I Z A Institute of Labor Economics

Initiated by Deutsche Post Foundation

DISCUSSION PAPER SERIES

IZA DP No. 12228

The Effect of Social Connectedness on Crime: Evidence from the Great Migration

Bryan A. Stuart
Evan J. Taylor

MARCH 2019

I Z A Institute
of Labor Economics
Initiated by Deutsche Post Foundation

DISCUSSION PAPER SERIES

IZA DP No. 12228

The Effect of Social Connectedness on Crime: Evidence from the Great Migration

Bryan A. Stuart
George Washington University and IZA

Evan J. Taylor
University of Chicago

MARCH 2019

ISSN: 2365-9793

IZA – Institute of Labor Economics

Schaumburg-Lippe-Straße 5–9
53113 Bonn, Germany

Phone: +49-228-3894-0
Email: publications@iza.org

www.iza.org

ABSTRACT

The Effect of Social Connectedness on Crime: Evidence from the Great Migration*

This paper estimates the effect of social connectedness on crime across U.S. cities from 1970 to 2009. Migration networks among African Americans from the South generated variation across destinations in the concentration of migrants from the same birth town. Using this novel source of variation, we find that social connectedness considerably reduces murders, rapes, robberies, assaults, burglaries, and motor vehicle thefts, with a one standard deviation increase in social connectedness reducing murders by 21 percent and motor vehicle thefts by 20 percent. Social connectedness especially reduces murders of adolescents and young adults committed during gang and drug activity.

JEL Classification: K42, N32, R23, Z13

Keywords: crime, social connectedness, Great Migration

Corresponding author:
Bryan A. Stuart
George Washington University
Department of Economics
2115 G Street NW
Washington, DC 20052
USA
E-mail: bastuart@gwu.edu

* Thanks to Martha Bailey, Dan Black, John Bound, Charlie Brown, Eric Chyn, John DiNardo, Alan Griffith, Mike Muller-Smith, Daniel Nagin, Seth Sanders, Jeff Smith, Lowell Taylor, Anthony Yezer, referees, and numerous seminar and conference participants for helpful comments. Thanks to Seth Sanders and Jim Vaupel for facilitating access to the Duke SSA/Medicare data. During work on this project, Stuart was supported in part by an NICHD training grant (T32 HD007339) and an NICHD center grant (R24 HD041028) to the Population Studies Center at the University of Michigan.

1 Introduction

For almost 200 years, the enormous variance of crime rates across space has intrigued social scientists and policy makers (Guerry, 1833; Quetelet, 1835; Weisburd, Bruinsma and Bernasco, 2009). Prior work finds that standard covariates explain less than one-third of the cross-city variation in crime rates, which suggests a potential role for social influences (Glaeser, Sacerdote and Scheinkman, 1996). One possible explanation is peer effects, whereby an individual is more likely to commit crime if his peers commit crime (e.g., Case and Katz, 1991; Damm and Dustmann, 2014). Another explanation is that cities differ in the degree of social connectedness, or the strength of relationships between individuals, including those unlikely to commit crime.

This paper uses a new source of variation in social connectedness to estimate its effect on crime. Migration networks among millions of African Americans who moved out of the U.S. South from 1915 to 1970 generated variation across destinations in the concentration of migrants from the same birth town. For example, consider Beloit, Wisconsin and Middletown, Ohio, two cities similar along many dimensions, including the total number of Southern black migrants that moved there. Around 18 percent of Beloit's black migrants came from Pontotoc, Mississippi, while less than five percent of Middletown's migrants came from any single town. Historical accounts trace the sizable migration from Pontotoc to Beloit to a single influential migrant, John McCord, getting a job in 1914 at a manufacturer in search of workers (Bell, 1933). Furthermore, ethnographic and newspaper accounts suggest that Southern birth town networks translated into strong community ties in the North (Stack, 1974; Associated Press, 1983; Laury, 1986; Crowder and Spencer, 2002; Smith, 2006). Guided by a simple economic model, we proxy for social connectedness using a Herfindahl-Hirschman Index of birth town to destination city population flows for African Americans born in the South from 1916 to 1936, who we observe in the Duke SSA/Medicare dataset. We focus on social connectedness among black migrants because birth town migration networks are especially strong among this group (Stuart and Taylor, 2018) and qualitative and quantitative evidence supports our empirical strategy.

We estimate regressions that relate cross-city differences in crime from 1970 to 2009 to cross-city differences in social connectedness. The historical literature suggests that, conditional on economic and social opportunities, variation in social connectedness stems from idiosyncratic factors, like the right migrant being in the right place at the right time. To exploit this variation, we control for population, manufacturing employment (the sector employing the largest number of African American migrants), and the black population share from 1920 to 1960. Our regressions also include the number of Southern black migrants that live in each city, to adjust for differences in the overall attractiveness of cities to black migrants, and contemporaneous population, land area, and state-by-year fixed effects. City-level crime counts come from FBI Uniform Crime Reports.

We find that social connectedness leads to sizable reductions in crime rates. The elasticity of the crime rate with respect to social connectedness ranges from -0.07 to -0.25 across the seven index crimes of murder, rape, robbery, assault, burglary, larceny, and motor vehicle theft, and is statistically distinguishable from zero for every crime besides larceny. At the mean, a one standard deviation increase in social connectedness leads to a precisely estimated 21 percent decrease in murder, the best measured crime in FBI data. Our estimates imply that replacing Middletown's social connectedness with that of Beloit would decrease murders, robberies, and motor vehicle thefts by 28 to 30 percent. By comparison, the estimates in Chalfin and McCrary (2018) imply that a similar decrease in murders would require a 44 percent increase in the number of police officers.

Because social connectedness arises from individuals' location decisions, a natural concern is whether our estimates reflect causal effects. The validity of our empirical strategy hinges upon whether social connectedness is correlated with unobserved determinants of crime from 1970 to 2009, conditional on the covariates described above. Historical accounts emphasize the importance of migrants who were well-connected in their birth town and who worked for an employer in search of labor in establishing concentrated migration flows from Southern birth towns to Northern cities (Scott, 1920; Bell, 1933; Gottlieb, 1987; Grossman, 1989). It is unlikely that these idiosyncratic factors are correlated with unobserved determinants of crime some 50 years later. These considerations provide qualitative support for our empirical strategy.

We marshal a wide range of quantitative support for our empirical strategy. First, 78 percent of the variation in social connectedness stems from a single birth town to destination city migration flow. This accords with historical accounts emphasizing the importance of idiosyncratic factors. Second, social connectedness is not correlated with murder rates from 1911 to 1916 or 1936 to 1939. This implies that connected groups of migrants did not simply move to low crime cities.

To provide even stronger support, we show that our results are robust to selection on both observed and unobserved variables. Our results are similar when including a battery of additional controls: contemporaneous economic and demographic factors, the number and concentration of white migrants and immigrants, and characteristics of counties from which migrants came. Our results also are robust to controlling for the share of migrants drawn to each destination by birth town migration network. This variable, which we estimate using a structural model of location decisions, controls for a range of unobserved migrant characteristics. Finally, we develop a more general test of selection on unobserved variables. The main threat to identification is that connected groups of migrants moved to cities with low crime rates, and unobserved determinants of crime persisted over time. In the presence of this unobserved selection, controlling for the 1960-1969 crime rate would eliminate the relationship between crime and social connectedness from 1970 to 2009. In contrast, if our empirical strategy is valid, then controlling for the 1960-1969 crime rate would partly attenuate the estimated effect of social connectedness and this attenuation would diminish over time; this is exactly what we find, which rules out the main threat to identification. All of this evidence supports our empirical strategy.

A number of additional results clarify the mechanisms through which social connectedness reduces crime. Social connectedness reduces crimes that are more and less likely to have witnesses, which suggests that an increased probability of detection is not the only operative mechanism. The effect of social connectedness on crime is not driven by effects on employment, education, homeownership, the prevalence of single parents, or crack cocaine use (which emerged in the mid 1980s). Other mechanisms, such as effects on non-cognitive skills, personality traits, and norms, likely matter. We see the largest reductions in murders of adolescents and young adults, committed by acquaintances or strangers, in the course of gang, drug, and other felonious activity.

Furthermore, the effect of social connectedness on crime is persistent: even in the 2000s, when many of the original Southern migrants were no longer alive, crime rates were lower in cities with higher social connectedness. Natural explanations for this persistence include changing norms or skills, which are passed down across generations, and path dependence in crime (Nagin and Paternoster, 1991).

There is widespread interest in the effects of social connectedness and the related concept of social capital.[1] This interest partly stems from the possibility that relationships between individuals can address market failures and generate desirable outcomes that are difficult to accomplish with government policies. However, estimating the effects of social connectedness and social capital has proven challenging. Some of the most influential evidence comes from correlations between outcomes, such as income and crime, and proxies for social capital, like individuals' participation in community organizations, their stated willingness to intervene in the community, and their stated willingness to trust others (Sampson, Raudenbush and Earls, 1997; Putnam, 2000). These proxies for social capital reflect individuals' contemporaneous decision to invest in their community, which raises the concern that these correlations reflect reverse causality or omitted variables bias. As a result, the empirical importance of social capital continues to be debated (Durlauf, 2002). This stands in contrast to several papers that credibly identify peer effects in crime.[2]

We use variation in social connectedness that has the unusual and attractive property of being established decades before we measure outcomes as the result of a known process - birth town migration networks. This facilitates our primary contribution, which is providing new, more credible evidence on the effect of social connectedness on crime. We also contribute to the literature in economics studying how social capital and trust relate to various outcomes, including growth and

[1] Although definitions of social capital vary, Portes (1998) argues that a consensus definition is "the ability of actors to secure benefits by virtue of membership in social networks or other social structures" (p. 6). In discussing social capital, authors typically emphasize the role of trust and reciprocity (Fukuyama, 1995; Putnam, 2000; Bowles and Gintis, 2002). Social connectedness is a broader concept than social capital, as Karlan (2007) also notes.

[2] Although both reflect social influences, peer effects and social connectedness are different concepts. Peer effects arise through interactions between individuals making decisions about whether to commit crime, while social connectedness in our setting is analogous to social cohesion among older individuals that arises because they share the same birth town. Recent research on peer effects in crime includes Ludwig, Duncan and Hirschfield (2001); Kling, Ludwig and Katz (2005); Ludwig and Kling (2007); Bayer, Hjalmarsson and Pozen (2009); Drago and Galbiati (2012); Damm and Dustmann (2014); Billings, Deming and Ross (2016); Corno (2017); Stevenson (2017)

development (Knack and Keefer, 1997; Miguel, Gertler and Levine, 2005), government efficiency and public good provision (La Porta et al., 1997; Alesina, Baqir and Easterly, 1999, 2000), financial development (Guiso, Sapienza and Zingales, 2004), microfinance (Karlan, 2005, 2007; Cassar, Crowley and Wydick, 2007; Feigenberg, Field and Pande, 2013), and intergenerational mobility (Chetty et al., 2014).

More broadly, there is enormous interest in the causes and consequences of criminal activity and incarceration in U.S. cities, especially for African Americans (Freeman, 1999; Neal and Rick, 2014; Evans, Garthwaite and Moore, 2016), and this paper demonstrates the importance of social connectedness in reducing crime. Our results imply that policies which lower social connectedness, including mass incarceration, could have more negative consequences than commonly understood. We also add to the literature on the consequences of the Great Migration for migrants and cities, which has not considered the effects of social connectedness before (e.g., Scroggs, 1917; Smith and Welch, 1989; Margo, 1990; Carrington, Detragiache and Vishwanath, 1996; Collins, 1997; Boustan, 2009, 2010; Hornbeck and Naidu, 2014; Black et al., 2015). Our work complements research on the effects of immigration on crime (for a recent review, see Bell and Machin, 2013). Much of this literature focuses on how crime depends on the number of immigrants and the types of countries from which immigrants originate. By comparison, this paper focuses on the role of social connectedness among a more homogenous group of domestic migrants. This paper draws on Stuart and Taylor (2018), which examines the role of birth town migration networks in more detail.

2 Historical Background on the Great Migration

The Great Migration saw nearly six million African Americans leave the South from 1910 to 1970 (United States Bureau of the Census, 1979).[3] Although migration was concentrated in certain destinations, like Chicago, Detroit, and New York, other cities also experienced dramatic changes. For example, Chicago's black population share increased from 2 to 32 percent from 1910 to 1970,

[3] Parts of this section come from Stuart and Taylor (2018).

while Racine, Wisconsin experienced an increase from 0.3 to 10.5 percent (Gibson and Jung, 2005). Migration out of the South increased from 1910 to 1930, slowed during the Great Depression, and then resumed forcefully from 1940 to 1970.

Several factors contributed to the exodus of African Americans from the South. World War I, which simultaneously increased labor demand among Northern manufacturers and decreased labor supply from European immigrants, helped spark the Great Migration (Scroggs, 1917; Scott, 1920; Gottlieb, 1987; Marks, 1989; Margo, 1990; Jackson, 1991; Collins, 1997; Gregory, 2005). However, many underlying causes existed long before the war, including a less developed Southern economy, the decline in agricultural labor demand due to the boll weevil's destruction of crops (Scott, 1920; Marks, 1989, 1991; Lange, Olmstead and Rhode, 2009), widespread labor market discrimination (Marks, 1991), and racial violence and unequal treatment under Jim Crow laws (Tolnay and Beck, 1991).

Migrants tended to follow paths established by railroad lines: Mississippi-born migrants predominantly moved to Illinois and other Midwestern states, and South Carolina-born migrants predominantly moved to New York and Pennsylvania (Scott, 1920; Carrington, Detragiache and Vishwanath, 1996; Collins, 1997; Boustan, 2010; Black et al., 2015). Labor agents, offering paid transportation, employment, and housing, directed some of the earliest migrants, but their role diminished after the 1920s, and most individuals paid for the relatively expensive train fares themselves (Gottlieb, 1987; Grossman, 1989).[4] African-American newspapers from the largest destinations circulated throughout the South, providing information on life in the North (Gottlieb, 1987; Grossman, 1989).[5]

Historical accounts and recent quantitative work indicate that birth town migration networks strongly affected location decisions during the Great Migration. Initial migrants, most of whom moved in the 1910s, chose their destination primarily in response to economic opportunity. Migrants who worked for an employer in search of labor and were well-connected in their birth town

[4]In 1918, train fare from New Orleans to Chicago cost $22 per person, when Southern farmers' daily wages typically were less than $1 and wages at Southern factories were less than $2.50 (Henri, 1975).

[5]The *Chicago Defender*, perhaps the most prominent African-American newspaper of the time, was read in 1,542 Southern towns and cities in 1919 (Grossman, 1989).

linked family, friends, and acquaintances to jobs and shelter in the North, sometimes leading to persistent birth town to destination city migration flows (Rubin, 1960; Gottlieb, 1987). Describing this behavior shortly after the start of the Great Migration, Scott (1920) wrote,

> "The tendency was to continue along the first definite path. Each member of the vanguard controlled a small group of friends at home, if only the members of his immediate family. Letters sent back, representing that section of the North and giving directions concerning the route best known, easily influenced the next groups to join their friends rather than explore new fields. In fact, it is evident throughout the movement that the most congested points in the North when the migration reached its height, were those favorite cities to which the first group had gone" (p. 69).

Consistent with these accounts, Stuart and Taylor (2018) provide quantitative evidence that birth town migration networks strongly influenced the location decisions of African American migrants from the South.

The experience of John McCord captures many important features of early black migrants' location decisions.[6] Born in Pontotoc, Mississippi, nineteen-year-old McCord traveled in search of higher wages in 1912 to Savannah, Illinois, where a fellow Pontotoc-native connected him with a job. McCord moved to Beloit, Wisconsin in 1914 after hearing of employment opportunities and quickly began work as a janitor at the manufacturer Fairbanks Morse and Company. After two years in Beloit, McCord spoke to his manager about returning home for a vacation. The manager asked McCord to recruit workers during the trip, and McCord returned with 18 unmarried men, all of whom were soon hired. Thus began a persistent flow of African Americans from Pontotoc to Beloit: among individuals born from 1916 to 1936, 14 percent of migrants from Pontotoc lived in Beloit's county in old age (Stuart and Taylor, 2018).[7]

Qualitative evidence documents the impact of social ties among African Americans from the same birth town on life in the North. For example, roughly 1,000 of Erie, Pennsylvania's 11,600

[6]The following paragraph draws on Bell (1933). See also Knowles (2010).
[7]This is 68 times larger than the total share of Mississippi migrants that lived there in old age.

African American residents once lived in Laurel, Alabama, and almost half had family connections there, leading an Erie resident to say, "I'm surrounded by so many Laurelites here, it's like a second home" (Associated Press, 1983). Nearly forty percent of the migrants in Decatur, Illinois came from Brownsville, Tennessee, and Brownsville high school reunions took place in Decatur from the 1980s to 2000s (Laury, 1986; Smith, 2006).[8] As described by a Brownsville native, "Decatur's a little Brownsville, really" (Laury, 1986).

Stack (1974) offers deeper insights into birth town and family social ties among African American migrants. This ethnography is set in an unidentified Midwestern city that lies on a major railroad connecting the South to Chicago. Stack documents "extensive networks of kin and friends" that originated during the Great Migration and continued to grow in the North (p. 28). These networks served many purposes, one of which was childrearing. Households typically contained three generations of kin (not all of whom were first or second degree relatives): "males and females beyond childbearing age, a middle generation of mothers raising their own children or children of close kin, and the children" (p. 123). Beyond the middle and older generations of adults in their household, children were raised by "discipliners . . . allowed to participate in the control of children," "trainers [who] not only discipline but teach moral values and respect for adults," and older children who learned these behaviors from adults (p. 84). This environment clearly could have reduced crime. Motivated by these accounts, we now turn to a systematic analysis of the effect of social connectedness on crime.

3 Conceptual Framework: Crime and Social Connectedness

To inform our empirical work, we next describe the ways in which social connectedness might affect crime. As a starting point, consider younger and older individuals, with the key distinction being that younger individuals might commit crime, while older individuals do not. In deciding whether to commit crime, younger individuals are influenced by non-social factors, peer effects, and social connectedness. The non-social determinants of crime include employment opportuni-

[8] The 40 percent figure comes from the Duke SSA/Medicare dataset, described below.

ties and the degree of policing, among other factors. Peer effects matter because individuals are influenced by the crime decisions of other youth. In our setting, social connectedness is analogous to social cohesion among older individuals that stems from a shared birth town.

Social connectedness could directly affect crime in several ways. For example, it might reduce crime by increasing the probability a criminal is identified and punished, helping individuals find jobs, increasing the stock of cognitive and non-cognitive skills that boost earnings in the non-crime labor market, or promoting anti-crime norms. Alternatively, social connectedness could increase crime by reinforcing unproductive norms or providing trust that facilitates criminal activity, as with the Ku Klux Klan, Mafia, or gangs (Fukuyama, 2000; Putnam, 2000). The work of Stack (1974) suggests that social connectedness might decrease crime in our setting, but this is ultimately an empirical question.

The total effect of social connectedness depends on these direct effects plus peer effects and spillovers. For example, suppose that social connectedness among Southern black migrants directly affects only African American youth (e.g., because of segregation). Social connectedness could indirectly affect non-black youth through peer effects that link the crime decisions of black and non-black youth (e.g., because of gang activity). These peer effects could amplify the direct effect of social connectedness, providing one reason why social connectedness among Southern black migrants could affect city-level crime rates. Another reason is that, although African Americans account for a minority of the population, they account for a majority of the crimes reported to police in the cities we study.

Appendix A contains a simple model that formalizes these forces. We show that if social connectedness reduces the crime rate of African Americans with ties to the South, then social connectedness weakly reduces the crime rate of all groups, as long as the equilibrium is stable and peer effects are non-negative. In this situation, the crime-reducing effect of social connectedness among Southern African Americans is not counteracted by higher crime rates among other groups. A symmetric result holds if social connectedness instead increases the crime rate of African Americans with ties to the South. An additional takeaway from the model is that the Herfindahl-Hirschman

Index emerges as a natural way to measure the degree of social connectedness in each destination city. Guided by this theoretical analysis, we next describe our empirical strategy for estimating the effect of social connectedness on crime. We return to mechanisms below.

4 Data and Empirical Strategy

4.1 Data on Crime, Social Connectedness, and Control Variables

We estimate the effect of social connectedness on crime from 1970 to 2009, since the Great Migration ended around 1970. We measure annual city-level crime counts using FBI Uniform Crime Reports (UCR) data, available from the Inter-university Consortium for Political and Social Research (ICPSR). UCR data contain voluntary monthly reports on the number of offenses reported to police, which we aggregate to the city-year level. We focus on the seven commonly studied index crimes: murder and non-negligent manslaughter ("murder"), forcible rape ("rape"), robbery, assault, burglary, larceny, and motor vehicle theft. Murder is the best measured crime, and robbery and motor vehicle theft are also relatively well-measured (Blumstein, 2000; Tibbetts, 2012). Missing crimes are indistinguishable from true zeros in the UCR. Because cities in our sample almost certainly experience property crime each year, in our main analysis we drop all city-years in which any of the three property crimes (burglary, larceny, and motor vehicle theft) equal zero.[9] We also use annual population estimates from the Census Bureau in the UCR data.

The Duke SSA/Medicare dataset provides the birth town to destination city population flows that underlie our measure of social connectedness. The data contain sex, race, date of birth, date of death (if deceased), and the ZIP code of residence at old age (death or 2001, whichever is earlier) for over 70 million individuals who received Medicare Part B from 1976 to 2001. In addition, the data include a 12-character string with self-reported birth town information from the Social Security Administration NUMIDENT file, which is matched to places, as described in Black et al. (2015). These data capture long-run location decisions, as we only observe individuals' location

[9] At least one property crime equals zero for about 4 percent of city-year observations.

at birth and old age.[10] As a result, our measure of social connectedness for each city does not vary over time. We focus on individuals born from 1916 to 1936 in the former Confederate states, which we refer to as the South. Out-migration rates for the 1916-1936 cohorts are among the highest of all cohorts in the Great Migration (Appendix Figure A.1), and coverage rates decline considerably for earlier and later cohorts (Black et al., 2015). We restrict our main analysis sample to cities with at least 25 Southern-born African American migrants in the Duke dataset to improve the reliability of our estimates.

