Online Appendix

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

Bryan A. Stuart and Evan J. Taylor

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

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

where $\text{HHI}^s \equiv \sum_j (N^s_{j,0}/N^s_0)^2$ is the Herfindahl-Hirschman Index of birth town to destination city population flows for African Americans from the South.[1] HHI^s approximately equals the probability that two randomly chosen members of the older generation share a birth town.[2] 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,

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

[1] 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.

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

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

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 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.[3] 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.[4] 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. *$d\bar{C}^s/d\text{HHI}^s \leq 0, d\bar{C}^n/d\text{HHI}^s \leq 0$, and $d\bar{C}^w/d\text{HHI}^s \leq 0$ if $\partial F^s/\partial \text{HHI}^s < 0$, 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,

[3] 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.

[4] 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).

$\bar{C}$, 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 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

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

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.[5] These data are used regularly in

[5] 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.

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.

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.9, our results are similar when we include these large cities.[6]

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.

[6] 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 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.[7]

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

$$\begin{aligned} \rho_{j,k} \equiv \mathbb{P}[D_{i,j,k} = 1 | D_{i-1,j,k} = 1, i \in j] &= \mathbb{P}[k \in H_i | i \in j] + \mathbb{P}[k \in M_i | i \in j] && \text{(A.14)} \\ &= h_{j,k} + m_{j,k}, && \text{(A.15)} \end{aligned}$$

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

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

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 $m_{j,k}$.[8] As a result, the share of migrants in destination k that

[7] 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.

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

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[9]

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

Substituting equation (A.22) into equation (A.21) and simplifying yields[10]

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

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

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

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

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.[11] 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 converges very rapidly in the vast majority of cases.[12]

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 0.04 to 0.60. The unweighted average of m_k across cities is 0.32, and the 1980 population weighted average is 0.36.

Appendix Table A.8 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

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

[12] For 42 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 42 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.

conditional on the log number of migrants.[13] 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.9 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.[14] Results are also similar when we estimate linear models where the dependent variable is the log number of crimes.[15]

Appendix Table A.10 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.

Appendix Table A.11 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 2. 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

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

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

[15] 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 2.

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.12. 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.

References

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

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.

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

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

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

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

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.

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."

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

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

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]."

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

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	264	209	53	114	199	355	717
Assault	1,245	691	394	742	1,097	1,596	2,522
Burglary	1,370	496	644	999	1,332	1,691	2,212
Larceny	3,372	1,301	1,588	2,460	3,332	4,099	5,031
Motor Vehicle Theft	639	424	205	323	463	904	1,415

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.
Source: United States Department of Justice, Federal Bureau of Investigation (2005)

Table A.3: 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.

Source: Ruggles et al. (2019)

Table A.4: Social Connectedness and Migration Flows, Southern Black Migrants, by Destination City

Rank	City	Percent of Migrants from Top Sending Town	HHI	Number of Migrants	Residualized Log HHI
1	Decatur, IL	37.5	0.144	686	2.64
2	Fort Wayne, IN	13.5	0.028	1,462	1.11
3	York, PA	22.7	0.059	194	0.97
4	Troy, NY	16.2	0.039	204	0.96
5	Erie, PA	13.4	0.029	647	0.90
6	Beloit, WI	18.1	0.058	342	0.87
7	Cincinnati, OH	4.1	0.008	6,565	0.75
8	Auburn, NY	25.0	0.083	44	0.74
9	Garfield, NJ	19.2	0.080	26	0.73
10	Waterbury, CT	11.2	0.021	713	0.73
11	Easton, PA	19.6	0.050	112	0.67
12	Niagara Falls, NY	7.7	0.019	742	0.67
13	Cleveland, OH	4.6	0.006	18,374	0.63
14	Waterloo, IA	12.0	0.030	435	0.61
15	Paterson, NJ	7.7	0.011	1,866	0.58
16	Newton, MA	11.1	0.035	45	0.58
17	Lima, OH	12.1	0.023	572	0.58
18	Richmond, IN	19.4	0.055	108	0.58
19	Duluth, MN	11.6	0.038	43	0.53
20	Aurora, IL	10.9	0.022	384	0.53
21	Anderson, IN	12.3	0.036	374	0.53
22	Joplin, MO	16.3	0.068	49	0.52
23	Inglewood, CA	5.9	0.009	3,058	0.50
24	Middletown, CT	11.2	0.028	143	0.47
25	Seattle, WA	4.4	0.005	2,970	0.46
26	Santa Barbara, CA	7.7	0.018	117	0.44
27	Dearborn, MI	8.1	0.033	37	0.43
28	Oakland, CA	4.8	0.006	11,506	0.42
29	East Chicago, IN	9.0	0.020	858	0.42
30	Racine, WI	11.8	0.022	773	0.41
31	Hoboken, NJ	11.6	0.039	43	0.40
32	Everett, WA	8.0	0.046	25	0.39
33	Burbank, CA	18.5	0.064	27	0.39
34	San Francisco, CA	5.2	0.007	6,632	0.38
35	Kalamazoo, MI	7.6	0.012	537	0.37
36	Hackensack, NJ	6.4	0.012	375	0.36
37	Indianapolis (balance), IN	4.2	0.006	6,922	0.32
38	Muskegon, MI	7.3	0.014	454	0.31
39	Cleveland Heights, OH	6.0	0.009	832	0.30
40	East St. Louis, IL	4.1	0.010	3,111	0.29
41	Warren, OH	8.0	0.015	733	0.29
42	Evansville, IN	11.0	0.022	264	0.29
43	Ogden, UT	6.2	0.019	112	0.28
44	East Cleveland, OH	5.6	0.007	2,194	0.28
45	New Rochelle, NY	5.5	0.008	621	0.27
46	Alhambra, CA	11.8	0.050	34	0.27

