciplinary History*, 1996, *29* (4), 141–156.

Fishback, Price V. and Joseph A. Cullen, "Second World War spending and local economic activity in US counties, 1939-58," *Economic History Review*, 2013, *66* (4), 975–992.

_ **, William C. Horrace, and Shawn Kantor**, "The impact of New Deal expenditures on mobility during the Great Depression," *Explorations in Economic History*, 2006, *43* (2), 179–222.

Flynn, George Q., *The Draft, 1940-1973*, University Press of Kansas, Lawrence, KS, 1993.

Fogel, Robert, "Public Use Tape on the Aging of Veterans of the Union Army: U.S. Federal Census Records, 1850, 1860, 1900, 1910," Technical Report C3, Center for Population Economics, University of Chicago Graduate School of Business, and Department of Economics, Brigham Young University 2000.

Forstall, Richard L., *Population of States and Counties of the United States: 1790-1990*, U.S. Bureau of the Census, Washington, D.C., 1996.

Fouka, Vasiliki, "Backlash: The Unintended Effects of Language Prohibition in US Schools after World War I," *mimeo*, 2018.

_ , "How do immigrants respond to discrimination? The case of Germans in the US during World War I," *American Political Science Review*, 2018, *forthcoming.*

Friedline, Terri, Rainier D. Masa, and Gina A.N. Chowa, "Transforming wealth: Using the inverse hyperbolic sine (IHS) and splines to predict youth's math achievement," *Social Science Research*, 2015, *49*, 264–287.

Fryer, Roland G., "Guess Who's Been Coming to Dinner? Trends in Interracial Marriage over the 20th Century," *Journal of Economic Perspectives*, 2007, *21* (2), 71–90.

Galdo, Jose, "The Long-Run Labor-Market Consequences of Civil War: Evidence from the Shining Path in Peru.," *Economic Development and Cultural Change*, 2013, *61* (4), 789 – 823.

Gertler, Paul, David I. Levine, and Minnies Ames, "Schooling and Parental Death," *The Review of Economics and Statistics*, 2004, *86* (1), 211–225.

Getis, Arthur and J. Keith Ord, "The Analysis of Spatial Association by Use of Distance Statistics," *Geographical Analysis*, 1992, *24* (3), 189–206.

Glass, Charles, *Deserter: A Hidden Story of the Second World War*, The Penguin Press, New York, NY, 2013.

Glasson, William H., *Federal Military Pensions in the United States*, New York: Oxford University Press, Oxford, UK, 1918.

Goldin, Claudia, "The Role of World War II in the Rise of Women's Employment," *American Economic Review*, 1991, *81* (4), 741–756.

— **and Claudia Olivetti**, "Shocking Labor Supply: A Reassessment of the Role of World War II on Womens Labor Supply," *American Economic Review Papers and Proceedings*, 2013, *103* (3), 257–262.

— **and Robert A. Margo**, "The Great Compression: The Wage Structure in the United States at Mid-Century," *Quarterly Journal of Economics*, 1992, *107* (1), 1–34.

Gruber, Jonathan, "Is Making Divorce Easier Bad for Children? The Long-Run Implications of Unilateral Divorce," *Journal of Labor Economics*, 2004, *22* (4), 799–833.

Hamermesh, Daniel S. and Jeff E. Biddle, "Beauty and the Labor Market," *American Economic Review*, 1994, *84* (5), 1174–1194.

Haulsee, William M., Frank G. Howe, and Alfred C. Doyle, *Soldiers of the Great War, Vols. 1-3*, Soldiers Record Publishing Association, Washington, DC, 1920.

Heedegaard, Morten S. and J.-R. Tyran, "The Price of Prejudice," *American Economic Journal: Applied Economics*, 2018, *10* (1), 40–63.

Higham, John, *Strangers in the Land*, Rutgers University Press, New Brunswick, NJ, 1998.

Hornbeck, Richard and Suresh Naidu, "When the Levee Breaks: Black Migration and Economic Development in the American South," *American Economic Review*, 2014, *104* (3), 963–990.

Hornung, Erik, "Immigration and the Diffusion of Technology: The Huguenot Diaspora in Prussia," *American Economic Review*, 2014, *104* (1), 84–122.

Imbens, Guido and Wilbert van der Klaauw, "Evaluating the Cost of Conscription in The Netherlands.," *Journal of Business and Economic Statistics*, 1995, *13* (2), 207 – 215.