Census county and city data books provide covariates each decade from 1920 to 2000. In 1920 and 1930, we have county-level covariates. Starting in 1940, we have city-level covariates for cities with at least 25,000 residents. Consequently, our main sample contains cities with at least 25,000 residents from 1940-forward. We limit our sample to cities in the Northeast, Midwest, and West Census regions to focus on the cross-region moves that characterize the Great Migration. Our main analysis sample excludes cities with especially severe measurement errors in the crime data, as described in Appendix B. Appendix Tables A.1 and A.2 provide summary statistics, and Appendix Figure A.2 shows the geographic distribution of our sample. All 224 cities in our sample are in a county with a railroad.[11]

4.2 Estimating the Effect of Social Connectedness on Crime

Our main estimating equation is

$$Y_{k,t} = \exp[\ln(\text{HHI}_k)\delta + \ln(N_k)\theta + X'_{k,t}\beta] + \epsilon_{k,t}, \tag{1}$$

where $Y_{k,t}$ is the number of crimes in city k in year t. The key variable of interest is our proxy for social connectedness among African Americans with ties to the South, $\text{HHI}_k = \sum_j (N_{j,k}/N_k)^2$, where $N_{j,k}$ is the number of migrants from birth town j that live in destination city k, and $N_k \equiv$

[10] As described in detail below, there is relatively little migration for our sample after leaving the South, so our ability to observe individuals' location only in old age is not particularly important.

[11] Of these, 112 cities are linked to the South via one railroad line, 111 cities are linked via two lines, and one city (Lynn, Massachusetts) is linked via three lines.

$\sum_j N_{j,k}$ is the total number of migrants. A Herfindahl-Hirschman Index is a natural way to measure social connectedness, as discussed in Section 3. $X_{k,t}$ is a vector of covariates, including log population and other variables described below, and $\epsilon_{k,t}$ captures unobserved determinants of crime.[12] We use an exponential function in equation (1) because there are no murders for many city-year observations (Appendix Table A.1).[13]

Our proxy for social connectedness varies only across cities, but the number of crimes varies across both cities and years. Instead of collapsing the data into city-level observations, we use equation (1) because our panel of cities is not balanced.[14] We cluster standard errors by city to allow for arbitrary autocorrelation in unobserved determinants of crime.

The key parameter of interest is δ, which we interpret as the elasticity of the crime *rate* with respect to HHI_k, because we control for log population and specify the conditional mean as an exponential function. If social connectedness reduces the city-level crime rate, then $\delta < 0$. We estimate δ using cross-city variation in social connectedness, conditional on the total number of migrants and $X_{k,t}$. The key identifying assumption is

$$\epsilon_{k,t} \perp\!\!\!\perp \text{HHI}_k | (N_k, X_{k,t}), \tag{2}$$

which states that, conditional on the number of migrants living in city k and the vector of control variables, social connectedness is independent of unobserved determinants of crime from 1970 to 2009. Condition (2) allows the total number of migrants, N_k, to depend arbitrarily on observed and unobserved determinants of crime.[15]

As discussed in Section 2, historical accounts suggest that variation in social connectedness,

[12]Because equation (1) includes $\ln(\text{HHI}_k)$, $\ln(N_k)$, and log population, our estimate of δ would be identical if we instead used city population as the denominator of HHI_k.

[13]We estimate the parameters in equation (1) using a Poisson quasi-maximum likelihood estimator. Consistent estimation of (δ, θ, β) requires the assumption that $E[Y_{k,t}|\cdot] = \exp[\ln(\text{HHI}_k)\delta + \ln(N_k)\theta + X'_{k,t}\beta]$, but does not require any restriction on the conditional variance of the error term (Wooldridge, 2002). Given this, we use the representation in equation (1) to facilitate discussion of our assumptions about unobserved determinants of crime.

[14]Estimating regressions on data collapsed to the city-level yields nearly identical results.

[15]Condition (2) does not guarantee identification of the other parameters in equation (1) besides δ. For example, identification of θ requires exogenous variation in the total number of migrants. Boustan (2010) provides one possible strategy for identifying θ, but we do not pursue that here.

conditional on economic and social opportunities, arose largely from idiosyncratic factors like the right migrant being in the right place at the right time. For example, social connectedness in Beloit, Wisconsin stemmed from John McCord's ability to convince 18 individuals from his birth town to come to Beloit in 1916. If John McCord had worked in a different city that offered similar opportunities, these migrants likely would have followed McCord there. If a less influential migrant had worked in Beloit, these migrants likely would not have followed.

We construct HHI_k and N_k for migrants born from 1916 to 1936. The vast majority of these individuals moved out of the South between 1940 and 1960 (Stuart and Taylor, 2018). For this generation, the historical literature highlights the role of previous migrants' location decisions, contemporaneous economic conditions, and moving costs as the main factors determining where individuals moved (Gottlieb, 1987; Grossman, 1989). Moving costs mattered in a specific way: migrants moved along vertical routes established by railroad lines, but along a railroad line, there was little variation in the cost of moving to different destinations.

Our main specification includes several variables that bolster the credibility of condition (2). We control for the log number of Southern black migrants to account for a broad set of factors that, via revealed preference, influenced the attractiveness of destinations to black migrants. We also control for log population, the African American population share, and log manufacturing employment from 1920 to 1960, as these variables could affect the strength of social connectedness and be correlated with later determinants of crime. We control for log population in year t and log land area, so that we also control for log population density. State-by-year fixed effects flexibly account for determinants of crime that vary over time at the state-level, due to changes in economic conditions, government spending, and other factors. Below, we examine the robustness of our results to a battery of additional covariates. We also examine selection on unobserved variables in two distinct ways. The results support the validity of condition (2).

We construct HHI_k and N_k using migrants' location in old age, measured from 1976 to 2001. In principle, migration after 1970, when we first measure crime, could influence HHI_k. If migrants with a higher concentration of friends and family nearby were less likely to out-migrate in response to higher crime shocks, then HHI_k would be larger in cities with greater unobserved determinants of crime, $\epsilon_{k,t}$. This would bias our estimate of δ upwards, making it more difficult to conclude that social connectedness reduces crime. Reassuringly, Table 1 reveals very low migration rates among African Americans who were born in the South from 1916 to 1936 and living in the North, Midwest, and West. Around 90 percent of individuals stayed in the same county for the five-year periods 1955-1960, 1965-1970, 1975-1980, 1985-1990, and 1995-2000. This suggests that our inability to construct HHI_k using migrants' location before 1970 is relatively unimportant.

4.3 Initial Evidence on the Validity of the Empirical Strategy

Before discussing our results, we present initial evidence that supports the validity of our empirical strategy. We first examine whether social connectedness stems from a large concentration of migrants from a single birth town. If idiosyncratic factors drive social connectedness, then a single sending town should account for most of the variation. Consistent with this, Figure 1 shows that 78 percent of the variation in log HHI is explained by the leading term of log HHI, which equals the log squared share of migrants from the top sending town.[16]

Second, we examine whether crime rates in the early twentieth century are correlated with social connectedness. If connected groups of migrants moved to cities with low crime rates, and these low crime rates persisted into the 1970s and beyond, then this would threaten our empirical strategy. Table 2 reports regressions of $\ln(\text{HHI}_k)$ on $\ln(N_k)$ and several covariates. Column 1 shows a negative relationship between log social connectedness and the log number of migrants. This relationship is mechanical: because birth towns are smaller than destination cities, a city must attract migrants from many birth towns to attract a large number of migrants. Column 2 shows that social connectedness is stronger in cities with more manufacturing employment in 1940.[17] The relationship between social connectedness and the African American population share is positive, but not statistically significant. Column 3, which includes the log mean murder rate from 1936

[16] Appendix Table A.3 lists the HHI and top sending town migrant share for each city.

[17] This is consistent with Stuart and Taylor (2018), who find that birth town migration networks brought African Americans to cities with more manufacturing employment.

to 1939, is the most important.[18] The point estimate is small and indistinguishable from zero. As a result, we find no evidence that cities with lower crime rates from 1936 to 1939 attracted more connected groups of migrants.[19]

5 The Effect of Social Connectedness on Crime

5.1 Main Results

Table 3 shows that social connectedness leads to sizable and statistically significant reductions in murder, rape, robbery, assault, burglary, and motor vehicle theft. The table reports estimates of equation (1) for an unbalanced panel of 224 cities.[20] As seen in column 1, the estimated elasticity of the murder rate with respect to HHI is -0.245 (0.064). The estimates for robbery and motor vehicle theft, two other well-measured crimes in the FBI data, are -0.234 (0.045) and -0.227 (0.083). At the mean, these estimates imply that a one standard deviation increase in social connectedness leads to a 21 percent decrease in murders and a 20 percent decrease in robberies and motor vehicle thefts. Summed over the 40 years from 1970 to 2009, a one standard deviation increase in social connectedness leads to 80 fewer murders, 2,529 fewer robberies, and 5,566 fewer motor vehicle thefts per 100,000 residents.

Simple examples help further illustrate the effects of social connectedness on crime. First, consider Middletown, Ohio and Beloit, Wisconsin. These cities are similar in their total number of Southern black migrants, 1980 population, and 1980 black population share, but Beloit's HHI

[18]We digitized FBI UCR data to construct this variable. UCR data are available for 81 cities from 1930 to 1936 (see Fishback, Johnson and Kantor (2010)) and not available before 1930. To examine crime rates before the Great Migration began, we construct log murder rates from 1911 to 1916 using historical mortality statistics for cities with at least 100,000 residents in 1920 (United States Bureau of the Census, 1922). As seen in Appendix Table A.4, we find no statistically or substantively significant relationship between social connectedness and early century murder rates, although power is limited by the smaller sample size. This conclusion holds when we use inverse probability weights to make this sample of cities, which has higher population, comparable to our main analysis sample on observed covariates.

[19]Results in Table 2 are extremely similar if we replace the 1940 covariates with 1950 or 1960 covariates. We use a single year of covariates to transparently describe the cross-sectional patterns that underlie our identification strategy. Because we include covariates from 1920 to 1960 in equation (1), our estimates of δ also control for changes in covariates across decades.

[20]Appendix Table A.5 displays results for all variables.

is over four times as large as Middletown's (0.057 versus 0.014).[21] The estimates in Table 3 imply that replacing Middletown's HHI with that of Beloit would decrease murders, robberies, and motor vehicle thefts by 28-30 percent. By comparison, the estimates in Chalfin and McCrary (2018) imply that a similar decrease in murders would require a 44 percent increase in the number of police officers.[22] The effect of social connectedness is even larger in other examples. HHI in Decatur, Illinois is almost twenty times larger than that of Albany, NY (0.118 versus 0.006).[23] Replacing Albany's HHI with that of Decatur would decrease murders by 63 percent, robberies by 60 percent, and motor vehicle thefts by 58 percent. While these effects are sizable, they are reasonable in light of the tremendous variation in crime rates across cities (Appendix Table A.2).

5.2 Addressing Threats to Empirical Strategy

Selection on Observed Variables

We first examine the robustness of our results to a battery of additional covariates. We focus on the effect of social connectedness on murder, given its importance for welfare and higher measurement quality. Column 1 of Table 4 repeats our baseline specification to facilitate comparisons. In column 2, we control for the contemporaneous share of the population that is African American and female.[24] In column 3, we control for the share of the population age 5-17, 18-64, and 65+, and the share with a high school and college degree. In column 4, we control for log median family income, the unemployment rate, the labor force participation rate, and log manufacturing employment. We add these variables because they could be correlated with social connectedness and unobserved determinants of crime, biasing our estimate of δ. However, social connectedness

[21] For Middletown and Beloit, the number of Southern black migrants is 376 and 407; the 1980 population is 35,207 and 43,719; and the 1980 percent black is 11.3 and 12.0.

[22] Chalfin and McCrary (2018) estimate an elasticity of murder with respect to police of -0.67, almost three times the size of our estimated elasticity for murder.

[23] For Decatur and Albany, the number of Southern black migrants is 760 and 874; the 1980 population is 94,081 and 101,727; and the 1980 percent black is 14.6 and 15.9.

[24] When explaining crime in year t, we use the variable from the decade in which t lies. Demographic and economic variables are available every ten years from 1970 to 2000, except for percent with a high school degree and college degree (not observed in 2000), log median family income (not observed in 2000), and manufacturing share (not observed in 2000). We use the 1990 values of these unavailable variables. Appendix B has additional details on the sample and data.

might affect some of these variables, in which case controlling for them would eliminate the original omitted variables bias while introducing another source of bias.[25] In practice, the coefficient on log HHI changes very little when including these variables.

Next, we control for log HHI and the log number of Southern white migrants and foreign immigrants, using country of origin for the latter group. In principle, social connectedness among these groups could affect crime. We focus on the results for Southern black migrants because previous work documents the importance of birth town migration networks (Stuart and Taylor, 2018) and we are most confident in the validity of condition (2) and the interpretation of HHI as reflecting social connectedness for this group. While we do not assign a causal interpretation to the additional variables, they could be correlated with omitted determinants of crime. As seen in column 5, our results are very similar when including these variables. Column 6 shows that our results also are similar when controlling for racial fragmentation (Alesina and Ferrara, 2000) plus the share of the population that is Hispanic, foreign born, speaks only English at home, and speaks English well or very well.[26]

Another possible concern is that our results reflect the characteristics of migrants' birth place, as opposed to social connectedness. To examine this, we construct migrant-weighted averages of Southern birth county characteristics. We use the 1920 Census to measure the black farm ownership rate, black literacy rate, black population density, percent black, and percent rural. We also measure exposure to Rosenwald schools, which increased educational attainment among African Americans in the South (Aaronson and Mazumder, 2011). Our results are very similar when adding these variables, as seen in column 7.[27]

Columns 2-7 of Table 4 demonstrate that our results are robust to controlling for many additional variables. Next, we provide additional support for our empirical strategy by examining selection on unobserved variables in two distinct ways.

[25] In the language of Angrist and Pischke (2009), this is a "proxy control" situation (p. 66).

[26] Following Alesina and Ferrara (2000), we define racial fragmentation as one minus a HHI of the share of population that is white, black, American Indian, and any other race.

[27] County-level homicide data are available from historical vital statistics reports starting in the late 1940s. Results are extremely similar if we also control for the log mean homicide rate in 1950 in migrants' county of birth.

Selection on Unobserved Variables: Using a Structural Migration Model

One concern is that our estimate of δ is biased by unobserved characteristics of certain migrants - those who chose the same destination as other migrants from their birth town. Census data reveal that Southern black migrants living in a state or metropolitan area with a higher share of migrants from their birth state have less education and income (Appendix Table A.6).[28] As a result, migrants who followed their birth town network likely had less education and earnings capacity than other migrants. This negative selection on education and earnings could generate a positive correlation between HHI_k and $\epsilon_{k,t}$, biasing our estimate of δ upwards, and making it harder to conclude that social connectedness reduces crime (e.g., Lochner and Moretti, 2004).

At the same time, migrants who followed their birth town network might display greater co-operation or other pro-social behaviors. To address this possibility, we estimate a structural model of location decisions, originally developed in Stuart and Taylor (2018), which generates the share of migrants that moved to each destination because of their birth town migration network. When used as a covariate in equation (1), this variable proxies for unobserved characteristics of migrants that chose to follow other migrants from their birth town. We sketch this model in the text, leaving some details to Appendix C.

In the model, migrants from birth town j are indexed on a circle by $i \in \{1, \dots, N_j\}$, where N_j is the total number of migrants. For migrant i, destination k belongs to one of three preference groups: high (H_i), medium (M_i), or low (L_i). The high preference group contains a single destination. In the absence of social interactions between migrants, the destination in H_i is most preferred, and destinations in M_i are preferred over those in L_i.[29] A migrant never moves to a destination in L_i. A migrant chooses a destination in M_i if and only if his neighbor, $i - 1$, chooses the same destination. A migrant chooses a destination in H_i if his neighbor chooses the same destination or

[28]Research on immigrants in the U.S. finds similar patterns of selection (Bartel, 1989; Bauer, Epstein and Gang, 2005; McKenzie and Rapoport, 2010).

[29]The assumption that H_i is a non-empty singleton ensures that migrant i has a well-defined location decision in the absence of social interactions. We could allow H_i to contain many destinations and specify a decision rule among the elements of H_i. This extension would complicate the model without adding new insights.

his neighbor selects a destination in L_i.[30]

Migrants from the same birth town can differ in their preferences over destinations. The probability that destination k is in the high preference group for a migrant from town j is $h_{j,k} \equiv \mathbb{P}[k \in H_i | i \in j]$, and the probability that destination k is in the medium preference group is $m_{j,k} \equiv \mathbb{P}[k \in M_i | i \in j]$. These probabilities arise from expected utility maximization problems solved by migrants. We do not need to specify migrants' utility functions, but expected wages and transportation costs are among the relevant factors. We also do not need to specify why some migrants choose the same destination as their neighbor. For example, neighbors might provide information about employment opportunities, or migrants might value living near friends and family.

The share of migrants from birth town j that chose destination k because of social interactions equals $m_{j,k}$.[31] As a result, the share of migrants that chose this destination because of social interactions is

$$m_k \equiv \sum_j N_{j,k} m_{j,k} / N_k. \tag{3}$$

By including m_k in equation (1), we can assess whether our results stem from social connectedness or unobserved characteristics of migrants that chose the same destination as other migrants from their birth town. Appendix C contains additional details on the model and estimation, which follows Stuart and Taylor (2018).

Conditional on the number of migrants in a destination (N_k) and the share of migrants that chose their destination because of social interactions (m_k), variation in social connectedness (HHI_k) continues to arise from concentrated birth town to destination city population flows. To see this,

[30] This model shares a similar structure as Glaeser, Sacerdote and Scheinkman (1996) in that some agents imitate their neighbors. However, we differ from Glaeser, Sacerdote and Scheinkman (1996) in that we model the interdependence between various destinations (i.e., this is a multinomial choice problem) and allow for more than two types of agents.

[31] The share of migrants from birth town j that chose destination k because of the network is $\mathbb{P}[k \in M_i | D_{i,j,k} = 1]$. By Bayes' theorem, this equals

$$\mathbb{P}[k \in M_i | D_{i,j,k} = 1] = \frac{\mathbb{P}[D_{i,j,k} = 1 | k \in M_i]\,\mathbb{P}[k \in M_i]}{\mathbb{P}[D_{i,j,k} = 1]} = \frac{\mathbb{P}[D_{i,j,k} = 1]\,\mathbb{P}[k \in M_i]}{\mathbb{P}[D_{i,j,k} = 1]} = m_{j,k}$$

because $\mathbb{P}[D_{i,j,k} = 1 | k \in M_i] = \mathbb{P}[D_{i-1,j,k} = 1] = \mathbb{P}[D_{i,j,k} = 1]$.

consider two hypothetical cities that each have 20 migrants, one-fourth of whom chose their destination because of social interactions ($m_k = 0.25$). In the low HHI city, the 20 migrants come from five birth towns. Each town sends four migrants, one of whom moves there because of social interactions. As a result, $\text{HHI}_{\text{Low}} = 0.2$. In the high HHI city, the 20 migrants also come from five birth towns. One town sends 12 migrants, three of whom move there because of social interactions. Two towns each send two migrants, one of whom moves there because of social interactions, and two towns each send two migrants, neither of whom is influenced by social interactions. As a result, $\text{HHI}_{\text{High}} = 0.4$. This example is consistent with Figure 1 in that variation in social connectedness arises from the top sending town.

Column 8 of Table 4 shows that the estimated effect of social connectedness changes little when we control for the share of migrants that chose their destination because of their birth town migration network.[32] We find little evidence that our results are driven by unobserved characteristics of certain migrants. For completeness, column 9 includes all of the additional covariates previously described. The effect of social connectedness is similar in magnitude and statistically significant. As noted above, column 9 is not our preferred specification, because some of the covariates could be affected by social connectedness.

Selection on Unobserved Variables: Using Lagged Crime Rates

Although we have addressed many potential concerns, it is possible that cities with higher social connectedness have lower unobserved determinants of crime, $\epsilon_{k,t}$, for some other reason. For example, if connected groups of migrants moved to cities with low crime rates, and unobserved determinants of crime persisted over time, then our estimate of δ could be biased downwards. We have already presented evidence against this threat by showing that log HHI is not correlated with murder rates from 1936 to 1939 (Table 2) or 1911 to 1916 (Appendix Table A.4).

To provide more direct evidence against this threat, we estimate the effect of social connectedness on crime for each five-year interval from 1970 to 2009 while controlling for the 1960-1969 log

[32] Results are very similar when we use quadratic, cubic, or quartic functions of this variable.

mean crime rate. If our results were driven by connected groups of migrants initially sorting into cities with low crime rates and unobserved determinants of crime persisting over time, then controlling for the 1960-1969 crime rate would eliminate the correlation between social connectedness and crime rates in later years. On the other hand, if condition (2) is valid and there is a true effect of social connectedness, then controlling for the 1960-1969 crime rate will not completely attenuate the estimate of δ; adding this control could attenuate estimates because unobserved determinants of crime are serially correlated, but the attenuation would diminish with time.

To see this more formally, consider a simple log-linear model,

$$\ln(Y_{k,t}) = \ln(\text{HHI}_k)\delta_t + \epsilon_{k,t} \tag{4}$$

$$\epsilon_{k,t} = \epsilon_{k,t-1}\rho + u_{k,t}, \tag{5}$$

where δ_t is the effect of social connectedness on crime in year t, $\rho \in (-1, 1)$ captures serial correlation in unobserved determinants of crime, $\mathbb{E}[u_{k,t}|\epsilon_{k,t-1}] = 0$, and we ignore other covariates. We use a linear model to simplify the analysis, but we have used Monte Carlo simulations to verify that the main conclusion holds with an exponential conditional mean function in equation (4). Because there is little migration after 1960 (Table 1), the main concern is that $\mathbb{C}[\ln(\text{HHI}_k), \epsilon_{k,1960}] < 0$ and $\rho \in (0, 1)$. We could have $\mathbb{C}[\ln(\text{HHI}_k), \epsilon_{k,1960}] < 0$ if connected groups of migrants moved to cities with low unobserved determinants of crime in 1960. If unobserved determinants of crime are positively correlated, then our estimate of δ in 1970 could be biased by this selection.

Consider estimating a regression on 1970-2009 data that controls for the 1960 crime rate,

$$Y_{k,t} = \ln(\text{HHI}_k)d_t + Y_{k,1960}b_t + e_{k,t}. \tag{6}$$

It is straightforward to show that

$$\text{plim } \hat{d}_t = \delta_t - \delta_{1960}\rho^t. \tag{7}$$

Equation (7) shows that controlling for the 1960 crime rate eliminates the selection bias that arises from $\mathbb{C}[\ln(\text{HHI}_k), \epsilon_{k,t}] \neq 0$. However, if there is an effect of social connectedness on crime in 1960 and unobserved determinants of crime are serially correlated, then $\hat{d}_t$ is a biased estimator of δ_t. As t increases, the bias declines as the correlation of $\epsilon_{k,t}$ with $\epsilon_{k,1960}$ declines. If $\hat{d}_t$ approaches the coefficient on $\ln(\text{HHI}_k)$ from the regression that does not control for $Y_{k,1960}$, then our results are not driven by selection of connected groups of migrants into cities with low $\epsilon_{k,1960}$. In contrast, if our results are driven by selection, so that $\delta_t = \delta_{1960} = 0$, then plim $\hat{d}_t = 0$.