Table A.4: Social Connectedness and Migration Flows, Southern Black Migrants, by Destination City

Rank	City	Percent of Migrants from Top Sending Town	HHI	Number of Migrants	Residualized Log HHI
47	Rockford, IL	5.2	0.011	1,295	0.26
48	Bayonne, NJ	5.6	0.023	124	0.26
49	Clifton, NJ	8.3	0.037	36	0.24
50	Ann Arbor, MI	3.5	0.007	370	0.22
51	Belleville CDP, NJ	4.8	0.026	42	0.22
52	Malden, MA	5.9	0.035	34	0.21
53	Beverly Hills, CA	9.5	0.034	42	0.21
54	Fitchburg, MA	10.7	0.048	28	0.19
55	Atlantic City, NJ	2.3	0.005	876	0.19
56	Medford, MA	4.8	0.028	42	0.19
57	Denver, CO	1.7	0.003	3,435	0.19
58	Pittsburgh, PA	4.9	0.006	3,728	0.19
59	Holyoke, MA	12.8	0.038	47	0.18
60	Alton, IL	7.2	0.018	335	0.17
61	Springfield, OH	5.2	0.015	484	0.17
62	Kansas City, KS	2.2	0.005	1,906	0.16
63	Norwalk, CT	6.2	0.010	530	0.16
64	Bristol, CT	10.4	0.035	48	0.16
65	Burlington, IA	7.7	0.050	26	0.15
66	Clinton, IA	14.8	0.059	27	0.15
67	Galesburg, IL	10.3	0.029	78	0.14
68	Hamilton, OH	13.8	0.051	29	0.14
69	Newport, RI	4.3	0.019	69	0.14
70	Buffalo, NY	3.5	0.004	6,811	0.14
71	Passaic, NJ	6.5	0.013	447	0.14
72	Pittsfield, MA	7.7	0.033	52	0.14
73	Lowell, MA	5.3	0.037	38	0.14
74	Topeka, KS	2.7	0.007	403	0.13
75	St. Louis, MO	3.3	0.006	11,317	0.13
76	Flint, MI	2.8	0.004	4,758	0.12
77	Lafayette, IN	14.3	0.048	35	0.12
78	Akron, OH	4.4	0.006	3,669	0.12
79	Sacramento, CA	3.7	0.004	3,317	0.11
80	Grand Rapids, MI	3.2	0.007	1,482	0.10
81	White Plains, NY	4.1	0.009	368	0.10
82	Port Huron, MI	5.5	0.017	145	0.09
83	Newburgh, NY	9.4	0.015	384	0.09
84	Hartford, CT	4.7	0.007	1,525	0.08
85	Woonsocket, RI	17.2	0.070	29	0.07
86	Zanesville, OH	13.3	0.053	30	0.07
87	Bakersfield, CA	4.7	0.008	488	0.06
88	Yakima, WA	3.2	0.015	93	0.06
89	University City, MO	3.4	0.006	1,086	0.05
90	Cedar Rapids, IA	9.2	0.029	87	0.05
91	Elyria, OH	5.7	0.013	470	0.04
92	Glendale, CA	12.2	0.034	49	0.04

Table A.4: Social Connectedness and Migration Flows, Southern Black Migrants, by Destination City