Jaworski, Taylor, "You're in the Army Now: The Impact of World War II on Women's Education, Work, and Family," *Journal of Economic History*, 2014, *74* (1), 169–195.

Kendrick, John W., *Productivity Trends in the United States*, Princeton University Press, Princeton, NJ, 1961.

Kondo, Keisuke, "Hot and cold spot analysis using Stata," *Stata Journal*, 2016, *16* (3), 612–631.

Kovac, Dejan, "Do Fathers Matter Paternal Mortality and Childrens Long-Run Outcomes," Working Paper 609, Princeton University, Industrial Relations Section 2017.

Kriner, Douglas and Francis Shen, *The Casualty Gap - The Causes and Consequences of American Wartime Inequalities*, Oxford University Press, Oxford, UK, 2010.

Larsen, Timothy B., "The Strange Career of Jim Crow: Labor Scarcity and Racial Treatment in the American South," *mimeo*, 2015.

Lee, Ulysses, *The Employment of Negro Troops*, in Conn, S. (eds.) United States Army in World War II, Center of Military History U.S. Army, Washington D.C., 1965.

Logan, Trevon D. and John M. Parman, "Segregation and Homeownership in the Early Twentieth Century," *American Economic Review Papers and Proceedings*, 2017, *107* (5), 410–414.

Long, Clarence D., "Wages by Occupational and Individual Characteristics," in "Wages and Earnings in the United States, 1860-1890" 1960, pp. 94–108.

Lübke, Frederick C., *Bonds of Loyalty*, Northern Illinois University Press, DeKalb, IL, 1974.

_ , *Germans in the New World: Essays in the History of Immigration*, University of Illinois Press, DeKalb, IL, 1999.

Maloney, Thomas N., "Wage Compression and Wage Inequality Between Black and White Males in the United States, 1940-1960," *Journal of Economic History*, 1994, *54* (2), 358–381.

_ **and Ken R. Smith**, "Parental Loss, Economic Inequality, and Intergenerational Mobility in the Long Run: Evidence from the Utah Population Database," Technical Report, mimeo 2018.

Margo, Robert A., "Explaining Black-White Wage Convergence, 1940-1950," *Industrial and Labor Relations Review*, 1995, *48* (3), 470–481.

Matthews, Donald and James Prothro, "Negro Political Participation Study, 1961-1962," *Ann Arbor, MI: Inter-university Consortium for Political and Social Research [distributor], 2006-08-15. https://doi.org/10.3886/ICPSR07255.v3*, 2006.

Mazumder, Soumyajit, "Becoming White: How Mass Warfare Turned Immigrants into Americans," *mimeo*, 2018.

McPherson, James M., *Battle Cry of Freedom: The American Civil War*, Oxford: Oxford University Press, Oxford, UK, 1988.

Meyer, Bruce D. and Nikolas Mittag, "Misclassification in binary choice models," *Journal of Econometrics*, 2017, *200* (2), 295–311.

Miguel, Edward and Gerard Roland, "The Long-Run Impact of Bombing Vietnam.," *Journal of Development Economics*, 2011, *96* (1), 1 – 15.

Miller, Conrad, "The Persistent Effect of Temporary Affirmative Action," *American Economic Journal: Applied Economics*, 2017, *9* (3), 152–190.

Miller, Thomas W., *Report of the Alien Property Custodian*, Government Printing Office, Washington, DC, 1922.

Moran, Patrick A.P., "Notes on Continuous Stochastic Phenomena," *Biometrika*, 1950, *37* (1/2), 17–23.

Moser, Petra, "Taste-based discrimination evidence from a shift in ethnic preferences after WWI," *Explorations in Economic History*, 2012, *49* (2), 167–188.

_ **and Alessandra Voena**, "Compulsory Licensing: Evidence from the Trading with the Enemy Act," *American Economic Review*, 2012, *102* (1), 396–427.

Myrdal, Gunnar, *An American Dilemma: The Negro Problem and Modern Democracy*, Harper and Brothers Publishers, New York, NY, 1944.

Olea, Jose L.M. and Carolin Pflueger, "A Robust Test for Weak Instruments," *Journal of Business and Economic Statistics*, 2013, *31* (3), 358–369.