Figure 2 plots coefficient estimates from our baseline specification and from a specification that includes the 1960-1969 log mean murder rate. The results are consistent with the prediction in equation (7) if our empirical strategy were valid: there is some attenuation, but this declines over time, and the two sets of point estimates converge. We conclude that our results are not driven by the sorting of connected groups of migrants into low crime cities, but instead reflect the effect of social connectedness on crime. This rules out a large set of threats to our empirical strategy.

Figure 2 also shows that the effects of social connectedness on crime are persistent. Even in the 2000s, when many of the individuals born from 1916 to 1936 were no longer alive, cities with higher social connectedness have lower murder rates. Natural explanations for this persistence include changing norms and non-cognitive skills, which are passed down across generations, and path dependence in criminal and gang activity (Nagin and Paternoster, 1991).

Appendix D describes several additional robustness tests, all of which support our findings.

5.3 Mechanisms

The previous results show that social connectedness reduces city-level crime rates, demonstrate the robustness of this finding, and support the validity of our empirical strategy. So far, we have estimated the overall effect of social connectedness on crime rates. We next present results that clarify our main finding and the underlying mechanisms.

Several potential mechanisms stem from previous theoretical and empirical work. For example, social connectedness could increase the probability that criminals are identified and punished, increase formal labor market opportunities through job referrals, increase homeownership through shared information or resources, increase the number of single-parent households (by providing additional resources for childrearing) or decrease the number of single-parent households (by increasing the costs faced by less-active parents), increase cognitive skills, increase non-cognitive skills such as self-control and forward-looking behavior, affect personality traits such as aggression, or increase pro-social norms. We consider these in turn.

If social connectedness reduces crime by increasing the probability that criminals are identified and punished, we should primarily see reductions in crimes that tend to be witnessed. However, Table 3 shows that social connectedness reduces crimes that are more and less likely to have witnesses: burglary and motor vehicle theft are less likely to have witnesses than robbery or assault, yet the estimates are similar in magnitude for all of these crimes.[33]

We partly examine other mechanisms by adding observed proxies to equation (1). For example, consider the black unemployment rate. If social connectedness increases the probability of employment for young adults and this decreases crime, then including the black unemployment rate in equation (1) would attenuate the coefficient on HHI. However, an attenuated coefficient would not necessarily imply that employment is a mechanism, as the reduction in crime could cause higher employment, or social connectedness could independently cause lower crime and higher employment. An attenuated coefficient only suggests the variable in question as a potential mechanism. On the other hand, if the estimated effect of HHI on crime does not change when adding a variable, this implies it is not a quantitatively important mechanism.

Table 5 explores several possible mechanisms. We focus on years 1980-1989 because African American-specific covariates from the Census are not available for 1960 or 1990, and the crack index from Fryer et al. (2013) is only available from 1980-forward. Panel A presents results for the 222 cities with non-missing African-American specific covariates, and Panel B contains results for the 66 cities for which the Fryer et al. (2013) crack index is also available.

Column 1 contains the estimate of δ from our baseline specification. In column 2, we add

[33]Unlike larceny or motor vehicle theft, a robbery features the use of force or threat of force. Consequently, robberies are witnessed by at least one individual (the victim).

black demographic and economic covariates, including the share of African Americans with a high school and college degree, and the black unemployment rate. Column 3 adds the black homeownership rate, column 4 adds the share of black households headed by a single female, and column 5 adds both of these variables. In column 6 of Panel B, we add the crack index from Fryer et al. (2013), and column 7 adds all variables. Estimates of δ are very similar across the specifications in Panel A. There is more variation in Panel B, which is not surprising given the smaller sample size. The most important takeaway from Panel B is the lack of evidence that crack cocaine use is an important mechanism. In sum, Table 5 indicates that the effect of social connectedness on crime is not mediated by short-run effects on employment, education, homeownership, the prevalence of single parents, or crack cocaine use.[34]

To better understand the underlying mechanisms, we use FBI Supplemental Homicide Reports (SHR), which provide additional details on offenders and victims for murders starting in 1976. Table 6 reports the estimated effect of social connectedness on murders, by victim characteristic and circumstance. The overall effect, in row 1, is very similar to the Table 3 estimate. Social connectedness leads to the largest reduction in murders committed during gang and drug activity. Social connectedness also leads to a sizable reduction in murders committed alongside other felonies. The effects are smaller, but still sizable and precisely estimated, for murders that occur during arguments. For 28 percent of murders, the circumstance is unknown, mainly because the murder is not cleared by arrest. The largest effects are on murders committed with guns and for victims age 10-25.[35] Finally, the effects are larger in magnitude for murders committed by acquaintances and strangers than those committed by romantic partners or family members.[36]

[34] Social connectedness also could affect the community's relationship with police. For example, individuals in more connected destinations might be more or less likely to report crimes to police or cooperate with investigations. Data limitations prevent a full examination of these issues, but the scope for under- or over-reporting of crimes is negligible for murder and relatively small for robbery and motor vehicle theft (Blumstein, 2000; Tibbetts, 2012). Consistent with this, we estimate similar effects on homicides measured in vital statistics data, which do not require a police report (Appendix D).

[35] We also estimate significant reductions in murders of individuals age 36 and older. Most of these victims are killed by younger offenders. Furthermore, social connectedness likely has persistent effects on individuals, by changing norms and skills or by reducing the tendency of crime to beget more crime (Nagin and Paternoster, 1991); these persistent effects would reduce the probability of committing crime in adulthood, thus reducing the chances of being murdered.

[36] Appendix Table A.12 distinguishes between black and non-black victims. Results are similar for both groups,

To further examine mechanisms, Table 7 reports the effect of social connectedness by offender race, victim race, and circumstance. Column 2 reports the effect of social connectedness on murders committed by African Americans. While African Americans make up 16 percent of the population in our sample, they account for 56 percent of the murders for which offender race is available. African Americans constitute 82 percent of the victims of black murderers. Among this group, social connectedness especially reduces murders that coincide with gang and drug activity and other felonies. Social connectedness leads to a smaller, but still sizable, reduction in murders that occur alongside arguments. Social connectedness also reduces murders of non-black victims by black offenders.[37] We again see particular reductions in murders that coincide with gang and drug activity, other felonies, and arguments. Furthermore, murders of non-black victims are more likely to occur in these circumstances, partly explaining the row 8 coefficient being larger than row 2. Because African Americans are the vast majority of victims, the effect of social connectedness on murders of black victims by black offenders is quantitatively the most important.

Column 4 reports the effect of social connectedness on murders committed by non-black offenders. This reduction is driven primarily by fewer murders of black victims by non-black offenders, especially those associated with gang and drug activity, felonies, and arguments. Social connectedness also reduces murders of non-black victims by non-black offenders; these reductions are concentrated in gang and drug activity and felonies.

Overall, the results in Table 7 are consistent with social connectedness among African Americans from the South having a direct effect on black offenders and an indirect effect on non-black offenders through peer effects and spillovers. The simple framework in Section 3 describes this equilibrium. While most murders are intraracial, the presence of interracial spillovers, as seen in the SHR data and qualitative accounts of drug and turf wars (Block and Block, 1993; Quadracci, 2007; Audi, 2011), means that these spillovers are relevant. As crime can lead to more crime (e.g.,

with the most notable difference being the relationship between victim and offender. For non-black victims, social connectedness has the largest effect on murders committed by strangers. For black victims, social connectedness has the largest effects on murders committed by family, acquaintances, and strangers, with a somewhat smaller effect on murders committed by romantic partners.

[37] While it would be interesting to distinguish non-black victims and offenders by Hispanic origin, SHR data do not identify individuals by both race and Hispanic origin before 2000.

through retaliatory murders), social connectedness can even reduce murders of non-black victims by non-black offenders. Further support for this interpretation comes from the fact that the interracial effects of social connectedness are on murders committed alongside gang, drug, and other felonious activity.

Based on all of these results and prior research (e.g., Stack, 1974; Nagin and Pogarsky, 2004; Heckman, Stixrud and Urzua, 2006; Heckman, Pinto and Savelyev, 2013; Heller et al., 2017; Stevenson, 2017), the most likely mechanisms appear to be non-cognitive skills, personality traits, and anti-crime norms. Prior research suggests that these factors play a large role in interactions that adolescents and young adults have with strangers and acquaintances. For example, non-cognitive skills such as self-control and forward-looking behavior could prevent the escalation of conflicts into violence.

6 Conclusion

This paper estimates the effect of social connectedness on crime across U.S. cities from 1970 to 2009. We use a new source of variation in social connectedness stemming from birth town migration networks among millions of African Americans from the South. A one standard deviation increase in social connectedness leads to a precisely estimated 21 percent decrease in murder and a 20 percent decrease in motor vehicle thefts. We find that social connectedness also leads to sizable and statistically significant reductions in rapes, robberies, assaults, and burglaries. Social connectedness reduces crimes that are more and less likely to have witnesses, which suggests that an increased detection probability is not the only mechanism. The effect of social connectedness on crime does not appear to be mediated by short-run effects on employment, education, homeownership, the prevalence of single parents, or crack cocaine use. Instead, effects on non-cognitive skills, personality traits, and norms are most likely. Social connectedness especially reduces murders of adolescents and young adults committed in the course of gang and drug activity.

Our results suggest that social connectedness, and the related concept of social capital, could help address market failures and generate desirable outcomes that are difficult to accomplish with government policies. The results also suggest that policies which disrupt social networks and communities, such as mass incarceration or the construction of interstate highways in the United States, could have negative consequences that are more severe and long-lasting than previously thought.

References

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

Alesina, Alberto, and Eliana La Ferrara. 2000. "Participation in Heterogeneous Communities." *Quarterly Journal of Economics*, 115(3): 847–904.

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

Alesina, Alberto, Reza Baqir, and William Easterly. 2000. "Redistributive Public Employment." *Journal of Urban Economics*, 48(2): 219–241.

Angrist, Joshua D., and Jörn-Steffen Pischke. 2009. *Mostly Harmless Econometrics: An Empiricist's Companion.* Princeton: Princeton University Press.

Associated Press. 1983. "Blacks in Pennsylvania Town Recall Southern Past." *The Baytown Sun*.

Audi, Tamara. 2011. "Latino Gang Targeted Blacks, U.S. Says." *Wall Street Journal*.

Bartel, Ann P. 1989. "Where do the New U.S. Immigrants Live?" *Journal of Labor Economics*, 7(4): 371–391.

Bauer, Thomas, Gil S. Epstein, and Ira N. Gang. 2005. "Enclaves, Language, and the Location Choice of Migrants." *Journal of Population Economics*, 18(4): 649–662.

Bayer, Patrick, Randi Hjalmarsson, and David Pozen. 2009. "Building Criminal Capital behind Bars: Peer Effects in Juvenile Corrections." *The Quarterly Journal of Economics*, 124(1): 105–147.

Bell, Brian, and Stephen Machin. 2013. "Crime and Immigration: What Do We Know?" In *Lessons from the Economics of Crime: What Reduces Offending?*. ed. Philip J. Cook, Stephen

Machin, Olivier Marie and Giovanni Mastrobuoni, 149–174. MIT Press.

Bell, Velma Fern. 1933. “The Negro in Beloit and Madison, Wisconsin.” Master’s diss. University of Wisconsin.

Billings, Stephen B., David J. Deming, and Stephen L. Ross. 2016. “Partners in Crime: Schools, Neighborhoods and the Formation of Criminal Networks.” National Bureau of Economic Research Working Paper 21962.

Black, Dan A., Seth G. Sanders, Evan J. Taylor, and Lowell J. Taylor. 2015. “The Impact of the Great Migration on Mortality of African Americans: Evidence from the Deep South.” *American Economic Review*, 105(2): 477–503.

Block, Carolyn Rebecca, and Richard Block. 1993. “Street Gang Crime in Chicago.” *National Institute of Justice Research in Brief*.

Blumstein, Alfred. 2000. “Disaggregating the Violence Trends.” In *The Crime Drop in America*. ed. Alfred Blumstein and Joel Wallman, 13–44. New York: Cambridge University Press.

Boustan, Leah Platt. 2009. “Competition in the Promised Land: Black Migration and Racial Wage Convergence in the North, 1940-1970.” *Journal of Economic History*, 69(3): 756–783.

Boustan, Leah Platt. 2010. “Was Postwar Suburbanization ‘White Flight’? Evidence from the Black Migration.” *Quarterly Journal of Economics*, 125(1): 417–443.

Bowles, Samuel, and Herbert Gintis. 2002. “Social Capital and Community Governance.” *Economic Journal*, 112(483): F419–F436.

Carrington, William J., Enrica Detragiache, and Tara Vishwanath. 1996. “Migration with Endogenous Moving Costs.” *American Economic Review*, 86(4): 909–930.

Case, Anne C., and Lawrence F. Katz. 1991. “The Company You Keep: The Effects of Family and Neighborhood on Disadvantaged Youths.” National Bureau of Economic Research Working Paper 3705.

Cassar, Alessandra, Luke Crowley, and Bruce Wydick. 2007. “The Effect of Social Capital on Group Loan Repayment: Evidence from Field Experiments.” *Economic Journal*, 117(517): F85–F106.

Chalfin, Aaron, and Justin McCrary. 2018. "Are U.S. Cities Underpoliced? Theory and Evidence." *The Review of Economics and Statistics*, 100(1): 167–186.

Chetty, Raj, Nathaniel Hendren, Patrick Kline, and Emmanuel Saez. 2014. "Where is the Land of Opportunity: The Geography of Intergenerational Mobility in the United States." *Quarterly Journal of Economics*, 129(4): 1553–1623.

Collins, William J. 1997. "When the Tide Turned: Immigration and the Delay of the Great Black Migration." *Journal of Economic History*, 57(3): 607–632.

Corno, Lucia. 2017. "Homelessness and Crime: Do Your Friends Matter?" *The Economic Journal*, 127(602): 959–995.

Crowder, Carla, and Thomas Spencer. 2002. "Perry County Natives Find Success Far from Home." *The Birmingham News*.

Damm, Anna Piil, and Christian Dustmann. 2014. "Does Growing Up in a High Crime Neighborhood Affect Youth Criminal Behavior?" *American Economic Review*, 104(6): 1806–1832.

Drago, Francesco, and Roberto Galbiati. 2012. "Indirect Effects of a Policy Altering Criminal Behavior: Evidence from the Italian Prison Experiment." *American Economic Journal: Applied Economics*, 4(2): 199–218.

Durlauf, Steven N. 2002. "On the Empirics of Social Capital." *Economic Journal*, 112(483): F459–F479.

Evans, William N., Craig Garthwaite, and Timothy J. Moore. 2016. "The White/Black Educational Gap, Stalled Progress, and the Long-Term Consequences of Crack Cocaine Markets." *Review of Economics and Statistics*, 98(5): 832–847.

Feigenberg, Benjamin, Erica Field, and Rohini Pande. 2013. "The Economic Returns to Social Interaction: Experimental Evidence from Microfinance." *Review of Economic Studies*, 80(4): 1459–1483.

Fishback, Price V., Ryan S. Johnson, and Shawn Kantor. 2010. "Striking at the Roots of Crime: The Impact of Welfare Spending on Crime during the Great Depression." *The Journal of Law and Economics*, 53(4): 715–740.

Freeman, Richard B. 1999. "The Economics of Crime." In *Handbook of Labor Economics*. Vol. 3C, ed. Orley Ashenfelter and David Card, 3529–3571. Amsterdam: North Holland.

Fryer, Jr., Roland G., Paul S. Heaton, Steven D. Levitt, and Kevin M. Murphy. 2013. "Measuring Crack Cocaine and Its Impact." *Economic Inquiry*, 51(3): 1651–1681.

Fukuyama, Francis. 1995. *Trust: The Social Virtues and the Creation of Prosperity.* New York, NY: The Free Press.

Fukuyama, Francis. 2000. "Social Capital and Civil Society." IMF Working Paper 00/74.

Gibson, Campbell, and Kay Jung. 2005. "Historical Census Statistics on Population Totals by Race, 1790 to 1990, and by Hispanic Origin, 1790 to 1990, For Large Cities and Other Urban Places in the United States." *U.S. Census Bureau Population Division Working Paper No. 76*.

Glaeser, Edward L., Bruce Sacerdote, and José A. Scheinkman. 1996. "Crime and Social Interactions." *Quarterly Journal of Economics*, 111(2): 507–548.

Gottlieb, Peter. 1987. *Making Their Own Way: Southern Blacks' Migration to Pittsburgh, 1916-1930.* Urbana: University of Illinois Press.

Gregory, James N. 2005. *The Southern Diaspora: How the Great Migrations of Black and White Southerners Transformed America.* Chapel Hill: University of North Carolina Press.

Grossman, James R. 1989. *Land of Hope: Chicago, Black Southerners, and the Great Migration.* Chicago: University of Chicago Press.

Guerry, André-Michel. 1833. *Essai sur la Statistique Morale de la France.* Paris: Crochard.

Guiso, Luigi, Paola Sapienza, and Luigi Zingales. 2004. "The Role of Social Capital in Financial Development." *American Economic Review*, 94(3): 526–556.

Haines, Michael R., and ICPSR. 2010. "Historical, Demographic, Economic, and Social Data: The United States, 1790-2002. Ann Arbor, MI: ICPSR [distributor]."

Heckman, James J., Jora Stixrud, and Sergio Urzua. 2006. "The Effects of Cognitive and Noncognitive Abilities on Labor Market Outcomes and Social Behavior." *Journal of Labor Economics*, 24(3): 411–482.

Heckman, James, Rodrigo Pinto, and Peter Savelyev. 2013. "Understanding the Mechanisms through Which an Influential Early Childhood Program Boosted Adult Outcomes." *American Economic Review*, 103(6): 2052–86.

Heller, Sara B., Anuj K. Shah, Jonathan Guryan, Jens Ludwig, Sendhil Mullainathan, and Harold A. Pollack. 2017. "Thinking, Fast and Slow? Some Field Experiments to Reduce Crime and Dropout in Chicago." *The Quarterly Journal of Economics*, 132(1): 1–54.

Henri, Florette. 1975. *Black Migration: Movement North, 1900-1920.* New York: Anchor Press/Doubleday.

Hornbeck, Richard, and Suresh Naidu. 2014. "When the Levee Breaks: Black Migration and Economic Development in the American South." *American Economic Review*, 104(3): 963–990.

Jackson, Blyden. 1991. "Introduction: A Street of Dreams." In *Black Exodus: The Great Migration from the American South*. ed. Alferdteen Harrison, xi–xvii. Jackson: University Press of Mississippi.

Karlan, Dean S. 2005. "Using Experimental Economics to Measure Social Capital and Predict Financial Decisions." *American Economic Review*, 95(5): 1688–1699.

Karlan, Dean S. 2007. "Social Connections and Group Banking." *Economic Journal*, 117(517): F52–F84.

Kling, Jeffrey R., Jens Ludwig, and Lawrence F. Katz. 2005. "Neighborhood Effects on Crime for Female and Male Youth: Evidence from a Randomized Housing Voucher Experiment." *Quarterly Journal of Economics*, 120(1): 87–130.

Knack, Stephen, and Philip Keefer. 1997. "Does Social Capital Have an Economic Payoff? A Cross-Country Investigation." *Quarterly Journal of Economics*, 112(4): 1251–1288.

Knowles, Lucas W. 2010. "Beloit, Wisconsin and the Great Migration the Role of Industry, Individuals, and Family in the Founding of Beloit's Black Community 1914 - 1955."

Lange, Fabian, Alan L. Olmstead, and Paul W. Rhode. 2009. "The Impact of the Boll Weevil, 1892-1932." *Journal of Economic History*, 69: 685–718.

La Porta, Rafael, Florencio Lopez de Silanes, Andrei Shleifer, and Robert W. Vishny. 1997. "Trust in Large Organizations." *American Economic Review*, 87(2): 333–338.

Laury, Susan. 1986. "Brownsville Folk Reunite in Decatur." *Herald and Review*.

Lochner, Lance, and Enrico Moretti. 2004. "The Effect of Education on Crime: Evidence from Prison Inmates, Arrests, and Self-Reports." *American Economic Review*, 94(1): 155–189.

Ludwig, Jens, and Jeffrey R. Kling. 2007. "Is Crime Contagious?" *Journal of Law and Economics*, 50(3): 491–518.

Ludwig, Jens, Greg J. Duncan, and Paul Hirschfield. 2001. "Urban Poverty and Juvenile Crime: Evidence from a Randomized Housing-Mobility Experiment." *Quarterly Journal of Economics*, 116(2): 655–679.

Manski, Charles F. 1993. "Identification of Endogenous Social Effects: The Reflection Problem." *Review of Economic Studies*, 60(3): 531–542.

Margo, Robert A. 1990. *Race and Schooling in the South, 1880-1950: An Economic History.* Chicago: The University of Chicago Press.

Marks, Carole. 1989. *Farewell, We're Good and Gone: The Great Black Migration.* Bloomington: Indiana University Press.

Marks, Carole. 1991. "The Social and Economic Life of Southern Blacks During the Migrations." In *Black Exodus: The Great Migration from the American South*. ed. Alferdteen Harrison, 36–50. Jackson: University Press of Mississippi.

McKenzie, David, and Hillel Rapoport. 2010. "Self-Selection Pattners in Mexico-U.S. Migration: The Role of Migration Networks." *Review of Economics and Statistics*, 92(4): 811–821.

Miguel, Edward, Paul Gertler, and David I. Levine. 2005. "Does Social Capital Promote Industrialization? Evidence from a Rapid Industrializer." *Review of Economics and Statistics*, 87(4): 754–762.

Mosher, Clayton J., Terance D. Miethe, and Timothy C. Hart. 2011. *The Mismeasure of Crime.* . 2 ed., Los Angeles: SAGE.

Nagin, Daniel S., and Greg Pogarsky. 2004. "Time and Punishment: Delayed Consequences and Criminal Behavior." *Journal of Quantitative Criminology*, 20(4): 295–317.

Nagin, Daniel S., and Raymond Paternoster. 1991. "On the Relationship of Past to Future Participation in Delinquency." *Criminology*, 29(2): 163–189.

Neal, Derek, and Armin Rick. 2014. "The Prison Boom & The Lack of Black Progress after Smith & Welch."

Portes, Alejandro. 1998. "Social Capital: Its Origins and Applications in Modern Sociology." *Annual Review of Sociology*, 24(1): 1–24.

Putnam, Robert D. 2000. *Bowling Alone: The Collapse and Revival of American Community.* New York, NY: Simon & Schuster.

Quadracci, Mario. 2007. "Gang Wars." *Milwaukee Magazine*.

Quetelet, Adolphe. 1835. *Sur l'Homme et le Developpement De Ses Facultes.* Paris: Bachelier.

Rokaw, William M., James A. Mercy, and Jack C. Smith. 1990. "Comparing Death Certificate Data with FBI Crime Reporting Statistics on U.S. Homicides." *Public Health Reports*, 105(5): 447–455.

Rubin, Morton. 1960. "Migration Patterns of Negroes from a Rural Northeastern Mississippi Community." *Social Forces*, 39(1): 59–66.