Rank	City	Percent of Migrants from Top Sending Town	HHI	Number of Migrants	Residualized Log HHI
93	Joliet, IL	3.8	0.008	965	0.04
94	Scranton, PA	11.4	0.040	35	0.03
95	Massillon, OH	5.9	0.020	205	0.03
96	San Bernardino, CA	2.2	0.004	1,291	0.03
97	Lincoln, NE	3.4	0.014	118	0.03
98	Tucson, AZ	1.7	0.004	929	0.02
99	San Diego, CA	2.6	0.003	4,173	0.02
100	West Orange CDP, NJ	1.8	0.011	112	0.00
101	Davenport, IA	7.4	0.019	215	0.00
102	Portland, OR	2.3	0.005	2,078	0.00
103	Albuquerque, NM	2.4	0.005	576	-0.00
104	Long Beach, CA	3.8	0.005	2,112	-0.01
105	Jersey City, NJ	2.2	0.004	2,645	-0.01
106	Bloomfield CDP, NJ	2.9	0.014	104	-0.01
107	Phoenix, AZ	3.3	0.004	1,996	-0.02
108	Omaha, NE	2.1	0.004	1,918	-0.03
109	Colorado Springs, CO	2.5	0.004	734	-0.05
110	Columbus, OH	2.1	0.003	5,174	-0.06
111	Riverside, CA	3.2	0.005	926	-0.06
112	Evanston, IL	2.9	0.006	734	-0.06
113	New Haven, CT	4.5	0.006	1,696	-0.06
114	Middletown, OH	5.3	0.014	380	-0.07
115	Williamsport, PA	8.1	0.039	37	-0.07
116	New Castle, PA	5.1	0.019	99	-0.07
117	Belleville, IL	4.3	0.014	116	-0.07
118	St. Joseph, MO	7.7	0.047	26	-0.07
119	Binghamton, NY	6.2	0.030	48	-0.08
120	New Bedford, MA	7.9	0.033	38	-0.08
121	Dayton, OH	2.7	0.005	4,107	-0.09
122	Bloomington, IL	6.5	0.021	93	-0.09
123	Portsmouth, OH	7.9	0.035	38	-0.09
124	Pasadena, CA	3.8	0.007	1,177	-0.09
125	Mount Vernon, NY	2.3	0.005	1,502	-0.09
126	Perth Amboy, NJ	6.7	0.016	149	-0.10
127	Rochester, NY	3.1	0.005	3,136	-0.11
128	East Orange, NJ	1.9	0.003	2,720	-0.12
129	Jamestown, NY	7.0	0.034	43	-0.12
130	Trenton, NJ	4.3	0.005	2,068	-0.13
131	Pueblo, CO	3.7	0.010	136	-0.13
132	Newark, NJ	1.6	0.003	7,905	-0.14
133	Fresno, CA	2.7	0.005	1,655	-0.14
134	South Gate, CA	8.8	0.042	34	-0.14
135	Berkeley, CA	5.2	0.007	1,874	-0.14
136	Spokane, WA	3.4	0.010	177	-0.15
137	Gary, IN	3.5	0.004	7,149	-0.15
138	Boston, MA	1.6	0.003	4,142	-0.15

Table A.4: Social Connectedness and Migration Flows, Southern Black Migrants, by Destination City