Olmstead, Alan and Paul W. Rhode, "Reshaping the Landscape: The Impact and Diffusion of the Tractor in American Agriculture, 1910-1960," *2001*, Journal of Economic History, *61* (3), 663–698.

Oster, Emily, "Unobservable Selection and Coefficient Stability: Theory and Evidence," *Journal of Business and Economic Statistics, in print. DOI: https://doi.org/10.1080/07350015.2016.1227711*, 2017.

Painter, Gary and David I. Levine, "Family Structure and Youths' Outcomes: Which Correlations are Causal?," *The Journal of Human Resources*, 2000, *35* (3), 524–549.

Pascali, Luigi, "Banks and Development: Jewish Communities in the Italian Renaissance and Current Economic Performance," *Review of Economics and Statistics*, 2016, *98* (1), 140–158.

Pei, Zhuan, Jörn-Steffen Pischke, and Hannes Schwandt, "Poorly Measured Confounders are More Useful on the Left than on the Right," *Journal of Business and Economic Statistics, forthcoming. DOI: https://doi.org/10.1080/07350015.2018.1462710*, 2018.

Roberts, Evan and Alexandra Burda, "Correlates and Consequences of American War Casualties in World War I," *University of Minnesota Working Paper*, 2018, *2018-3.*

Ruggles, Steven, Sarah Flood, Ronald Goeken, Josiah Grover, Erin Meyer, Jose Pacas, and Matthew Sobek, "IPUMS USA," *Version 8.0 [dataset]. Minneapolis, MN: IPUMS, 2018. https://doi.org/10.18128/D010.V8.0*, 2018.

Salisbury, Laura, "Women's income and marriage markets in the United States: Evidence from the Civil War Pension," *The Journal of Economic History*, 2017, *77* (1), 1–37.

Selcer, Richard F., *Civil War America, 1850 to 1875* Almanacs of American life, Facts On File, New York, NY, 2006.

Senne, Jean-Noel, "Death and schooling decisions over the short and long run in rural Madagascar," *Journal of Population Economics*, April 2014, *27* (2), 497–528.

Serneels, Pieter and Marijke Verpoorten, "The Impact of armed conflict on economic performance Evidence from Rwanda.," *Journal of Conflict Resolution*, 2015, *59* (4).

Shatnawi, Dina and Price V. Fishback, "The Impact of World War II on the Demand for Female Workers in Manufacturing," *Journal of Economic History*, 2018, *78* (2), 539–574.

Skocpol, Theda, *Protecting soldiers and mothers: The political origins of social policy in the United States*, Harvard University Press, Cambridge, MA, 1992.

Smith, James P., "Race and Human Capital," *American Economic Review*, 1984, *74* (4), 685–698.

Testa, Patrick, "The Economic Legacy of Expulsion: Lessons from Postwar Czechoslovakia," *SSRN Working Paper*, 2018, *https://ssrn.com/abstract=3251541.*

Turner, Sarah and John Bound, "Closing the Gap or Widening the Divide: The Effects of the G.I. Bill and World War II onthe Educational Outcomes of Black Americans," *Journal of Economic History*, 2003, *63* (1), 145–177.

UNICEF Press Center, "Orphans," https://www.unicef.org/media/media_45279.html 2017. Accessed: 2017-09-15.

United States Department of Commerce, "County and City Data Book [United States]

Consolidated File: County Data, 1947-1977. ICPSR07736-v2," *Ann Arbor, MI: Inter-university Consortium for Political and Social Research [distributor], 2012-09-18. DOI: https://doi.org/10.3886/ICPSR07736.v2*, 2012.

Washington, Thomas, *Officers and Enlisted Men of the United States Navy Who Lost Their Lives during the World War*, Government Printing Office, Washington, DC, 1920.

Wasi, Nada and Aaron Flaaen, "Record Linkage Using Stata: Pre-processing, Linking, and Reviewing Utilities," *Stata Journal*, 2015, *15* (3), 672–697.

Weaver, Robert C., "Negro Employment in the Aircraft Industry," *Quarterly Journal of Economics*, 1945, *59* (4), 597–625.

Wolfbein, Seymour L., "Postwar trends in Negro employment," *Monthly Labor Review*, 1947, *December*, 663–665.

Wright, Gavin, *Sharing the Prize*, Harvard University Press, Cambridge, MA, 2013.