Ruggles, Steven, Sarah Flood, Ronald Goeken, Josiah Grover, Erin Meyer, Jose Pacas, and Matthew Sobek. 2019. "IPUMS USA: Version 9.0 [dataset]. Minneapolis, MN: IPUMS."

Sampson, Robert J., Stephen W. Raudenbush, and Felton Earls. 1997. "Neighborhoods and Violent Crime: A Multilevel Study of Collective Efficacy." *Science*, 277(918): 918–924.

Scott, Emmett J. 1920. *Negro Migration During the War.* New York: Oxford University Press.

Scroggs, William O. 1917. "Interstate Migration of Negro Population." *Journal of Political Economy*, 25(10): 1034–1043.

Smith, James P., and Finis Welch. 1989. "Black Economic Progress After Myrdal." *Journal of Economic Literature*, 27(2): 519–564.

Smith, Sheila. 2006. "All-class Reunion Recalls Decatur's Ties to Brownsville, Tenn." *Herald & Review*.

Stack, Carol. 1974. *All our Kin.* New York: Basic Books.

Stevenson, Megan. 2017. "Breaking Bad: Mechanisms of Social Influence and the Path to Criminality in Juvenile Jails." *The Review of Economics and Statistics*, 99(5): 824–838.

Stuart, Bryan A., and Evan J. Taylor. 2018. "Social Interactions and Location Decisions: Evidence from U.S. Mass Migration."

Tibbetts, Stephen G. 2012. *Criminological Theory: The Essentials.* Los Angeles: SAGE Publications.

Tolnay, Stewart E., and E. M. Beck. 1991. "Rethinking the Role of Racial Violence in the Great Migration." In *Black Exodus: The Great Migration from the American South*. ed. Alferdteen Harrison, 20–35. Jackson: University Press of Mississippi.

United States Bureau of the Census. 1922. "Mortality Statistics, 1920." *Twenty-First Annual Report*.

United States Bureau of the Census. 1979. "The Social and Economic Status of the Black Population in the United States, 1790-1978." *Current Population Reports, Special Studies Series P-23 No. 80*.

United States Bureau of the Census. 2008. "County and City Data Book [United States] Consolidated File: City Data, 1944-1977. ICPSR Version. Ann Arbor, MI: ICPSR [distribution]."

United States Bureau of the Census. 2012. "County and City Data Book [United States] Consolidated File: County Data, 1947-1977. ICPSR Version. Ann Arbor, MI: ICPSR [distribution]."

United States Department of Health and Human Services, Centers for Disease Control and Prevention. 2010. "National Center for Health Statistics. Mortality Detail Files, 1968-1991. Ann Arbor, MI: ICPSR [distributor]."

United States Department of Justice, Federal Bureau of Investigation. 2005. "Uniform Crime Reporting Program Data [United States]: Offenses Known and Clearances by Arrest, Various Years. Ann Arbor, MI: ICPSR [distributor]."

United States Department of Justice, Federal Bureau of Investigation. 2006. "Uniform Crime Reporting Program Data [United States]: Supplementary Homicide Reports, Various Years. Ann Arbor, MI: ICPSR [distributor]."

United States Department of Justice, Federal Bureau of Investigation. 2009. "Uniform Crime

Reporting Program Data [United States]: Arrests by Age, Sex, and Race, Various Years. Ann Arbor, MI: ICPSR [distributor]."

Weisburd, David, Gerben J.N. Bruinsma, and Wim Bernasco. 2009. "Units of Analysis in Geographic Criminology: Historical Development, Critical Issues, and Open Questions." In *Putting Crime in its Place: Units of Analysis in Geographic Criminology*. ed. David Weisburd, Gerben J.N. Bruinsma and Wim Bernasco, 3–31. New York: Springer.

Wooldridge, Jeffrey M. 2002. *Econometric Analysis of Cross Section and Panel Data.* Cambridge, MA: MIT Press.

Table 1: Five-Year Migration Rates, Southern Black Migrants Living Outside of the South

	1955-60 (1)	1965-70 (2)	1975-80 (3)	1985-90 (4)	1995-2000 (5)
Percent living in same state	93.1	95.5	96.2	96.0	95.9
Same county	86.4	90.4	93.8	77.2	93.8
Same house	33.0	54.0	72.8	77.2	79.1
Different house	53.4	36.4	21.0	-	14.7
Different county	-	4.3	2.4	-	2.1
Unknown	6.7	0.8	-	18.8	-
Percent living in different state	6.9	4.5	3.8	4.0	4.1
Not in South	4.0	2.8	1.4	1.2	1.0
In South	2.9	1.6	2.4	2.9	3.1

Notes: Sample restricted to African Americans who were born in the South from 1916-1936 and were living in the North, Midwest, or West Census regions five years prior to the census year. The 1990 data do not contain detailed information on within-state moves. The 2000 data contain information on public use microdata areas (PUMAs), which are defined by the Census Bureau and contain at least 100,000 residents, instead of counties. Sources: Ruggles et al. (2019)

Table 2: Key Correlates of Social Connectedness

Dependent variable: Log HHI, Southern black migrants	(1)	(2)	(3)
Log number, Southern black migrants	-0.412	-0.456	-0.452
	(0.020)	(0.031)	(0.033)
Log population, 1940		-0.150	-0.154
		(0.094)	(0.095)
Percent black, 1940		0.577	-0.033
		(1.067)	(1.445)
Log manufacturing employment, 1940		0.255	0.250
		(0.065)	(0.069)
Log mean murder rate, 1936-1939			0.050
			(0.063)
State fixed effects	x	x	x
N (cities)	195	195	195
R2	0.731	0.751	0.752

Notes: Sample contains cities in the North, Midwest, and West Census regions with at least 25,000 residents from 1940 to 2000 for which the mean murder rate from 1936-1939 is available. Log HHI and log number of migrants are measured between 1976 and 2001. Heteroskedastic-robust standard errors in parentheses. Sources: Duke SSA/Medicare data, Haines and ICPSR (2010), United States Bureau of the Census (2008), United States Department of Justice, Federal Bureau of Investigation (2005)

Table 3: The Effect of Social Connectedness on Crime, 1970-2009

	Dependent variable: Number of offenses reported to police						
	Murder	Rape	Robbery	Assault	Burglary	Larceny	Motor Vehicle Theft
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Log HHI, Southern black migrants	-0.245	-0.105	-0.234	-0.221	-0.149	-0.069	-0.227
	(0.064)	(0.048)	(0.045)	(0.047)	(0.032)	(0.043)	(0.083)
Log population and log land area	x	x	x	x	x	x	x
Log number, Southern black migrants	x	x	x	x	x	x	x
1920-1960 covariates	x	x	x	x	x	x	x
State-year fixed effects	x	x	x	x	x	x	x
Pseudo R2	0.823	0.871	0.947	0.914	0.952	0.945	0.935
N (city-years)	8,345	8,345	8,345	8,345	8,345	8,345	8,345
Cities	224	224	224	224	224	224	224

Notes: Table displays estimates of equation (1). 1920-1960 covariates are log population, percent black, and log manufacturing employment. Standard errors, clustered at the city level, are in parentheses. Sources: Duke SSA/Medicare data, Haines and ICPSR (2010), United States Bureau of the Census (2008), United States Department of Justice, Federal Bureau of Investigation (2005)

Table 4: The Effect of Social Connectedness on Murder, 1970-2009, Addressing Threats to Empirical Strategy

	Dependent variable: Number of murders reported to police								
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Log HHI, Southern black migrants	-0.245	-0.239	-0.207	-0.247	-0.256	-0.209	-0.223	-0.208	-0.170
	(0.064)	(0.063)	(0.056)	(0.057)	(0.061)	(0.059)	(0.065)	(0.071)	(0.057)
Log population and log land area	x	x	x	x	x	x	x	x	x
Log number, Southern black migrants	x	x	x	x	x	x	x	x	x
1920-1960 covariates	x	x	x	x	x	x	x	x	x
State-year fixed effects	x	x	x	x	x	x	x	x	x
Percent black and percent female		x							x
Age and education covariates			x						x
Economic covariates				x					x
Log HHI, Southern white migrants					x				x
Log number, Southern white migrants					x				x
Log HHI, immigrants					x				x
Log number, immigrants					x				x
Racial fragmentation, percent Hispanic, percent foreign, English language skills						x			x
Birth county covariates							x		x
Share of Southern black migrants influenced by birth town migration network								x	x
Pseudo R2	0.823	0.825	0.836	0.832	0.827	0.832	0.824	0.823	0.840
N (city-years)	8,345	8,345	8,345	8,345	8,345	8,345	8,345	8,345	8,345
Cities	224	224	224	224	224	224	224	224	224

Notes: Table displays estimates of equation (1). 1920-1960 covariates are log population, percent black, and log manufacturing employment. Age and education covariates are percent age 5-17, 18-64, and 65+, percent with high school degree, and percent with college degree. Economic covariates are log median family income, unemployment rate, labor force participation rate, and log manufacturing employment. Racial fragmentation is one minus an HHI of racial population shares. English language skills are the share of people age 5+ who speak only English at home and the share who speak English well or very well (including those who only speak English). Birth county covariates are migrant-weighted averages of the black farm ownership rate, black literacy rate, black population density, percent black, and percent rural, all measured in the 1920 Census, plus Rosenwald school exposure. Column 8 includes an estimate of the share of migrants that chose their destination because of their birth town migration network. We estimate this variable using a structural model of location decisions, as described in the text. Standard errors, clustered at the city level, are in parentheses. Sources: Aaronson and Mazumder (2011), Duke SSA/Medicare data, Haines and ICPSR (2010), United States Bureau of the Census (2008), United States Department of Justice, Federal Bureau of Investigation (2005)

Table 5: The Effect of Social Connectedness on Murder, 1980-1989, Possible Mechanisms

	Dependent variable: Number of murders reported to police						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Panel A: All cities with African American-specific covariates							
Log HHI, Southern black migrants	-0.232	-0.249	-0.232	-0.244	-0.243		
	(0.063)	(0.065)	(0.061)	(0.064)	(0.059)		
Log population and log land area	x	x	x	x	x		
Log number, Southern black migrants	x	x	x	x	x		
1920-1960 covariates	x	x	x	x	x		
State-year fixed effects	x	x	x	x	x		
Black demographic and economic covariates		x			x		
Black homeownership rate			x		x		
Share of black households headed by single woman				x	x		
Pseudo R2	0.829	0.833	0.829	0.830	0.834		
N (city-years)	2,202	2,202	2,202	2,202	2,202		
Cities	222	222	222	222	222		
Panel B: All cities with African American-specific covariates and crack index							
Log HHI, Southern black migrants	-0.166	-0.177	-0.068	-0.157	-0.095	-0.165	-0.095
	(0.105)	(0.109)	(0.078)	(0.107)	(0.075)	(0.106)	(0.075)
Log population and log land area	x	x	x	x	x	x	x
Log number, Southern black migrants	x	x	x	x	x	x	x
1920-1960 covariates	x	x	x	x	x	x	x
State-year fixed effects	x	x	x	x	x	x	x
Black demographic and economic covariates		x			x		x
Black homeownership rate			x		x		x
Share of black households headed by single woman				x	x		x
Crack index						x	x
Pseudo R2	0.821	0.827	0.824	0.822	0.832	0.822	0.832
N (city-years)	660	660	660	660	660	660	660
Cities	66	66	66	66	66	66	66

Notes: Table displays estimates of equation (1). 1920-1960 covariates are log population, percent black, and log manufacturing employment. Black demographic and economic covariates include percent age 5-17, 18-64, and 65+, percent female, percent with a high school degree, percent with a college degree, and unemployment rate. Crack index is from Fryer et al. (2013). Standard errors, clustered at the city level, are in parentheses. Sources: Duke SSA/Medicare data, Fryer et al. (2013), Haines and ICPSR (2010), United States Bureau of the Census (2008), United States Department of Justice, Federal Bureau of Investigation (2005)

Table 6: The Effect of Social Connectedness on Murder, 1976-2009, By Circumstance and Victim Characteristic

		Share of all victims	Coefficient on Log HHI, Southern black migrants
(1)	Total victims	1.00	-0.287
			(0.071)
	Circumstance		
(2)	Gang and drug activity	0.10	-0.652
			(0.171)
(3)	Felony	0.18	-0.328
			(0.094)
(4)	Argument	0.31	-0.193
			(0.077)
(5)	Other	0.13	-0.151
			(0.078)
(6)	Unknown	0.28	-0.348
			(0.136)
	Weapon		
(7)	Gun	0.63	-0.400
			(0.104)
(8)	Other	0.33	-0.138
			(0.045)
(9)	Unknown	0.04	-0.218
			(0.112)
	Age of victim		
(10)	0-9	0.04	-0.180
			(0.066)
(11)	10-17	0.06	-0.410
			(0.110)
(12)	18-25	0.30	-0.358
			(0.091)
(13)	26-35	0.27	-0.297
			(0.070)
(14)	36+	0.32	-0.223
			(0.064)
	Relationship between victim and offender		
(15)	Romantic partner	0.09	-0.124
			(0.049)
(16)	Family	0.06	-0.118
			(0.070)
(17)	Known, not family	0.30	-0.178
			(0.067)
(18)	Stranger	0.16	-0.284
			(0.101)
(19)	Unknown	0.40	-0.483
			(0.142)

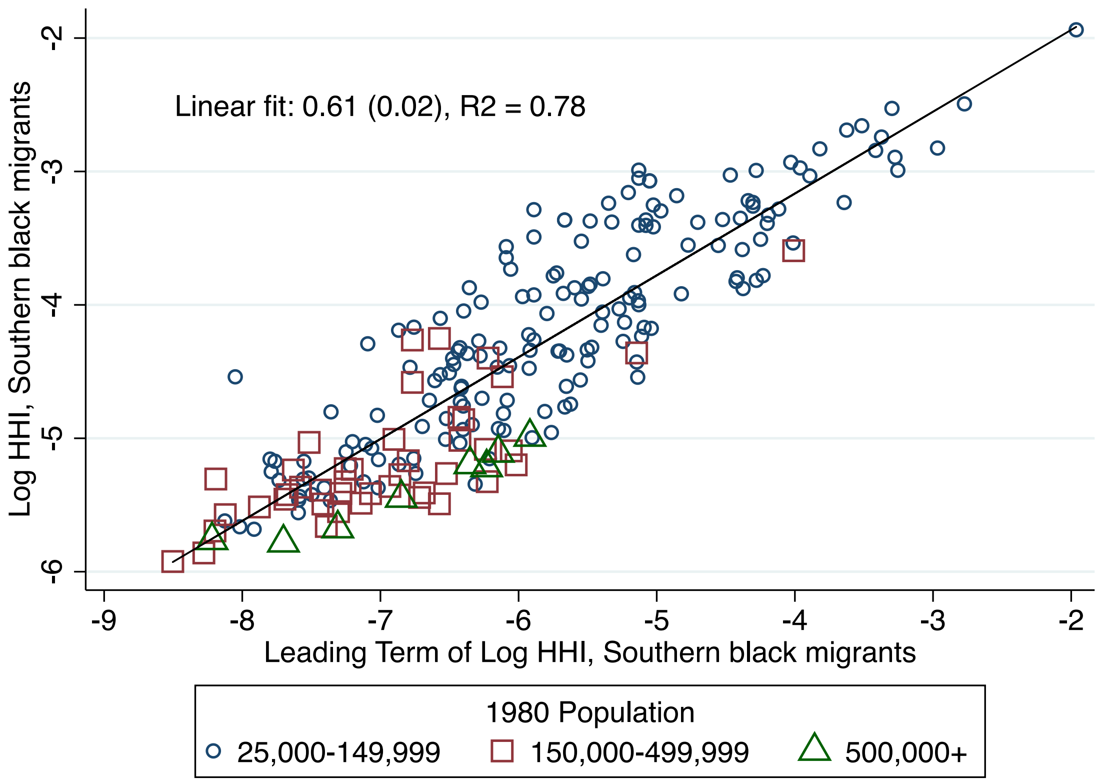
Notes: Table displays estimates of equation (1), using the same specification as Table 3. The dependent variable is the number of murders, by the indicated circumstance or victim characteristic. Standard errors, clustered at the city level, are in parentheses. Sources: Duke SSA/Medicare data, Haines and ICPSR (2010), United States Bureau of the Census (2008), United States Department of Justice, Federal Bureau of Investigation (2006)

Table 7: The Effect of Social Connectedness on Murder, 1976-2009, By Offender Race, Victim Race, and Circumstance

		Black offenders		Non-black offenders	
		Share of all victims (1)	Coefficient on Log HHI, S. black migrants (2)	Share of all victims (3)	Coefficient on Log HHI, S. black migrants (4)
(1)	All victims	0.35	-0.201	0.28	-0.214
			(0.065)		(0.077)
(2)	Black victims	0.29	-0.176	0.03	-0.350
			(0.067)		(0.128)
(3)	Gang and drug	0.03	-0.404	<0.01	-0.902
			(0.187)		(0.362)
(4)	Felony	0.03	-0.380	0.01	-0.334
			(0.110)		(0.240)
(5)	Argument	0.14	-0.168	0.01	-0.436
			(0.085)		(0.123)
(6)	Other	0.05	-0.101	<0.01	-0.051
			(0.088)		(0.180)
(7)	Unknown	0.05	-0.065	<0.01	-0.139
			(0.201)		(0.219)
(8)	Non-black victims	0.06	-0.302	0.25	-0.193
			(0.067)		(0.078)
(9)	Gang and drug	0.01	-0.421	0.02	-0.781
			(0.214)		(0.230)
(10)	Felony	0.02	-0.391	0.04	-0.319
			(0.108)		(0.112)
(11)	Argument	0.02	-0.275	0.11	-0.155
			(0.097)		(0.105)
(12)	Other	0.01	-0.128	0.06	-0.118
			(0.097)		(0.063)
(13)	Unknown	0.01	-0.260	0.03	-0.093
			(0.197)		(0.099)

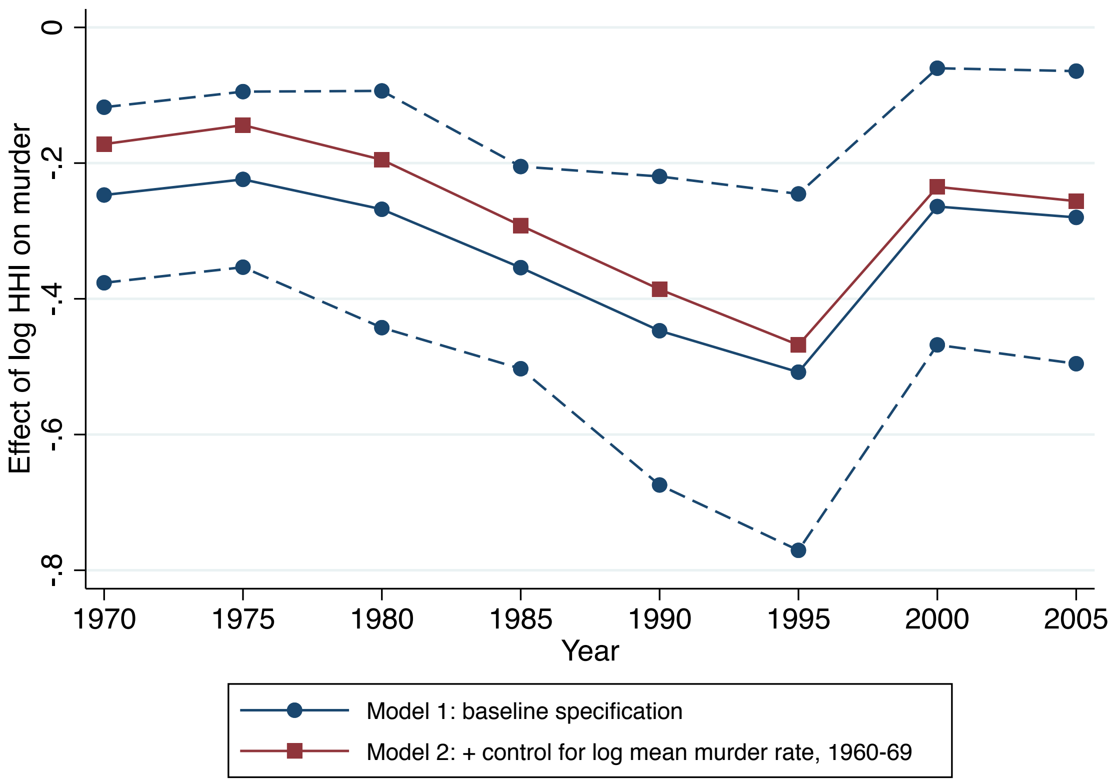
Notes: Table displays estimates of equation (1), using the same specification as Table 3. The dependent variable is the number of murders, by the indicated offender race, victim race, and circumstance. To ensure that all estimates in Table 7 can be interpreted as elasticities of the crime rate, we add log black population to the models in rows 2-7 and constrain the coefficient to equal one. In rows 8-13, we do the same for the log non-black population. We construct an annual black population estimate by combining Census annual population estimates with a linear extrapolation of the African American population share from decennial census data. Standard errors, clustered at the city level, are in parentheses. Sources: Duke SSA/Medicare data, Haines and ICPSR (2010), United States Bureau of the Census (2008), United States Department of Justice, Federal Bureau of Investigation (2006)

Figure 1: The Top Sending Town Accounts for Most of the Variation in Social Connectedness



Notes: The leading term of HHI equals the log squared percent of migrants from the top sending town. Figure contains 224 cities. Source: Duke SSA/Medicare data

Figure 2: The Effect of Social Connectedness on Murder, Assessing Threats to Empirical Strategy by Controlling for 1960-1969 Crime Rate



Notes: Figure shows point estimates and 95-percent confidence intervals from estimating equation (1) separately for years 1970-74, 1975-79, and so on. Model 1 includes the same covariates used in Table 3, and model 2 additionally controls for the log mean crime rate from 1960-69. Sources: Duke SSA/Medicare data, Haines and ICPSR (2010), United States Bureau of the Census (2008), United States Department of Justice, Federal Bureau of Investigation (2005)

Appendices - For Online Publication

A A Simple Model of Crime and Social Connectedness

In this appendix, we use a simple economic model to derive an empirical measure of social connectedness, and we show how the overall effect of social connectedness on crime depends on peer effects and related spillovers. This complements the more intuitive discussion in Section 3.

A.1 Individual Crime Rates

We focus on a single city and characterize individuals by their age and social ties. For simplicity, we consider a static model in which each younger individual makes a single decision about whether to commit crime, while older individuals do not commit crime. Each individual belongs to one of three groups: African Americans with ties to the South ($\tau = s$), African Americans without ties to the South ($\tau = n$), and non-black individuals ($\tau = w$). Older individuals have a tie to the South if they were born there. Younger individuals have a tie to the South if at least one parent, who is an older individual, was born in the South. We index younger individuals by i and older individuals by o.

For a younger individual who is black with ties to the South, we model the probability of committing crime as

$$\mathbb{E}[C_i|\tau_i = s, j_i = j] = \alpha^s + \beta^s\, \mathbb{E}[C_{-i}] + \sum_o \gamma^s_{i,o,j}, \tag{A.1}$$

where $C_i = 1$ if person i commits crime and $C_i = 0$ otherwise, and j_i denotes the birth town of i's parents. Equation (A.1) is a linear approximation to the optimal crime rule from a utility-maximizing model in which the relative payoff of committing crime depends on three factors. First, α^s, which is common to all individuals of type s, captures all non-social determinants of crime (e.g., due to the number of police or employment opportunities). Second, an individual's decision to commit crime depends on the average crime rate among peers, $\mathbb{E}[C_{-i}]$, because of peer effects or other spillovers, such as retaliatory gang violence. Finally, the effect of social connectedness is $\sum_o \gamma^s_{i,o,j}$, where $\gamma^s_{i,o,j}$ is the influence of older individual o on younger individual i. This reduced-form representation captures several possible channels through which social connectedness might affect crime, as discussed in Section 3.

Motivated by the qualitative evidence described in Section 2, we model social connectedness as a function of whether the parents of individual i share a birth town with individual o. In particular, $\gamma^s_{i,o,j} = \gamma^s_H$ if the individuals share a birth town connection, $j_i = j_o$, and $\gamma^s_{i,o,j} = \gamma^s_L$ otherwise. We assume that younger African Americans with ties to the South are only influenced by older African Americans with ties to the South, so that $\gamma^s_{i,o,j} = 0$ if $\tau_i \neq \tau_o$. Given these assumptions, the effect of social connectedness on person i is a weighted average of the high connectedness effect (γ^s_H) and the low connectedness effect (γ^s_L),

$$\sum_o \gamma^s_{i,o,j} = \frac{N^s_{j,0}}{N^s_0}\gamma^s_H + \left(1 - \frac{N^s_{j,0}}{N^s_0}\right)\gamma^s_L, \tag{A.2}$$

where $N^s_{j,0}$ is the number of older individuals of type s from birth town j, and $N^s_0 = \sum_j N^s_{j,0}$ is the total number of older individuals in the city. Through social connectedness, the older generation's migration decisions lead to differences in expected crime rates for younger individuals with ties to different birth towns.

The Herfindahl-Hirschman Index emerges as a natural way to measure social connectedness in this model. In particular, the probability that a randomly chosen African American with ties to the South commits crime is

$$\mathbb{E}[C_i|\tau_i = s] = \alpha^s + \beta^s\,\mathbb{E}[C_{-i}] + \gamma^s_L + (\gamma^s_H - \gamma^s_L)\mathrm{HHI}^s, \tag{A.3}$$

where $\mathrm{HHI}^s \equiv \sum_j (N^s_{j,0}/N^s_0)^2$ is the Herfindahl-Hirschman Index of birth town to destination city population flows for African Americans from the South.[38] HHI^s approximately equals the probability that two randomly chosen members of the older generation share a birth town.[39] The direct effect of social connectedness on the type s crime rate is $\gamma^s_H - \gamma^s_L$. One reasonable case is $\gamma^s_H < \gamma^s_L < 0$, so that older individuals discourage younger individuals from committing crime, and the effect is stronger among individuals who share a birth town connection. Expressions analogous to equation (A.3) exist for African American youth without ties to the South ($\tau = n$) and non-black youth ($\tau = w$).

A.2 City-Level Crime Rates

In the equilibrium of this model, peer effects and spillovers, which we refer to as peer effects for simplicity, can magnify or diminish the effect of social connectedness on crime. We use HHI to measure social connectedness and allow peer effects to differ by the type of peer, leading to the following equilibrium,

$$\bar{C}^s = F^s(\alpha^s, \mathrm{HHI}^s, \bar{C}^s, \bar{C}^n, \bar{C}^w) \tag{A.4}$$
$$\bar{C}^n = F^n(\alpha^n, \mathrm{HHI}^n, \bar{C}^s, \bar{C}^n, \bar{C}^w) \tag{A.5}$$
$$\bar{C}^w = F^w(\alpha^w, \mathrm{HHI}^w, \bar{C}^s, \bar{C}^n, \bar{C}^w), \tag{A.6}$$

where $\bar{C}^\tau$ is the crime rate among younger individuals of type τ, and F^τ characterizes the equilibrium crime rate responses. The equilibrium crime rate vector $(\bar{C}^s, \bar{C}^n, \bar{C}^w)$ is a fixed point of equations (A.4)-(A.6).

We are interested in the effect of social connectedness among African Americans with ties to

[38] In deriving equation (A.3), we assume that each Southern birth town accounts for the same share of individuals in the younger and older generations, so that $N^s_{j,0}/N^s_0 = N^s_{j,1}/N^s_1 \forall j$, where $N^s_{j,1}$ is the number of younger individuals of type s with a connection to birth town j, and $N^s_1 = \sum_j N^s_{j,1}$ is the total number of younger individuals.

[39] The probability that two randomly chosen members of the older generation share a birth town is

$$\mathbb{P}[j_o = j_{o'}] = \sum_j \mathbb{P}[j_o = j_{o'}|j_{o'} = j]\,\mathbb{P}[j_{o'} = j] = \sum_j \left(\frac{N^s_{j,0} - 1}{N^s_0 - 1}\right)\left(\frac{N^s_{j,0}}{N^s_0}\right) \approx \mathrm{HHI}^s.$$

the South, HHI^s, on equilibrium crime rates. Equations (A.4)-(A.6) imply that

$$\frac{d\bar{C}^s}{d\text{HHI}^s} = \frac{\partial F^s}{\partial \text{HHI}^s}\left(\frac{(1-J_{22})(1-J_{33}) - J_{23}J_{32}}{\det(I-J)}\right) \equiv \frac{\partial F^s}{\partial \text{HHI}^s} m^s \tag{A.7}$$

$$\frac{d\bar{C}^n}{d\text{HHI}^s} = \frac{\partial F^s}{\partial \text{HHI}^s}\left(\frac{J_{23}J_{31} + J_{21}(1-J_{33})}{\det(I-J)}\right) \equiv \frac{\partial F^s}{\partial \text{HHI}^s} m^n \tag{A.8}$$

$$\frac{d\bar{C}^w}{d\text{HHI}^s} = \frac{\partial F^s}{\partial \text{HHI}^s}\left(\frac{J_{21}J_{32} + J_{31}(1-J_{22})}{\det(I-J)}\right) \equiv \frac{\partial F^s}{\partial \text{HHI}^s} m^w, \tag{A.9}$$

where I is the 3×3 identity matrix and J, a sub-matrix of the Jacobian of equations (A.4)-(A.6), captures the role of peer effects.[40] Equations (A.7)-(A.9) depend on the direct effect of HHI^s on crime among African Americans with ties to the South, $\partial F^s / \partial \text{HHI}^s$, and peer effect multipliers, m^s, m^n, and m^w. We assume the equilibrium is stable, which essentially means that peer effects are not too large.[41] For example, if $J_{11} \equiv \partial F^s / \partial \bar{C}^s \geq 1$, and there are no cross-group peer effects, then a small increase in the crime rate among type s individuals leads to an equilibrium where all type s individuals commit crime. In a stable equilibrium, a small change in any group's crime rate does not lead to a corner solution.

Our main theoretical result is that if social connectedness reduces the crime rate of African Americans with ties to the South, then social connectedness reduces the crime rate of all groups, as long as the equilibrium is stable and peer effects (i.e., elements of J) are non-negative.

Proposition 1. *$d\bar{C}^s / dHHI^s \leq 0, d\bar{C}^n / dHHI^s \leq 0$, and $d\bar{C}^w / dHHI^s \leq 0$ if $\partial F^s / \partial HHI^s < 0$, the equilibrium is stable, and peer effects are non-negative.*

In a stable equilibrium with non-negative peer effects, the crime-reducing effect of social connectedness among Southern African Americans is not counteracted by higher crime rates among other groups. Hence, equilibrium crime rates of all groups weakly decrease in Southern black social connectedness. With negative cross-group peer effects, the reduction in crime rates among Southern African Americans could lead to higher crime by other groups. A symmetric result holds if social connectedness instead increases the crime rate of African Americans with ties to the South. Proposition 1 is not surprising, and we provide a proof in Appendix A.3.

Because of data limitations, most of our empirical analysis examines the city-level crime rate, $\bar{C}$, which is a weighted average of the three group-specific crime rates,

$$\bar{C} = P^b[P^{s|b}\bar{C}^s + (1 - P^{s|b})\bar{C}^n] + (1 - P^b)\bar{C}^w, \tag{A.10}$$

where P^b is the black population share and $P^{s|b}$ is the share of the black population with ties to the South. Proposition 1 provides sufficient, but not necessary, conditions to ensure that Southern

[40] In particular,

$$J \equiv \begin{bmatrix} \partial F^s / \partial \bar{C}^s & \partial F^s / \partial \bar{C}^n & \partial F^s / \partial \bar{C}^w \\ \partial F^n / \partial \bar{C}^s & \partial F^n / \partial \bar{C}^n & \partial F^n / \partial \bar{C}^w \\ \partial F^w / \partial \bar{C}^s & \partial F^w / \partial \bar{C}^n & \partial F^w / \partial \bar{C}^w \end{bmatrix},$$

and J_{ab} is the (a,b) element of J. m^s is the $(1,1)$ element of $(I-J)^{-1}$, m^n is the $(2,1)$ element, and m^w is the $(3,1)$ element.

[41] The technical assumption underlying stability is that the spectral radius of J is less than one. This condition is analogous to the requirement in linear-in-means models that the slope coefficient on the endogenous peer effect is less than one in absolute value (e.g., Manski, 1993).

black social connectedness decreases the city-level crime rate, $\bar{C}$, when the direct effect is negative. There exist situations in which cross-group peer effects are negative, but an increase in HHI^s still decreases the city-level crime rate.

A.3 Proof of Proposition 1

To prove Proposition 1, we show that the assumptions of a stable equilibrium and non-negative peer effects (i.e., elements of J) imply that the peer effect multipliers m^s, m^n, and m^w are non-negative.

Let $\lambda_1, \lambda_2, \lambda_3$ be the eigenvalues of the 3×3 matrix J. The spectral radius of J is defined as $\rho(J) \equiv \max\{|\lambda_1|, |\lambda_2|, |\lambda_3|\}$. To ensure the equilibrium is stable, we assume that $\rho(J) < 1$.

The on-diagonal elements of J (J_{11}, J_{22}, J_{33}) are less than one in a stable equilibrium. This follows from the facts that the spectral radius is less than one if and only if $\lim_{k\to\infty} J^k = 0$ and $\lim_{k\to\infty} J^k = 0$ implies that the on-diagonal elements of J are less than one.

In a stable equilibrium, we also have that $\det(I - J) > 0$, where I is the 3×3 identity matrix. This follows from our assumption that $\rho(J) < 1$, the fact that $\det(J) = \lambda_1\lambda_2\lambda_3$, and the fact that $\det(J) = \lambda_1\lambda_2\lambda_3$ if and only if $\det(I - J) = (1 - \lambda_1)(1 - \lambda_2)(1 - \lambda_3)$.

It is straightforward to show that

$$\begin{aligned} \det(I - J) &= (1 - J_{11})[(1 - J_{22})(1 - J_{33}) - J_{23}J_{32}] && \text{(A.11)} \\ &\quad - J_{12}[J_{23}J_{31} + J_{21}(1 - J_{33})] - J_{13}[J_{21}J_{32} + J_{31}(1 - J_{22})] \\ &= (1 - J_{11})m^s - J_{12}m^n - J_{13}m^w, && \text{(A.12)} \end{aligned}$$

where the second equality uses the peer effect multipliers defined in equations (A.7)-(A.9). Because the off-diagonal elements of J are non-negative (by assumption) and the on-diagonal elements of J are less than 1 (as implied by a stable equilibrium), we have that m^n and m^w are non-negative. As a result,

$$0 < \det(I - J) \leq (1 - J_{11})m^s. \tag{A.13}$$

Because $J_{11} < 1$, this implies that m^s is non-negative. QED.

B Additional Details on Data and Sample

Our primary measure of crime is annual city-level crime counts from FBI Uniform Crime Reports (UCR) data for 1970-2009. UCR data contain voluntary monthly reports on the number offenses reported to police, which we aggregate to the city-year level.[42] These data are used regularly in the literature and represent the best source of city crime rates. However, the UCR data are not perfect. Missing crimes are indistinguishable from true zeros in the UCR. Because cities in our sample almost certainly experience property crime each year, in our main analysis sample we drop all city-years in which any of the three property crimes (burglary, larceny, and motor vehicle theft) equal zero.

[42] We use Federal Information Processing System (FIPS) place definitions of cities. We follow Chalfin and McCrary (2018) in decreasing the number of murders for year 2001 in New York City by 2,753, the number of victims of the September 11 terrorist attack.

An alternative source of city-level crime counts is the FBI Age-Sex-Race (ASR) data, which report the number of offenses resulting in arrest by age, sex, and race beginning in 1980. The UCR data also report the number of offenses resulting in arrest. In principle, these two data sets, which both rely on reports from police agencies, should lead to similar crime counts. In practice, we found substantial differences between these data sets, especially for large cities.

Appendix Figure A.3 plots the difference between the number of murders in the FBI UCR versus ASR data by city population. For reference, we draw a vertical line at 500,000 residents and horizontal lines at crime differences of -100 and 100. We classify each city into one of two groups, based on whether the city has at least five "severe errors," which we define as years in which the absolute value of the difference in the number of crimes is at least 100. While somewhat arbitrary, this classification identifies the most severe instances of disagreement between the UCR and ASR data.

There are six cities with at least five severe errors: Chicago, Detroit, Los Angeles, Milwaukee, New York, and Philadelphia. Appendix Figure A.4 plots the number of murders from the UCR and ASR data for these cities over time. There does not appear to be a clear explanation for the differences between the two data sets. As a result, we drop these six cities from our main analysis sample. However, as seen in Panel A of Appendix Table A.8, our results are similar when we include these large cities.[43]

We further limit our main analysis sample to cities in the Census city data books that are published each decade. We use covariates from the 1940, 1950 and 1960 Census city data books. There are 409 cities in the U.S. that had at least 25,000 residents in 1940. Of these cities, 313 are not in the South census region and thus can receive long-distance Southern migrants. 230 of the 313 cities received at least 25 migrants in the Duke data. Our main analysis sample results from removing the six cities with severe errors in the UCR data, leaving a total of 224 cities. For nine cities, some covariates are missing in some years (percent black in 1960 is missing for six cities, and the manufacturing employment share in 1940 is missing for three). We impute covariates using adjacent decades in these cases.

We also use FBI Supplemental Homicide Reports (SHR) data. SHR data contain 25 different circumstances, which we collapse into four groups. The circumstances in gang and drug activity are gangland killing, youth gang killing, narcotics laws, and brawl under drugs. The circumstances in felony are rape, robbery, burglary, larceny, auto theft, arson, prostitution, other sex offense, gambling, institution killing, sniper attack, other felony, and suspected felony. The circumstances in argument are brawl under alcohol, argument over money, and other arguments. The circumstances in other are lovers' triangle, abortion, killed by babysitter, and other.

C Estimating a Model of Social Interactions in Location Decisions

This appendix provides additional details on the model of social interactions in location decisions discussed in Section 5.2. The model allows us to estimate the share of migrants that chose their destination because of social interactions. We include this variable in our regressions to examine whether the effect of social connectedness is driven by variation across cities in unobserved characteristics of migrants.

[43] Mosher, Miethe and Hart (2011) discuss measurement error in the UCR data in detail, but do not discuss the discrepancies we have identified between the UCR and ASR data.

C.1 Model of Social Interactions in Location Decisions

In the model, the probability that migrant i moves to destination k given that his neighbor moves there is

$$\rho_{j,k} \equiv \mathbb{P}[D_{i,j,k} = 1 | D_{i-1,j,k} = 1, i \in j] = \mathbb{P}[k \in H_i | i \in j] + \mathbb{P}[k \in M_i | i \in j] \tag{A.14}$$
$$= h_{j,k} + m_{j,k}, \tag{A.15}$$

where $D_{i,j,k}$ equals one if migrant i moves from j to k and zero otherwise.

The probability that destination k is in the medium preference group, conditional on not being in the high preference group, is $\nu_{j,k} \equiv \mathbb{P}[k \in M_i | k \notin H_i, i \in j]$. The conditional probability definition for $\nu_{j,k}$ implies that $m_{j,k} = \nu_{j,k}(1 - h_{j,k})$. We use $\nu_{j,k}$ to derive a simple sequential estimation approach.

In equilibrium, the probability that a randomly chosen migrant i moves from j to k is

$$P_{j,k} \equiv \mathbb{P}[D_{i,j,k} = 1] = \mathbb{P}[D_{i-1,j,k} = 1, k \in H_i] + \mathbb{P}[D_{i-1,j,k} = 1, k \in M_i]$$
$$+ \sum_{k' \neq k} \mathbb{P}[D_{i-1,j,k'} = 1, k \in H_i, k' \in L_i] \tag{A.16}$$
$$= P_{j,k} h_{j,k} + P_{j,k} \nu_{j,k}(1 - h_{j,k}) + \sum_{k' \neq k} P_{j,k'} h_{j,k}(1 - \nu_{j,k'}) \tag{A.17}$$
$$= P_{j,k} \nu_{j,k} + \left(\sum_{k'=1}^{K} P_{j,k'}(1 - \nu_{j,k'}) \right) h_{j,k}. \tag{A.18}$$

The first term on the right hand side of equation (A.16) is the probability that a migrant's neighbor moves to k, and k is in the migrant's high preference group; in this case, social interaction reinforces the migrant's desire to move to k. The second term is the probability that a migrant follows his neighbor to k because of social interactions. The third term is the probability that a migrant resists the pull of social interactions because town k is in the migrant's high preference group and the neighbor's chosen destination is in the migrant's low preference group.

The share of migrants from birth town j living in destination k that chose their destination because of social interactions equals $m_{j,k}$. As a result, the share of migrants in destination k that chose this destination because of social interactions is

$$m_k \equiv \sum_j N_{j,k} m_{j,k} / N_k, \tag{A.19}$$

where $N_{j,k}$ is the number of migrants that moved from j to k. Our goal is to estimate m_k for each destination.

C.2 Estimation

To facilitate estimation, we connect this model to the social interactions (SI) index introduced by Stuart and Taylor (2018). The SI index, $\Delta_{j,k}$, is the expected increase in the number of people from birth town j that move to destination k when an arbitrarily chosen person i is observed to make the same move,

$$\Delta_{j,k} \equiv \mathbb{E}[N_{-i,j,k}|D_{i,j,k}=1] - \mathbb{E}[N_{-i,j,k}|D_{i,j,k}=0], \tag{A.20}$$

where $N_{-i,j,k}$ is the number of people who move from j to k, excluding person i. A positive value of $\Delta_{j,k}$ indicates positive social interactions in moving from j to k, while $\Delta_{j,k}=0$ indicates the absence of social interactions. Stuart and Taylor (2018) show that the SI index can be expressed as

$$\Delta_{j,k} = \frac{C_{j,k}(N_j-1)}{P_{j,k}(1-P_{j,k})}, \tag{A.21}$$

where $C_{j,k}$ is the average covariance of location decisions between migrants from town j, $C_{j,k} \equiv \sum_{i\neq i' \in j} \mathbb{C}[D_{i,j,k}, D_{i',j,k}]/(N_j(N_j-1))$.

The model implies that $C_{j,k}$ equals[44]

$$C_{j,k} = \frac{2P_{j,k}(1-P_{j,k})\sum_{s=1}^{N_j-1}(N_j-s)\left(\frac{\rho_{j,k}-P_{j,k}}{1-P_{j,k}}\right)^s}{N_j(N_j-1)}. \tag{A.22}$$

Substituting equation (A.22) into equation (A.21) and simplifying yields[45]

$$\Delta_{j,k} = \frac{2(\rho_{j,k}-P_{j,k})}{1-\rho_{j,k}}, \tag{A.23}$$

which can be rearranged to show that

$$\rho_{j,k} = \frac{2P_{j,k}+\Delta_{j,k}}{2+\Delta_{j,k}}. \tag{A.24}$$

We follow the approach described in Stuart and Taylor (2018) to estimate $P_{j,k}$ and $\Delta_{j,k}$ using information on migrants' location decisions from the Duke SSA/Medicare data.[46] We then use equation (A.24) to estimate $\rho_{j,k}$ with our estimates of $P_{j,k}$ and $\Delta_{j,k}$.

Equations (A.15) and (A.18), plus the fact that $m_{j,k} = \nu_{j,k}(1-h_{j,k})$, imply that

$$\rho_{j,k} = \nu_{j,k} + \frac{P_{j,k}(1-\nu_{j,k})^2}{\sum_{k'=1}^{K} P_{j,k'}(1-\nu_{j,k'})}. \tag{A.25}$$

We use equation (A.25) to estimate $\nu_j \equiv (\nu_{j,1},\ldots,\nu_{j,K})$ using our estimates of $(P_{j,1},\ldots,P_{j,K}, \rho_{j,1},\ldots,\rho_{j,K})$. We employ a computationally efficient algorithm that leverages the fact that equation (A.25) is a quadratic equation in $\nu_{j,k}$, conditional on $\sum_{k'=1}^{K} P_{j,k'}(1-\nu_{j,k'})$. We initially assume that $\sum_{k'=1}^{K} P_{j,k'}(1-\nu_{j,k'}) = \sum_{k'=1}^{K} P_{j,k'} = 1$, then solve for $\nu_{j,k}$ using the quadratic formula, then construct an updated estimate of $\sum_{k'=1}^{K} P_{j,k'}(1-\nu_{j,k'})$, and then solve again for $\nu_{j,k}$ using the quadratic formula. We require that each estimate of $\nu_{j,k}$ lies in $[0,1]$. This iterated algorithm

[44] This follows from the fact that the covariance of location decisions for individuals i and $i+n$ is $\mathbb{C}[D_{i,j,k}, D_{i+n,j,k}] = P_{j,k}(1-P_{j,k})\left(\frac{\rho_{j,k}-P_{j,k}}{1-P_{j,k}}\right)^n$.

[45] Equation (A.23) results from taking the limit as $N_j \to \infty$, and so relies on N_j being sufficiently large.

[46] We use cross validation to define birth town groups. See Stuart and Taylor (2018) for details.

converges very rapidly in the vast majority of cases.[47]

We use equation (A.18) to estimate $h_{j,k}$ with our estimates of $\rho_{j,k}$ and $\nu_{j,k}$. Finally, we estimate $m_{j,k}$ using the fact that $m_{j,k} = \rho_{j,k} - h_{j,k}$. We use equation (A.19) to estimate our parameter of interest, m_k, using estimates of $m_{j,k}$ and observed migration flows, $N_{j,k}$.

C.3 Results

Appendix Figure A.5 displays a histogram of our estimates of the share of migrants that chose their destination because of social interactions, m_k, for cities in the North, Midwest, and West regions. The estimates range from 0.03 to 0.62. The unweighted average of m_k across cities is 0.26, and the 1980 population weighted average is 0.39.

Appendix Table A.7 examines the relationship between log HHI, the log number of migrants, and m_k. The raw correlation between log HHI and m_k is negative, but when we control for the log number of migrants, log HHI and m_k are positively correlated, as expected. This relationship is similar when including state fixed effects.

Appendix Figure A.6 further describes the relationship between log HHI and m_k. Panel A plots the unconditional relationship between log HHI and m_k, while Panel B plots the relationship conditional on the log number of migrants.[48] When we control for m_k in equation (1), we identify the effect of social connectedness on crime using variation in the vertical dimension of Panel B.

D Additional Robustness Checks

This appendix discusses a number of robustness tests.

Appendix Table A.8 shows that our conclusions are similar when including the six large cities excluded from our main analysis sample because of especially severe measurement error in crime (see Appendix B), estimating negative binomial models, dropping crime outliers, and measuring HHI using birth county to destination city population flows.[49] Results are also similar when we estimate linear models where the dependent variable is the log number of crimes.[50]

Appendix Table A.9 examines robustness to sample restrictions on the number of migrants. Our main analysis sample only includes cities that received at least 25 Southern black migrants according to the Duke data. The results are highly robust to the choice of cutoff.

[47]For 10 birth towns, the algorithm does not converge because our estimates of $P_{j,k}$ and $\rho_{j,k}$ do not yield a real solution to the quadratic formula. We examined the sensitivity of our results to these cases by (1) dropping birth towns for which the algorithm did not converge, (2) estimating $\nu_{j,k}$ and $\sum_{k'=1}^{K} P_{j,k'}(1-\nu_{j,k'})$ as the average of the values in the final four iterations, and (3) forcing $\hat{\nu}_{j,k}$ to equal zero for any (j,k) observation for which the quadratic formula solution does not exist. The motivation for (3) is that our estimates of $P_{j,k}$ and $\rho_{j,k}$ in these 10 cases were consistent with negative values of $\nu_{j,k}$, even though this is not a feasible solution. All three options yielded nearly identical estimates of our variable of interest, m_k. This is not surprising because these 10 birth towns account for a negligible share of the over 5,000 birth towns used to estimate m_k.

[48]In particular, Panel B plots the residuals from regressing log HHI and m_k on the log number of migrants.

[49]We prefer equation (1) over a negative binomial model because it requires fewer assumptions to generate consistent estimates of δ (e.g., Wooldridge, 2002).

[50]From log linear models, the estimate of δ is -0.245 (0.060) for robbery, -0.195 (0.045) for assault, -0.178 (0.040) for burglary, -0.089 (0.038) for larceny, and -0.163 (0.058) for motor vehicle theft. These are very similar to the estimates in Table 3.

Appendix Table A.10 examines robustness to our exclusion of city-year observations in which any property crime (burglary, larceny, or motor vehicle theft) equals zero, which is indistinguishable from missing data in the UCR. Panel A reprints our main estimates from Table 3. In Panel B, we drop city-year observations only if *all* three property crime variables are zero/missing. There are only 13 city-year observations for which one of the three property crimes is zero, but one or both of the other property crime variables is non-zero. This suggests that most of the instances in which any property crime is zero are years in which the city did not report these crimes. In Panel C, we do not drop city-year observations on the basis of zero/missing crime counts. The estimates are extremely similar across panels.

We also examine whether our results are similar when we measure murders using vital statistics data from the National Center for Health Statistics (NCHS). The key potential benefit of these data is that they do not rely on murders being reported to police. The public-use files contain the number of homicides at the county-level from 1970-1988.

Appendix Figure A.7 shows the average annual difference in murders in the NCHS and UCR data for counties in our baseline sample. Positive numbers indicate that, on average, the NCHS data contain more murders than the UCR. For over 90 percent of counties, the average difference is less than 6.5 murders in absolute value. However, there are some counties with larger differences. Most noteworthy are the three counties on the far left, where the mean difference is -44.5 (Franklin, OH, containing the city of Columbus), -36.5 (Alameda, CA, near San Francisco and Oakland), and -17.4 (Summit, OH, containing the city of Akron). For these counties, the UCR has more murders reported than the NCHS. This is somewhat surprising. As discussed by Rokaw, Mercy and Smith (1990), most of the explanations suggest that the UCR should have fewer murders than the NCHS. We have not been able to determine the explanation for these discrepancies.

In our sample, the UCR data contain 98.0 percent of the total number of murders reported in the NCHS data. The correlation between the number of murders in the UCR and NCHS is 0.98 across county-year observations. When we exclude the three counties with the largest differences, the correlation increases to 0.99 (while the UCR data contain 95.9 percent of the murders in the NCHS data).

We have also estimated regressions that use the NCHS number of homicides as the dependent variable. The results are in Appendix Table A.11. Columns 1 and 2 show results for all counties in our baseline sample. The coefficient on log HHI is similar in both regressions, although somewhat smaller when we use the NCHS data. In columns 3 and 4, we exclude the three counties with the largest mean differences in murders (Alameda, CA; Franklin, OH; Summit, OH). The coefficient on log HHI is identical from both data sets. Overall, this evidence indicates that the FBI data do a good job of capturing the number of murders. Given the similarity between the results for murder and other types of crime, we do not believe that our results are driven by differences in crime reporting.

Table A.1: Summary Statistics: Crime and Social Connectedness, 1970-2009

	Mean	SD	First Quartile	Third Quartile	Fraction Zero
Offenses reported to police per 100,000 residents					
Murder	9.4	10.3	3.0	12.0	0.096
Rape	38	33	14	53	0.057
Robbery	313	279	124	411	0.000
Assault	1,575	1,273	589	2,295	0.000
Burglary	1,534	791	958	1,992	0.000
Larceny	3,794	1,899	2,593	4,758	0.000
Motor Vehicle Theft	710	589	311	931	0.000
Population	139,712	165,960	46,815	150,819	-
HHI, Southern black migrants	0.018	0.018	0.006	0.023	-
Log HHI, Southern black migrants	-4.396	0.865	-5.172	-3.761	-
Top sending town share, Southern black migrants	0.062	0.045	0.032	0.076	-
Number, Southern black migrants	1,152	2,156	98	1,212	-

Notes: Each observation is a city-year. HHI and migrant counts are calculated among all individuals born in the former Confederacy states from 1916-1936.

Sources: Duke SSA/Medicare dataset, United States Department of Justice, Federal Bureau of Investigation (2005)

Table A.2: Summary Statistics: Average Crime Rates Per 100,000 Residents

			Percentile				
	Mean	SD	5	25	50	75	95
Murder	8.2	7.8	1.8	3.6	6.1	9.7	23.4
Rape	33.5	20.8	6.7	17.7	29.9	46.0	73.9
Robbery	263.8	209.2	53.3	114.3	198.6	355.1	716.8
Assault	1245.3	690.7	394.1	742.0	1096.5	1595.7	2521.8
Burglary	1370.4	496.1	644.4	998.5	1331.6	1690.8	2212.1
Larceny	3371.6	1301.4	1588.3	2460.3	3332.4	4099.0	5030.8
Motor Vehicle Theft	638.5	424.1	204.9	323.2	463.2	903.6	1414.5

Notes: For each city, we construct an average crime rate across years 1970-2009. Table A.2 reports summary statistics of these average crime rates.
Sources: United States Department of Justice, Federal Bureau of Investigation (2005)

Table A.3: Social Connectedness and Migration Flows, Southern Black Migrants, by Destination City

City	Percent of Migrants from Top Sending Town	HHI	Number of Migrants	Residualized Log HHI
Decatur, IL	37.5	0.144	686	2.64
Fort Wayne, IN	13.5	0.028	1462	1.11
York, PA	22.7	0.059	194	0.97
Troy, NY	16.2	0.039	204	0.96
Erie, PA	13.4	0.029	647	0.90
Beloit, WI	18.1	0.058	342	0.87
Cincinnati, OH	4.1	0.008	6565	0.75
Auburn, NY	25.0	0.083	44	0.74
Garfield, NJ	19.2	0.080	26	0.73
Waterbury, CT	11.2	0.021	713	0.73
Easton, PA	19.6	0.050	112	0.67
Niagara Falls, NY	7.7	0.019	742	0.67
Cleveland, OH	4.6	0.006	18374	0.63
Waterloo, IA	12.0	0.030	435	0.61
Paterson, NJ	7.7	0.011	1866	0.58
Newton, MA	11.1	0.035	45	0.58
Lima, OH	12.1	0.023	572	0.58
Richmond, IN	19.4	0.055	108	0.58
Duluth, MN	11.6	0.038	43	0.53
Aurora, IL	10.9	0.022	384	0.53
Anderson, IN	12.3	0.036	374	0.53
Joplin, MO	16.3	0.068	49	0.52
Inglewood, CA	5.9	0.009	3058	0.50
Middletown, CT	11.2	0.028	143	0.47
Seattle, WA	4.4	0.005	2970	0.46
Santa Barbara, CA	7.7	0.018	117	0.44
Dearborn, MI	8.1	0.033	37	0.43
Oakland, CA	4.8	0.006	11506	0.42
East Chicago, IN	9.0	0.020	858	0.42
Racine, WI	11.8	0.022	773	0.41
Hoboken, NJ	11.6	0.039	43	0.40
Everett, WA	8.0	0.046	25	0.39
Burbank, CA	18.5	0.064	27	0.39
San Francisco, CA	5.2	0.007	6632	0.38
Kalamazoo, MI	7.6	0.012	537	0.37
Hackensack, NJ	6.4	0.012	375	0.36
Indianapolis, IN	4.2	0.006	6922	0.32
Muskegon, MI	7.3	0.014	454	0.31
Cleveland Heights, OH	6.0	0.009	832	0.30
East St. Louis, IL	4.1	0.010	3111	0.29
Warren, OH	8.0	0.015	733	0.29
Evansville, IN	11.0	0.022	264	0.29
Ogden, UT	6.3	0.019	112	0.28
East Cleveland, OH	5.6	0.007	2194	0.28
New Rochelle, NY	5.5	0.008	621	0.27
Alhambra, CA	11.8	0.050	34	0.27

Table A.3: Social Connectedness and Migration Flows, Southern Black Migrants, by Destination City

City	Percent of Migrants from Top Sending Town	HHI	Number of Migrants	Residualized Log HHI
Rockford, IL	5.2	0.011	1295	0.26
Bayonne, NJ	5.6	0.023	124	0.26
Clifton, NJ	8.3	0.037	36	0.24
Ann Arbor, MI	3.5	0.007	370	0.22
Belleville CDP, NJ	4.8	0.026	42	0.22
Malden, MA	5.9	0.035	34	0.21
Beverly Hills, CA	9.5	0.034	42	0.21
Fitchburg, MA	10.7	0.048	28	0.19
Atlantic City, NJ	2.3	0.005	876	0.19
Medford, MA	4.8	0.028	42	0.19
Denver, CO	1.7	0.003	3435	0.19
Pittsburgh, PA	4.9	0.006	3728	0.19
Holyoke, MA	12.8	0.038	47	0.18
Alton, IL	7.2	0.018	335	0.17
Springfield, OH	5.2	0.015	484	0.17
Kansas City, KS	2.2	0.005	1906	0.16
Norwalk, CT	6.2	0.010	530	0.16
Bristol, CT	10.4	0.035	48	0.16
Burlington, IA	7.7	0.050	26	0.15
Clinton, IA	14.8	0.059	27	0.15
Galesburg, IL	10.3	0.029	78	0.14
Hamilton, OH	13.8	0.051	29	0.14
Newport, RI	4.3	0.019	69	0.14
Buffalo, NY	3.5	0.004	6811	0.14
Passaic, NJ	6.5	0.013	447	0.14
Pittsfield, MA	7.7	0.033	52	0.14
Lowell, MA	5.3	0.037	38	0.14
Topeka, KS	2.7	0.007	403	0.13
St. Louis, MO	3.3	0.006	11317	0.13
Flint, MI	2.8	0.004	4758	0.12
Lafayette, IN	14.3	0.048	35	0.12
Akron, OH	4.4	0.006	3669	0.12
Sacramento, CA	3.7	0.004	3317	0.11
Grand Rapids, MI	3.2	0.007	1482	0.10
White Plains, NY	4.1	0.009	368	0.10
Port Huron, MI	5.5	0.017	145	0.09
Newburgh, NY	9.4	0.015	384	0.09
Hartford, CT	4.7	0.007	1525	0.08
Woonsocket, RI	17.2	0.070	29	0.07
Zanesville, OH	13.3	0.053	30	0.07
Bakersfield, CA	4.7	0.008	488	0.06
Yakima, WA	3.2	0.015	93	0.06
University City, MO	3.4	0.006	1086	0.05
Cedar Rapids, IA	9.2	0.029	87	0.05
Elyria, OH	5.7	0.013	470	0.04
Glendale, CA	12.2	0.034	49	0.04

Table A.3: Social Connectedness and Migration Flows, Southern Black Migrants, by Destination City

City	Percent of Migrants from Top Sending Town	HHI	Number of Migrants	Residualized Log HHI
Joliet, IL	3.8	0.008	965	0.04
Scranton, PA	11.4	0.040	35	0.03
Massillon, OH	5.9	0.020	205	0.03
San Bernardino, CA	2.2	0.004	1291	0.03
Lincoln, NE	3.4	0.014	118	0.03
Tucson, AZ	1.7	0.004	929	0.02
San Diego, CA	2.6	0.003	4173	0.02
West Orange CDP, NJ	1.8	0.011	112	0.00
Davenport, IA	7.4	0.019	215	0.00
Albuquerque, NM	2.4	0.005	576	0.00
Portland, OR	2.3	0.005	2078	0.00
Long Beach, CA	3.8	0.005	2112	-0.01
Jersey City, NJ	2.2	0.004	2645	-0.01
Bloomfield CDP, NJ	2.9	0.014	104	-0.01
Phoenix, AZ	3.3	0.004	1996	-0.02
Omaha, NE	2.1	0.004	1918	-0.03
Colorado Springs, CO	2.5	0.004	734	-0.05
Columbus, OH	2.1	0.003	5174	-0.06
Riverside, CA	3.2	0.005	926	-0.06
Evanston, IL	2.9	0.006	734	-0.06
New Haven, CT	4.5	0.006	1696	-0.06
Middletown, OH	5.3	0.014	380	-0.07
Williamsport, PA	8.1	0.039	37	-0.07
New Castle, PA	5.1	0.019	99	-0.07
Belleville, IL	4.3	0.014	116	-0.07
St. Joseph, MO	7.7	0.047	26	-0.07
Binghamton, NY	6.3	0.030	48	-0.08
New Bedford, MA	7.9	0.033	38	-0.08
Dayton, OH	2.7	0.005	4107	-0.09
Bloomington, IL	6.5	0.021	93	-0.09
Portsmouth, OH	7.9	0.035	38	-0.09
Pasadena, CA	3.8	0.007	1177	-0.09
Mount Vernon, NY	2.3	0.005	1502	-0.09
Perth Amboy, NJ	6.7	0.016	149	-0.10
Rochester, NY	3.1	0.005	3136	-0.11
East Orange, NJ	1.9	0.003	2720	-0.12
Jamestown, NY	7.0	0.034	43	-0.12
Trenton, NJ	4.3	0.005	2068	-0.13
Pueblo, CO	3.7	0.010	136	-0.13
Newark, NJ	1.6	0.003	7905	-0.14
Fresno, CA	2.7	0.005	1655	-0.14
South Gate, CA	8.8	0.042	34	-0.14
Berkeley, CA	5.2	0.007	1874	-0.14
Spokane, WA	3.4	0.010	177	-0.15
Gary, IN	3.5	0.004	7149	-0.15
Boston, MA	1.6	0.003	4142	-0.15

Table A.3: Social Connectedness and Migration Flows, Southern Black Migrants, by Destination City

City	Percent of Migrants from Top Sending Town	HHI	Number of Migrants	Residualized Log HHI
Hammond, IN	3.6	0.009	416	-0.16
Bay City, MI	6.5	0.034	31	-0.16
Norwood, OH	6.1	0.021	82	-0.16
St. Paul, MN	4.0	0.007	596	-0.16
Norristown borough, PA	5.2	0.013	289	-0.17
Albany, NY	2.0	0.006	790	-0.17
Wilkes-Barre, PA	5.9	0.035	34	-0.17
Irvington CDP, NJ	2.2	0.004	1248	-0.17
Sioux City, IA	4.8	0.024	62	-0.18
Jackson, MI	4.4	0.009	321	-0.18
Marion, OH	5.7	0.023	70	-0.19
Meriden, CT	4.1	0.017	98	-0.19
Santa Ana, CA	4.0	0.008	299	-0.20
Terre Haute, IN	6.8	0.022	74	-0.21
Providence, RI	7.6	0.013	524	-0.21
Chester, PA	2.7	0.005	1144	-0.21
Moline, IL	8.0	0.046	25	-0.21
Lynn, MA	5.3	0.020	76	-0.22
Michigan City, IN	5.9	0.013	388	-0.22
Cambridge, MA	4.0	0.013	125	-0.22
Rome, NY	3.8	0.017	80	-0.23
West Allis, WI	7.4	0.043	27	-0.23
Lancaster, PA	7.6	0.020	132	-0.24
Danville, IL	4.1	0.013	266	-0.25
Peoria, IL	3.0	0.008	1038	-0.25
Utica, NY	5.9	0.010	321	-0.25
Montclair CDP, NJ	2.0	0.005	590	-0.25
Stamford, CT	2.9	0.006	581	-0.26
Reading, PA	7.8	0.014	296	-0.27
New London, CT	2.5	0.008	198	-0.27
Youngstown, OH	3.4	0.005	2360	-0.27
Mansfield, OH	7.3	0.016	219	-0.27
Lansing, MI	2.5	0.005	974	-0.28
Brockton, MA	3.8	0.011	160	-0.28
Salt Lake City, UT	3.7	0.014	107	-0.28
Elizabeth, NJ	3.0	0.006	767	-0.29
Cicero town, IL	5.3	0.030	38	-0.29
Wichita, KS	2.7	0.005	941	-0.29
Oak Park village, IL	4.1	0.007	442	-0.30
Kansas City, MO	2.6	0.004	5818	-0.31
Maywood village, IL	3.2	0.006	1579	-0.31
Newark, OH	4.2	0.021	72	-0.31
Worcester, MA	4.5	0.012	157	-0.32
New Britain, CT	3.4	0.011	238	-0.32
Springfield, MO	3.4	0.015	88	-0.32
Battle Creek, MI	4.6	0.007	605	-0.32

Table A.3: Social Connectedness and Migration Flows, Southern Black Migrants, by Destination City

City	Percent of Migrants from Top Sending Town	HHI	Number of Migrants	Residualized Log HHI
Yonkers, NY	1.7	0.005	721	-0.32
Saginaw, MI	2.5	0.004	2223	-0.32
Elmira, NY	4.0	0.013	149	-0.33
New Albany, IN	7.5	0.027	53	-0.34
Orange CDP, NJ	3.0	0.005	868	-0.34
San Jose, CA	4.4	0.005	992	-0.35
Springfield, MA	2.9	0.004	1270	-0.36
Bethlehem, PA	6.4	0.021	78	-0.36
Muncie, IN	5.8	0.013	329	-0.37
Pontiac, MI	2.8	0.005	1513	-0.37
Minneapolis, MN	1.9	0.004	1129	-0.37
Marion, IN	7.8	0.015	204	-0.37
Stockton, CA	2.3	0.004	1464	-0.38
Springfield, IL	4.0	0.009	372	-0.38
Syracuse, NY	2.6	0.004	1414	-0.38
Huntington Park, CA	6.9	0.039	29	-0.38
Santa Monica, CA	4.6	0.011	217	-0.39
Madison, WI	4.7	0.011	213	-0.39
Poughkeepsie, NY	4.8	0.009	293	-0.42
Toledo, OH	2.5	0.003	3786	-0.42
Plainfield, NJ	1.8	0.003	1212	-0.43
Steubenville, OH	6.7	0.017	163	-0.43
Camden, NJ	1.7	0.004	1454	-0.43
South Bend, IN	4.0	0.007	1391	-0.44
Lorain, OH	4.2	0.007	570	-0.45
Schenectady, NY	3.9	0.012	204	-0.45
Elgin, IL	4.8	0.012	166	-0.48
Harrisburg, PA	2.1	0.005	717	-0.49
Elkhart, IN	4.3	0.012	277	-0.49
Bridgeport, CT	2.4	0.004	1358	-0.50
Canton, OH	2.7	0.006	825	-0.57
Alameda, CA	3.9	0.011	129	-0.59
Kokomo, IN	4.7	0.013	172	-0.61
Kenosha, WI	6.4	0.013	188	-0.66
Rock Island, IL	4.0	0.010	272	-0.68
New Brunswick, NJ	2.1	0.006	388	-0.70
Waukegan, IL	2.3	0.006	699	-0.71
Allentown, PA	3.9	0.012	127	-0.73
Tacoma, WA	1.4	0.003	983	-0.77
Des Moines, IA	2.3	0.007	300	-0.78

Notes: This table shows cities ranked by residuals of log HHI against our base covariates in a linear regression.
Sources: Duke SSA/Medicare data, Haines and ICPSR (2010), United States Bureau of the Census (2008)

Table A.4: Key Correlates of Social Connectedness, with 1911-1916 Murder Rate

	Dependent variable: Log HHI, Southern black migrants					
	(1)	(2)	(3)	(4)	(5)	(6)
Log number, Southern black migrants	-0.232	-0.515	-0.516	-0.212	-0.477	-0.478
	(0.074)	(0.117)	(0.120)	(0.074)	(0.111)	(0.114)
Log population, 1940		0.240	0.243		0.252	0.264
		(0.201)	(0.204)		(0.192)	(0.207)
Percent black, 1940		-2.610	-2.695		-3.732	-3.898
		(4.740)	(4.916)		(4.224)	(4.335)
Log manufacturing employment, 1940		0.306	0.308		0.275	0.273
		(0.205)	(0.210)		(0.211)	(0.213)
Log mean murder rate, 1911-1916			0.027			0.049
			(0.262)			(0.253)
State fixed effects	x	x	x	x	x	x
N (cities)	46	46	46	46	46	46
R2	0.67	0.80	0.80	0.72	0.84	0.84
Inverse probability weighted				x	x	x

Notes: The sample contains cities in the North, Midwest, and West Census regions with at least 100,000 residents in 1920. We exclude murder rates based on less than five deaths in constructing the mean murder rate from 1911-1916. In columns 4-6, we use inverse probability weights (IPWs) because the sample of cities for which we observe murder rates from 1911-1916 differs on observed characteristics from our main analysis sample. We construct IPWs using fitted values from a logit model, where the dependent variable is an indicator for a city having murder rate data for at least one year from 1911-1916, and the explanatory variables are log population and log land area in 1980, plus the 1920-1960 covariates used in Table 3. Heteroskedastic-robust standard errors in parentheses.

Sources: United States Bureau of the Census (1922, p. 64-65), Duke SSA/Medicare data, United States Bureau of the Census (2008)

Table A.5: The Effect of Social Connectedness on Crime, 1970-2009, Results for All Explanatory Variables

	Dependent variable: Number of offenses reported to police						
	Murder (1)	Rape (2)	Robbery (3)	Assault (4)	Burglary (5)	Larceny (6)	Motor Vehicle Theft (7)
Log HHI, Southern black migrants	-0.245	-0.105	-0.234	-0.221	-0.149	-0.069	-0.227
	(0.064)	(0.048)	(0.045)	(0.047)	(0.032)	(0.043)	(0.083)
Log number, Southern black migrants	0.188	0.069	0.191	0.046	0.056	0.026	0.046
	(0.047)	(0.044)	(0.035)	(0.038)	(0.026)	(0.030)	(0.048)
Log population	0.650	0.955	0.996	0.686	0.712	0.736	1.106
	(0.138)	(0.120)	(0.137)	(0.140)	(0.100)	(0.124)	(0.154)
Log land area	-0.114	0.032	-0.293	-0.000	0.009	-0.038	-0.076
	(0.074)	(0.047)	(0.060)	(0.054)	(0.037)	(0.039)	(0.061)
Log population, 1920 (county)	-0.390	-0.032	-0.410	-0.113	-0.222	0.156	-0.472
	(0.377)	(0.233)	(0.305)	(0.270)	(0.184)	(0.172)	(0.488)
Percent black, 1920 (county)	-2.457	11.846	6.112	0.014	2.592	3.661	-12.675
	(3.885)	(3.527)	(3.471)	(5.037)	(2.031)	(3.387)	(3.901)
Log manufacturing employment, 1920 (county)	-0.244	-0.289	-0.308	-0.167	-0.038	-0.183	0.220
	(0.149)	(0.094)	(0.141)	(0.124)	(0.077)	(0.143)	(0.163)
Log population, 1930 (county)	0.154	-0.324	0.333	-0.336	-0.018	-0.318	0.605
	(0.312)	(0.204)	(0.260)	(0.207)	(0.159)	(0.142)	(0.371)
Percent black, 1930 (county)	2.977	-10.287	-4.309	-0.831	-3.234	-3.356	10.663
	(3.687)	(3.688)	(3.525)	(5.569)	(1.780)	(3.098)	(3.968)
Log manufacturing employment, 1930 (county)	0.282	0.487	0.290	0.364	0.149	0.207	-0.295
	(0.149)	(0.120)	(0.142)	(0.123)	(0.091)	(0.135)	(0.174)
Log population, 1940	0.616	0.316	0.575	0.397	0.174	0.152	0.082
	(0.399)	(0.247)	(0.333)	(0.320)	(0.176)	(0.205)	(0.384)
Percent black, 1940	7.496	-2.925	7.136	3.476	6.642	3.335	7.348
	(2.839)	(2.380)	(1.960)	(3.098)	(1.607)	(2.178)	(2.700)
Log manufacturing employment, 1940	-0.194	-0.149	0.066	0.102	0.038	-0.021	0.395
	(0.227)	(0.176)	(0.152)	(0.190)	(0.106)	(0.141)	(0.168)
Log population, 1950	-0.488	0.461	-0.066	0.007	-0.007	0.174	0.191
	(0.633)	(0.391)	(0.584)	(0.526)	(0.320)	(0.313)	(0.555)
Percent black, 1950	-10.967	-2.408	-9.889	-7.715	-8.805	-2.833	-9.370
	(2.729)	(2.594)	(2.064)	(3.039)	(1.423)	(2.409)	(2.578)

Table A.5: The Effect of Social Connectedness on Crime, 1970-2009, Results for All Explanatory Variables

	Dependent variable: Number of offenses reported to police						
	Murder (1)	Rape (2)	Robbery (3)	Assault (4)	Burglary (5)	Larceny (6)	Motor Vehicle Theft (7)
Log manufacturing employment, 1950	0.511	-0.176	0.263	-0.098	0.153	-0.034	-0.166
	(0.286)	(0.225)	(0.221)	(0.235)	(0.154)	(0.190)	(0.243)
Log population, 1960	-0.077	-0.714	-0.347	-0.121	0.099	0.194	-0.463
	(0.452)	(0.321)	(0.415)	(0.429)	(0.240)	(0.286)	(0.425)
Percent black, 1960	7.413	6.148	4.728	4.422	4.009	0.578	4.124
	(0.934)	(0.922)	(0.745)	(1.245)	(0.642)	(0.909)	(1.132)
Log manufacturing employment, 1960	-0.039	0.275	-0.266	0.034	-0.228	-0.176	0.007
	(0.228)	(0.141)	(0.177)	(0.202)	(0.108)	(0.137)	(0.188)
State fixed effects	x	x	x	x	x	x	x
Pseudo R2	0.823	0.871	0.947	0.914	0.952	0.945	0.935
N (city-years)	8,345	8,345	8,345	8,345	8,345	8,345	8,345
Cities	224	224	224	224	224	224	224

Notes and Sources: See note to Table 3.

Table A.6: Negative Selection of Southern Black Migrants into Connected Destinations, 1960 and 1970

Sample:	Men and Women			Men			Women		
Dependent variable:	Years of Schooling	Log Income	Log Income	Years of Schooling	Log Income	Log Income	Years of Schooling	Log Income	Log Income
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Panel A: Selection into state of residence									
Share of migrants from birth state in state of residence	-1.594	-0.107	-0.041	-1.768	-0.058	0.019	-1.516	-0.025	0.090
	(0.154)	(0.031)	(0.030)	(0.176)	(0.022)	(0.019)	(0.152)	(0.051)	(0.052)
Years of schooling			0.041			0.044			0.076
			(0.002)			(0.001)			(0.005)
N	97,132	77,760	77,760	45,187	42,960	42,960	51,945	34,800	34,800
R2	0.080	0.084	0.099	0.082	0.120	0.147	0.082	0.110	0.150
Panel B: Selection into metropolitan area of residence									
Share of migrants from birth state in metro of residence	-1.990	-0.182	-0.108	-2.057	-0.118	-0.036	-1.995	-0.154	-0.002
	(0.117)	(0.044)	(0.044)	(0.108)	(0.035)	(0.036)	(0.154)	(0.057)	(0.059)
Years of schooling			0.036			0.039			0.070
			(0.002)			(0.001)			(0.006)
N	66,359	52,958	52,958	30,533	29,201	29,201	35,826	23,757	23,757
R2	0.084	0.070	0.081	0.086	0.102	0.125	0.088	0.096	0.131
Quartic in age	x	x	x	x	x	x	x	x	x
Birth year fixed effects	x	x	x	x	x	x	x	x	x
Birth state fixed effects	x	x	x	x	x	x	x	x	x
State/metro of residence fixed effects	x	x	x	x	x	x	x	x	x
Survey year fixed effects	x	x	x	x	x	x	x	x	x

Notes: Sample limited to African Americans born in the South from 1916-1936 who are living in the North, Midwest, or West regions. Standard errors, clustered by state of residence, are in parentheses.

Sources: Ruggles et al. (2019)

Table A.7: Relationship between Social Connectedness, the Number of Migrants, and the Share of Migrants that Chose their Destination Because of Social Interactions

	Dependent variable: Log HHI, Southern black migrants			
	(1)	(2)	(3)	(4)
Log number, Southern black migrants	-0.447		-0.630	-0.625
	(0.018)		(0.031)	(0.035)
Share of migrants who chose location		-2.483	3.277	3.271
because of social interactions		(0.498)	(0.397)	(0.464)
State fixed effects				x
R2	0.682	0.128	0.792	0.823
N (cities)	224	224	224	224

Notes: We estimate the share of migrants that chose their destination because of social interactions using a structural model, as described in the text.
Sources: Duke SSA/Medicare data

Table A.8: The Effect of Social Connectedness on Crime, 1970-2009, Additional Robustness Checks

	Dependent variable: Number of offenses reported to police						
	Murder (1)	Rape (2)	Robbery (3)	Assault (4)	Burglary (5)	Larceny (6)	Motor Vehicle Theft (7)
Panel A: Including large cities with most extensive measurement error in crime							
Log HHI, Southern black migrants	-0.201	-0.122	-0.211	-0.219	-0.115	-0.078	-0.352
	(0.053)	(0.044)	(0.041)	(0.046)	(0.029)	(0.032)	(0.049)
Pseudo R2	0.945	0.921	0.984	0.943	0.976	0.974	0.971
N (city-years)	8,585	8,585	8,585	8,585	8,585	8,585	8,585
Cities	230	230	230	230	230	230	230
Panel B: Negative binomial model							
Log HHI, Southern black migrants	-0.204	-0.118	-0.211	-0.187	-0.158	-0.078	-0.129
	(0.054)	(0.049)	(0.047)	(0.039)	(0.036)	(0.035)	(0.048)
Pseudo R2	0.271	0.196	0.179	0.124	0.148	0.131	0.157
N (city-years)	8,345	8,345	8,345	8,345	8,345	8,345	8,345
Cities	224	224	224	224	224	224	224
Panel C: Drop observations if dependent variable is below 1/6 or above 6 times city mean							
Log HHI, Southern black migrants	-0.208	-0.103	-0.227	-0.216	-0.143	-0.064	-0.218
	(0.060)	(0.046)	(0.044)	(0.049)	(0.032)	(0.043)	(0.080)
Pseudo R2	0.820	0.880	0.949	0.915	0.955	0.950	0.937
N (city-years)	7,526	7,708	8,302	7,760	8,303	8,315	8,293
Cities	224	224	224	224	224	224	224
Panel D: Drop observations if dependent variable is below 1/6 or above 6 times city median							
Log HHI, Southern black migrants	-0.221	-0.107	-0.227	-0.209	-0.143	-0.064	-0.218
	(0.061)	(0.047)	(0.044)	(0.049)	(0.032)	(0.043)	(0.080)
Pseudo R2	0.822	0.882	0.949	0.916	0.955	0.950	0.937
N (city-years)	7,546	7,715	8,303	7,733	8,306	8,315	8,297
Cities	224	224	224	224	224	224	224
Panel E: Measure HHI using birth county to destination city population flows							
Log HHI, Southern black migrants	-0.178	-0.202	-0.200	-0.259	-0.225	-0.281	-0.191
	(0.091)	(0.118)	(0.110)	(0.093)	(0.101)	(0.115)	(0.139)
Pseudo R2	0.769	0.777	0.862	0.807	0.832	0.785	0.817
N (city-years)	8,345	8,345	8,345	8,345	8,345	8,345	8,345
Cities	224	224	224	224	224	224	224

Notes: In Panel B, we estimate a negative binomial model instead of equation (1). For Panels C and D, we construct mean and median number of crimes for each city from 1970-2009. Regressions include the same covariates used in Table 3. Standard errors, clustered at the city level, are in parentheses.

Sources: Duke SSA/Medicare data, Haines and ICPSR (2010), United States Bureau of the Census (2008), United States Department of Justice, Federal Bureau of Investigation (2005)

Table A.9: The Effect of Social Connectedness on Crime, 1970-2009, Robustness to Minimum Number of Migrants in Each City

	Dependent variable: Number of offenses reported to police						
	Murder	Rape	Robbery	Assault	Burglary	Larceny	Motor Vehicle Theft
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Panel A: At Least 5 Southern Black Migrants (9,966 city-years, 267 cities)							
Log HHI, Southern black migrants	-0.215	-0.123	-0.218	-0.213	-0.137	-0.079	-0.141
	(0.059)	(0.043)	(0.041)	(0.041)	(0.030)	(0.038)	(0.074)
Panel B: At Least 10 Southern Black Migrants (9,582 city-years, 257 cities)							
Log HHI, Southern black migrants	-0.224	-0.123	-0.219	-0.215	-0.138	-0.081	-0.158
	(0.060)	(0.044)	(0.042)	(0.043)	(0.031)	(0.039)	(0.077)
Panel C: At Least 25 Southern Black Migrants - Baseline Approach (8,345 city-years, 2224 cities)							
Log HHI, Southern black migrants	-0.245	-0.105	-0.234	-0.221	-0.149	-0.069	-0.227
	(0.064)	(0.048)	(0.045)	(0.047)	(0.032)	(0.043)	(0.083)
Panel D: At Least 50 Southern Black Migrants (6,871 city-years, 184 cities)							
Log HHI, Southern black migrants	-0.266	-0.132	-0.239	-0.231	-0.139	-0.070	-0.258
	(0.066)	(0.048)	(0.043)	(0.049)	(0.032)	(0.044)	(0.081)
Panel E: At Least 100 Southern Black Migrants (6,218 city-years, 166 cities)							
Log HHI, Southern black migrants	-0.267	-0.142	-0.239	-0.243	-0.143	-0.073	-0.248
	(0.066)	(0.048)	(0.044)	(0.050)	(0.032)	(0.045)	(0.084)

Notes: Table displays estimates of equation (1). The sample in each panel differs based on the minimum number of Southern black migrants in each city. Regression includes same covariates as in Table 3. Standard errors, clustered at the city level, are in parentheses.
Sources: Duke SSA/Medicare data, Haines and ICPSR (2010), United States Bureau of the Census (2008), United States Department of Justice, Federal Bureau of Investigation (2005)

Table A.10: The Effect of Social Connectedness on Crime, 1970-2009, Robustness to Dropping Cities with Zero Crimes

	Dependent variable: Number of offenses reported to police						
	Murder	Rape	Robbery	Assault	Burglary	Larceny	Motor Vehicle Theft
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Panel A: Drop City-Year Observation if Any Property Crime Missing/Zero - Baseline Approach (8,345 city-years, 224 cities)							
Log HHI, Southern black migrants	-0.245	-0.105	-0.234	-0.221	-0.149	-0.069	-0.227
	(0.064)	(0.048)	(0.045)	(0.047)	(0.032)	(0.043)	(0.083)
Panel B: Drop City-Year Observation if All Property Crimes Missing/Zero (8,358 city-years, 224 cities)							
Log HHI, Southern black migrants	-0.244	-0.105	-0.233	-0.218	-0.147	-0.068	-0.226
	(0.064)	(0.048)	(0.045)	(0.047)	(0.032)	(0.043)	(0.083)
Panel C: Do Not Drop City-Year Observation if Property Crimes Missing/Zero (8,770 city-years, 224 cities)							
Log HHI, Southern black migrants	-0.245	-0.117	-0.230	-0.229	-0.148	-0.073	-0.224
	(0.065)	(0.048)	(0.044)	(0.048)	(0.033)	(0.044)	(0.081)

Notes: Table displays estimates of equation (1). The sample in each panel differs based on the minimum number of Southern black migrants in each city. Regression includes same covariates as in Table 3. Standard errors, clustered at the city level, are in parentheses.
Sources: Duke SSA/Medicare data, Haines and ICPSR (2010), United States Bureau of the Census (2008), United States Department of Justice, Federal Bureau of Investigation (2005)

Table A.11: The Effect of Social Connectedness on Murder, County-Level Analysis from 1970-1988, Comparing UCR and NCHS Data

	All counties in baseline sample		Excluding Alameda, CA; Franklin, OH; Summit, OH	
Source of dependent variable (number of murders):	UCR (1)	NCHS (2)	UCR (3)	NCHS (4)
Log HHI, Southern black migrants	-0.167	-0.119	-0.154	-0.154
	(0.064)	(0.066)	(0.057)	(0.053)
Pseudo R2	0.831	0.821	0.829	0.832
N (county-years)	3,888	3,888	3,831	3,831
Counties	207	207	204	204

Notes: Table displays estimates of equation (1). We use county-level data for this analysis, as this is the smallest level of geographic detail in the publicly available NCHS (vital statistics) homicide data. Regression includes same covariates as in Table 3. Columns 3-4 exclude three counties (Alameda, CA; Franklin, OH; and Summit, OH) that have the largest mean difference in the number of murders in the UCR and NCHS data. The table shows that, aside from these three counties, our results are nearly identical in both data sets. Standard errors, clustered at the county level, are in parentheses.

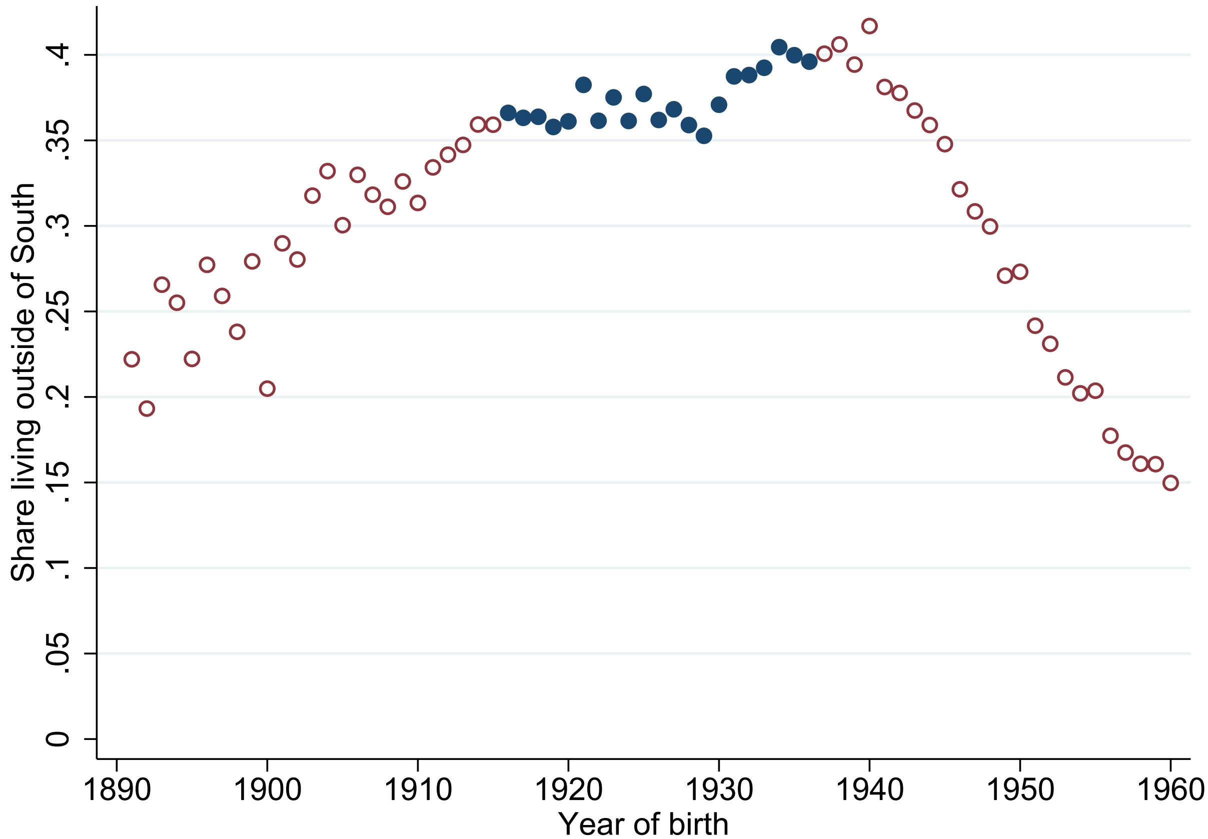
Sources: Duke SSA/Medicare data, Haines and ICPSR (2010), United States Department of Health and Human Services, Centers for Disease Control and Prevention (2010), United States Bureau of the Census (2012), United States Department of Justice, Federal Bureau of Investigation (2005)

Table A.12: The Effect of Social Connectedness on Murder, 1976-2009, By Victim Race, Characteristic, and Circumstance

		Black victims		Non-black victims	
		Share of all black victims	Coefficient on Log HHI, S. black migrants	Share of all non-black victims	Coefficient on Log HHI, S. black migrants
(1)	Total victims	1.00	-0.286	1.00	-0.288
			(0.092)		(0.073)
	Circumstance				
(2)	Gang and drug activity	0.11	-0.546	0.09	-0.813
			(0.209)		(0.199)
(3)	Felony	0.13	-0.299	0.22	-0.347
			(0.135)		(0.088)
(4)	Argument	0.32	-0.204	0.30	-0.189
			(0.083)		(0.091)
(5)	Other	0.12	-0.147	0.15	-0.150
			(0.104)		(0.062)
(6)	Unknown	0.32	-0.389	0.22	-0.310
			(0.185)		(0.115)
	Weapon				
(7)	Gun	0.70	-0.359	0.54	-0.455
			(0.127)		(0.110)
(8)	Other	0.26	-0.132	0.40	-0.136
			(0.051)		(0.052)
(9)	Unknown	0.04	-0.218	0.04	-0.199
			(0.156)		(0.107)
	Age of victim				
(10)	0-9	0.04	-0.156	0.04	-0.236
			(0.094)		(0.082)
(11)	10-17	0.07	-0.393	0.06	-0.429
			(0.150)		(0.139)
(12)	18-25	0.33	-0.317	0.25	-0.393
			(0.111)		(0.104)
(13)	26-35	0.29	-0.255	0.24	-0.363
			(0.087)		(0.085)
(14)	36+	0.27	-0.292	0.37	-0.193
			(0.087)		(0.062)
	Relationship between victim and offender				
(15)	Romantic partner	0.08	-0.128	0.09	-0.145
			(0.065)		(0.059)
(16)	Family	0.06	-0.210	0.07	-0.076
			(0.078)		(0.073)
(17)	Known, not family	0.31	-0.179	0.28	-0.177
			(0.078)		(0.083)
(18)	Stranger	0.12	-0.179	0.20	-0.363
			(0.134)		(0.103)
(19)	Unknown	0.44	-0.474	0.34	-0.495
			(0.187)		(0.125)

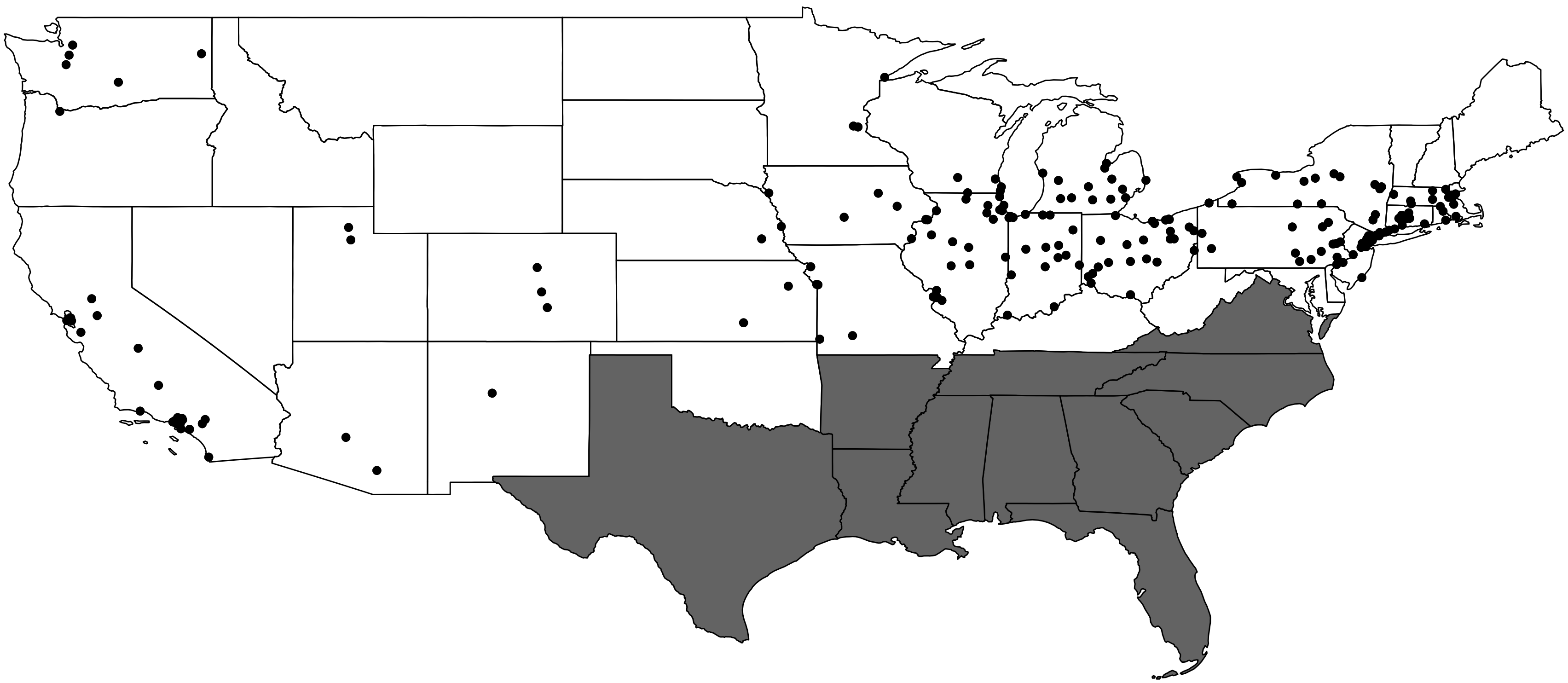
Notes: Table displays estimates of equation (1), using the same specification as Table 3. The dependent variable is the number of murders, by the indicated characteristic or circumstance. Standard errors, clustered at the city level, are in parentheses.
Sources: Duke SSA/Medicare data, Haines and ICPSR (2010), United States Bureau of the Census (2008), United States Department of Justice, Federal Bureau of Investigation (2006)

Figure A.1: Share of African Americans Born in the South Living Outside the South in Their 40s



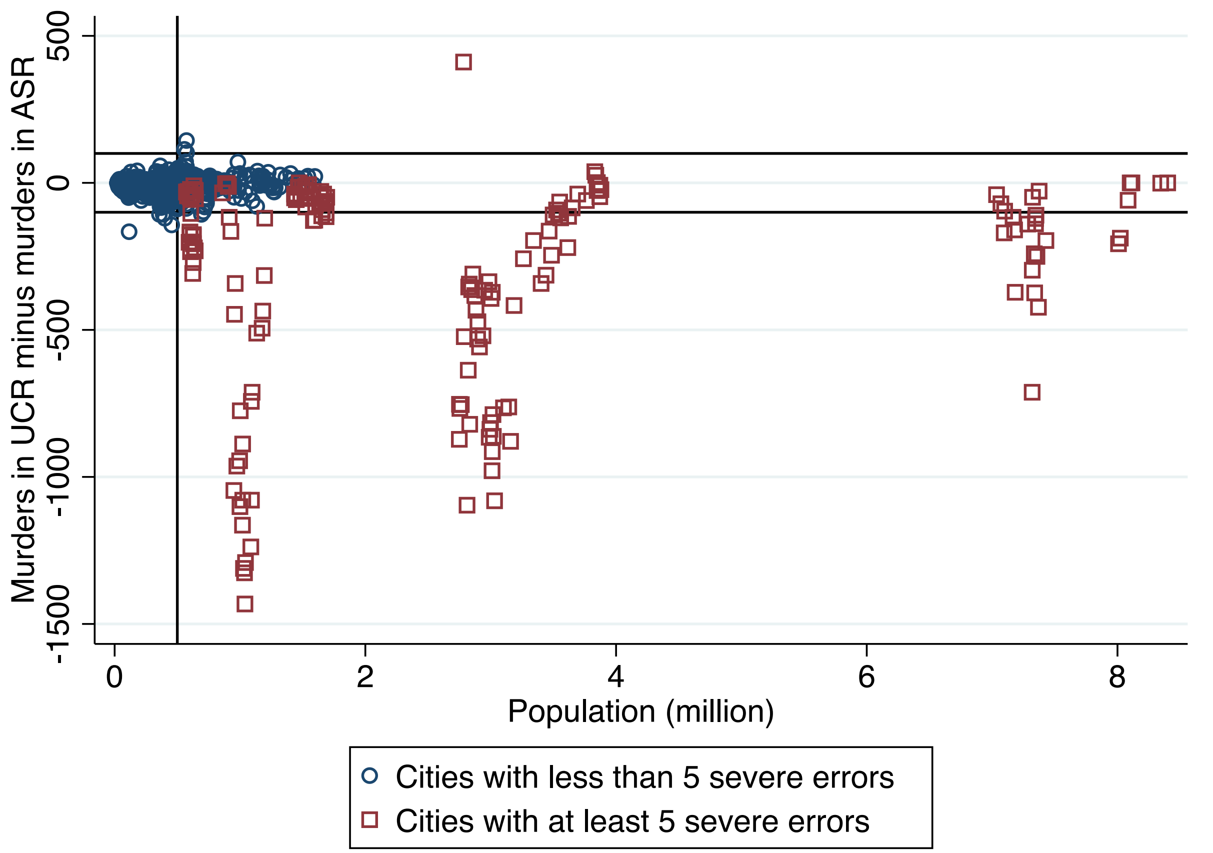
Notes: Sample contains African Americans from the eleven former Confederate states. For individuals born from 1891-1900, we measure their location using the 1940 Census. For individuals born from 1901-1910, we use the 1950 Census, and so forth. The shaded circles correspond to individuals born from 1916-1936, who comprise our sample from the Duke SSA/Medicare data.
Source: Ruggles et al. (2019)

Figure A.2: Cities in our Main Sample



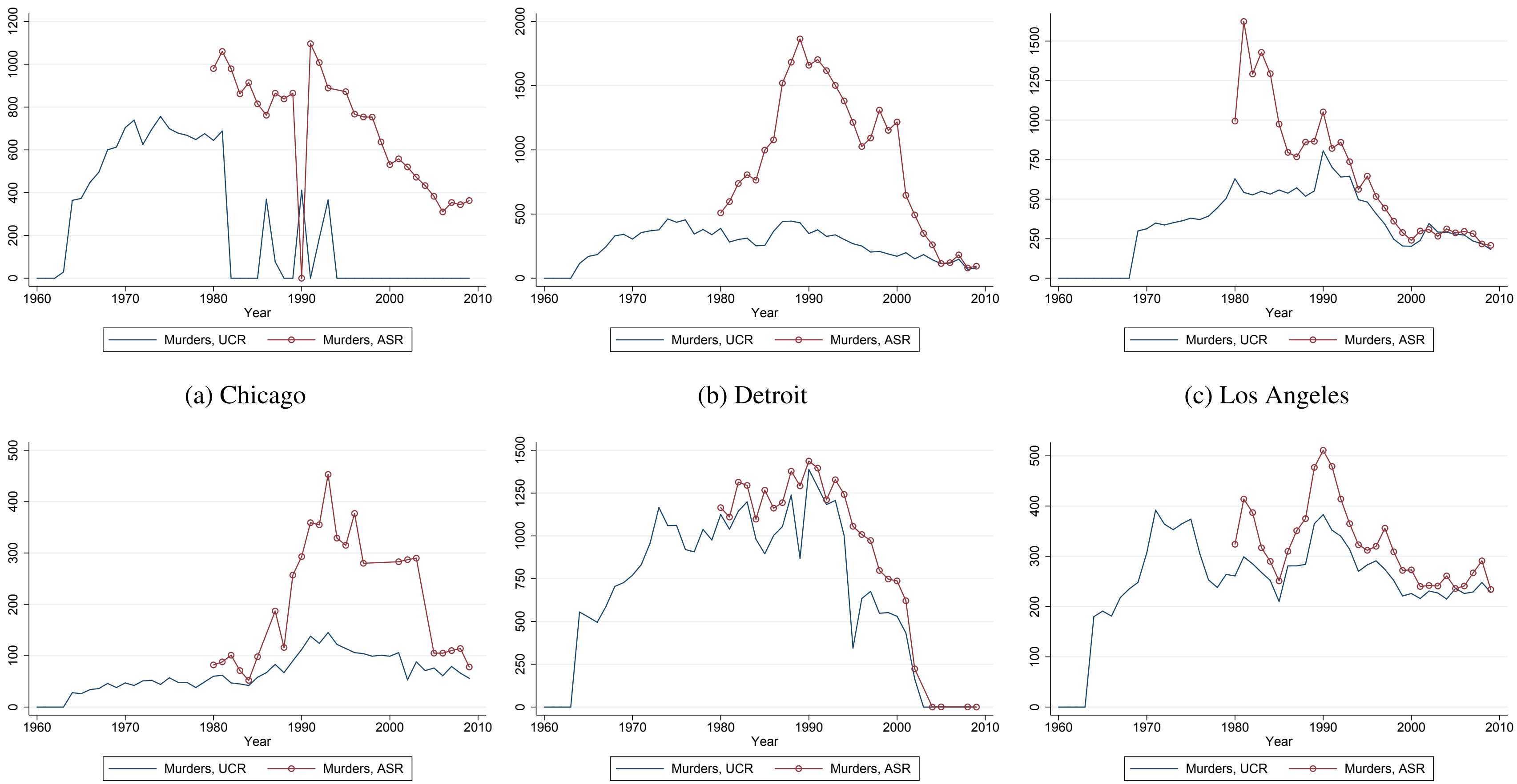
Notes: Figure displays the 224 cities in our main analysis sample.

Figure A.3: Comparison of Murders Cleared by Arrest in FBI UCR versus ASR Data



Notes: We classify a "severe error" as a year in which the absolute value of the difference between murders in the UCR and ASR data is at least 100. The six cities that would be in our main analysis sample except for the presence of at least five severe errors are Chicago, Detroit, Los Angeles, Milwaukee, New York, and Philadelphia.
Source: United States Department of Justice, Federal Bureau of Investigation (2005, 2009)

Figure A.4: The Relationship Between the Number of Murders Cleared by Arrest in UCR and ASR Data, 1960-2009, Severe Measurement Error Cities



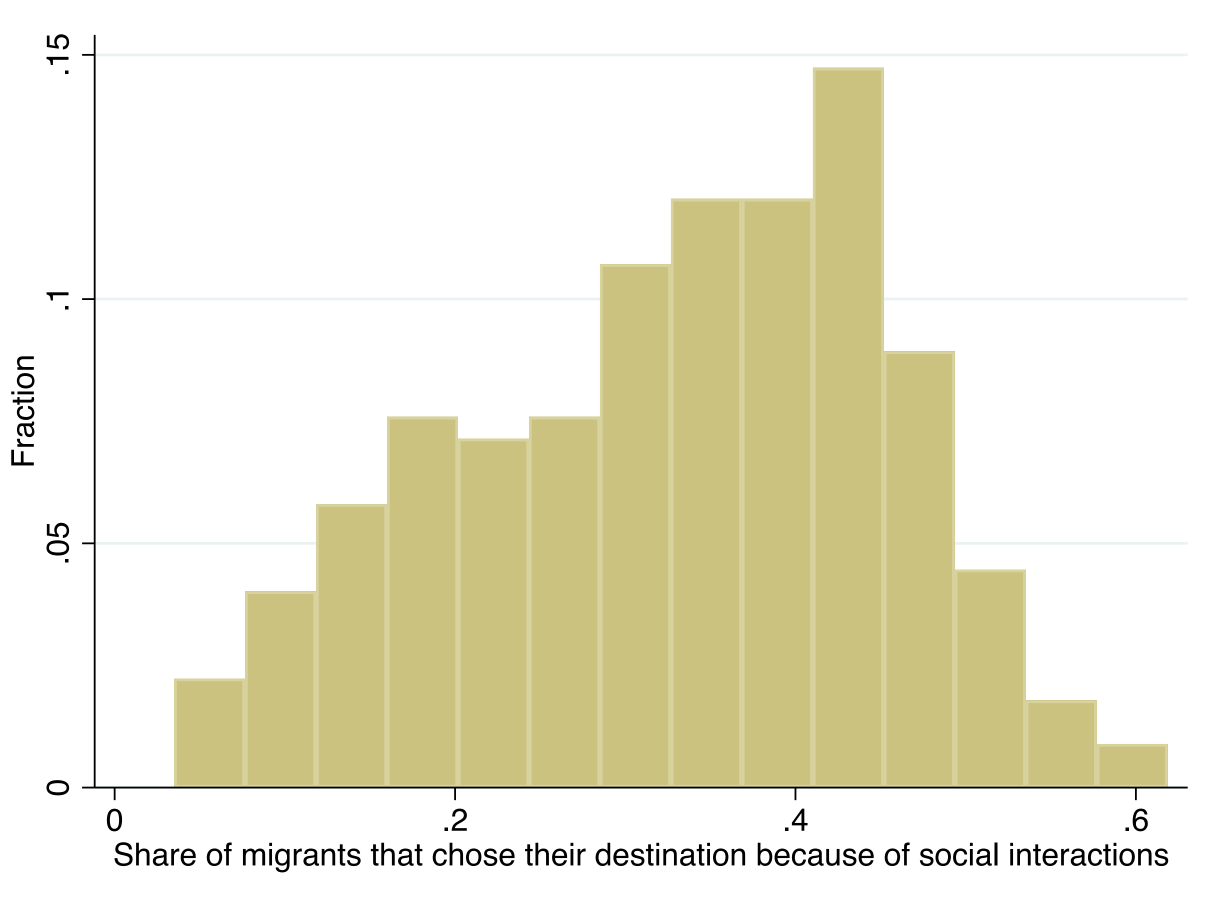
(a) Chicago (b) Detroit (c) Los Angeles

(d) Milwaukee (e) New York (f) Philadelphia

XXX

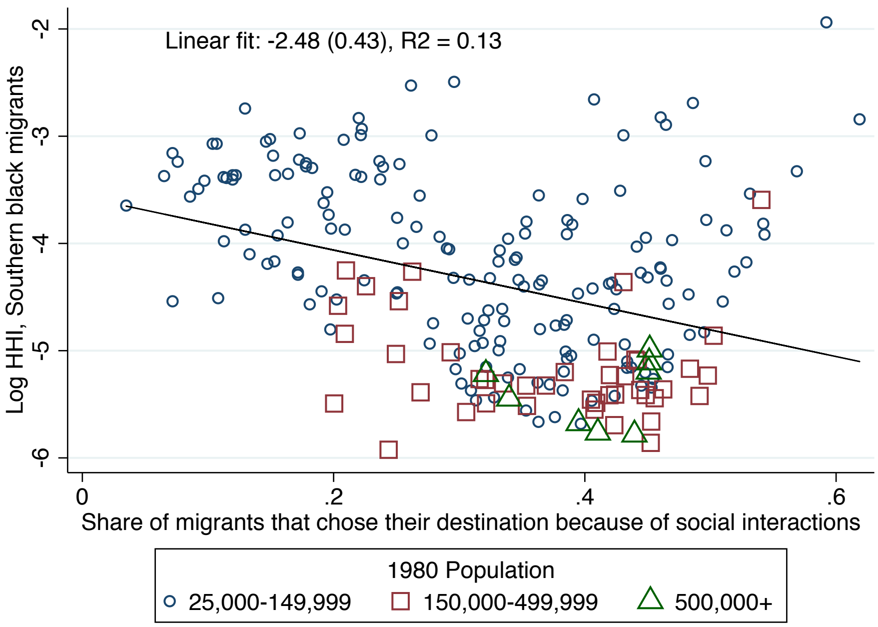
Notes: ASR data are first available in 1980. The cities in Appendix Figure A.4 are those for which the absolute value of the difference in murders between UCR and ASR data is at least 100 for at least five years.
Source: United States Department of Justice, Federal Bureau of Investigation (2005, 2009)

Figure A.5: Share of Migrants that Chose their Destination Because of Social Interactions

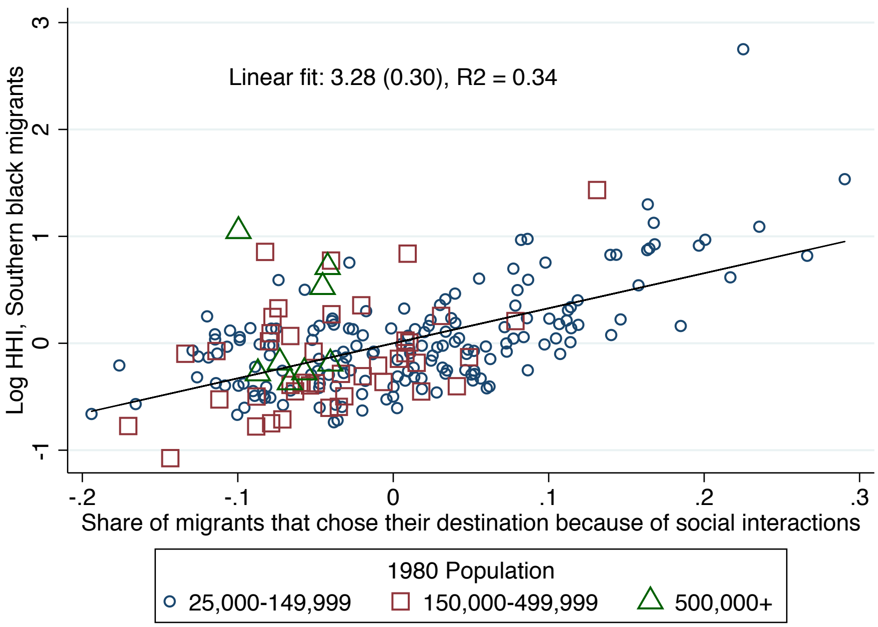


Notes: We estimate the share of migrants that chose their destination because of social interactions using a structural model, as described in the text.
Source: Duke SSA/Medicare data

Figure A.6: The Relationship between Social Connectedness and the Share of Migrants that Chose their Destination Because of Social Interactions



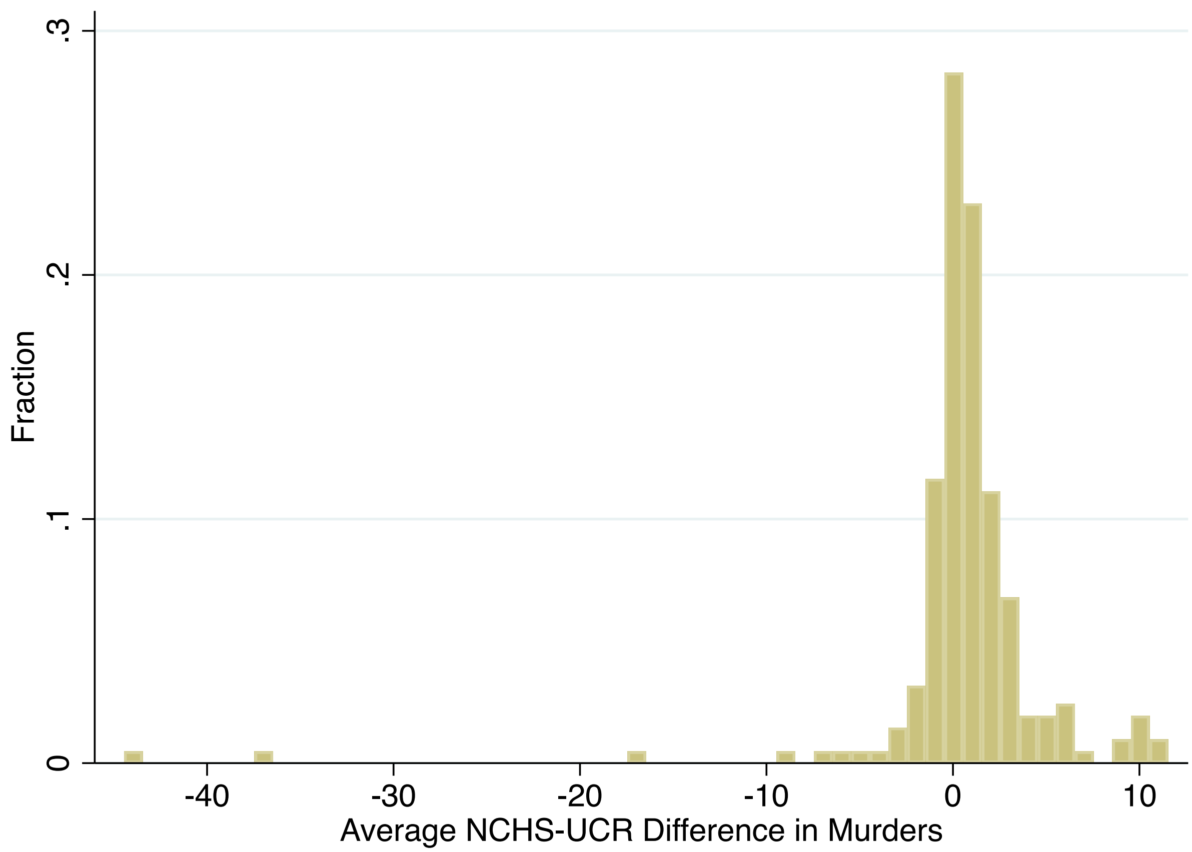
(a) Unconditional



(b) Conditional on Log Number, Southern Black Migrants

Notes: We estimate the share of migrants that chose their destination because of social interactions using a structural model, as described in the text. Panel B plots the residuals from regressing log HHI and the share of migrants that chose their destination because of social interactions on the log number of migrants.
Source: Duke SSA/Medicare data

Figure A.7: Average Difference in Murders in NCHS Relative to UCR Data



Notes: Figure reports the average difference in murders in the NCHS and UCR data. Positive numbers indicate that the NCHS data contain more murders on average than the UCR data. Sample limited to counties in our baseline sample. Source: United States Department of Health and Human Services, Centers for Disease Control and Prevention (2010), United States Department of Justice, Federal Bureau of Investigation (2005)