Rank	City	Percent of Migrants from Top Sending Town	HHI	Number of Migrants	Residualized Log HHI
139	Hammond, IN	3.6	0.009	416	-0.16
140	Bay City, MI	6.5	0.034	31	-0.16
141	Norwood, OH	6.1	0.021	82	-0.16
142	St. Paul, MN	4.0	0.007	596	-0.16
143	Norristown borough, PA	5.2	0.013	289	-0.17
144	Albany, NY	2.0	0.006	790	-0.17
145	Wilkes-Barre, PA	5.9	0.035	34	-0.17
146	Irvington CDP, NJ	2.2	0.004	1,248	-0.17
147	Sioux City, IA	4.8	0.024	62	-0.18
148	Jackson, MI	4.4	0.009	321	-0.18
149	Marion, OH	5.7	0.023	70	-0.19
150	Meriden, CT	4.1	0.017	98	-0.19
151	Santa Ana, CA	4.0	0.008	299	-0.20
152	Terre Haute, IN	6.8	0.022	74	-0.21
153	Providence, RI	7.6	0.013	524	-0.21
154	Chester, PA	2.7	0.005	1,144	-0.21
155	Moline, IL	8.0	0.046	25	-0.21
156	Lynn, MA	5.3	0.020	76	-0.22
157	Michigan City, IN	5.9	0.013	388	-0.22
158	Cambridge, MA	4.0	0.013	125	-0.22
159	Rome, NY	3.8	0.017	80	-0.23
160	West Allis, WI	7.4	0.043	27	-0.23
161	Lancaster, PA	7.6	0.020	132	-0.24
162	Danville, IL	4.1	0.013	266	-0.25
163	Peoria, IL	3.0	0.008	1,038	-0.25
164	Utica, NY	5.9	0.010	321	-0.25
165	Montclair CDP, NJ	2.0	0.005	590	-0.25
166	Stamford, CT	2.9	0.006	581	-0.26
167	Reading, PA	7.8	0.014	296	-0.27
168	New London, CT	2.5	0.008	198	-0.27
169	Youngstown, OH	3.4	0.005	2,360	-0.27
170	Mansfield, OH	7.3	0.016	219	-0.27
171	Lansing, MI	2.5	0.005	974	-0.28
172	Brockton, MA	3.8	0.011	160	-0.28
173	Salt Lake City, UT	3.7	0.014	107	-0.28
174	Elizabeth, NJ	3.0	0.006	767	-0.29
175	Cicero town, IL	5.3	0.030	38	-0.29
176	Wichita, KS	2.7	0.005	941	-0.29
177	Oak Park village, IL	4.1	0.007	442	-0.30
178	Kansas City, MO	2.6	0.004	5,818	-0.31
179	Maywood village, IL	3.2	0.006	1,579	-0.31
180	Newark, OH	4.2	0.021	72	-0.31
181	Worcester, MA	4.5	0.012	157	-0.32
182	New Britain, CT	3.4	0.011	238	-0.32
183	Springfield, MO	3.4	0.015	88	-0.32
184	Battle Creek, MI	4.6	0.007	605	-0.32

Table A.4: Social Connectedness and Migration Flows, Southern Black Migrants, by Destination City

Rank	City	Percent of Migrants from Top Sending Town	HHI	Number of Migrants	Residualized Log HHI
185	Yonkers, NY	1.7	0.005	721	-0.32
186	Saginaw, MI	2.5	0.004	2,223	-0.32
187	Elmira, NY	4.0	0.013	149	-0.33
188	New Albany, IN	7.5	0.027	53	-0.34
189	Orange CDP, NJ	3.0	0.005	868	-0.34
190	San Jose, CA	4.4	0.005	992	-0.35
191	Springfield, MA	2.9	0.004	1,270	-0.36
192	Bethlehem, PA	6.4	0.021	78	-0.36
193	Muncie, IN	5.8	0.013	329	-0.37
194	Pontiac, MI	2.8	0.005	1,513	-0.37
195	Minneapolis, MN	1.9	0.004	1,129	-0.37
196	Marion, IN	7.8	0.015	204	-0.37
197	Stockton, CA	2.3	0.004	1,464	-0.38
198	Springfield, IL	4.0	0.009	372	-0.38
199	Syracuse, NY	2.6	0.004	1,414	-0.38
200	Huntington Park, CA	6.9	0.039	29	-0.38
201	Santa Monica, CA	4.6	0.011	217	-0.39
202	Madison, WI	4.7	0.011	213	-0.39
203	Poughkeepsie, NY	4.8	0.009	293	-0.42
204	Toledo, OH	2.5	0.003	3,786	-0.42
205	Plainfield, NJ	1.8	0.003	1,212	-0.43
206	Steubenville, OH	6.7	0.017	163	-0.43
207	Camden, NJ	1.7	0.004	1,454	-0.43
208	South Bend, IN	4.0	0.007	1,391	-0.44
209	Lorain, OH	4.2	0.007	570	-0.45
210	Schenectady, NY	3.9	0.012	204	-0.45
211	Elgin, IL	4.8	0.012	166	-0.48
212	Harrisburg, PA	2.1	0.005	717	-0.49
213	Elkhart, IN	4.3	0.012	277	-0.49
214	Bridgeport, CT	2.4	0.004	1,358	-0.50
215	Canton, OH	2.7	0.006	825	-0.57
216	Alameda, CA	3.9	0.011	129	-0.59
217	Kokomo, IN	4.7	0.013	172	-0.61
218	Kenosha, WI	6.4	0.013	188	-0.66
219	Rock Island, IL	4.0	0.010	272	-0.68
220	New Brunswick, NJ	2.1	0.006	388	-0.70
221	Waukegan, IL	2.3	0.006	699	-0.71
222	Allentown, PA	3.9	0.012	127	-0.73
223	Tacoma, WA	1.4	0.003	983	-0.77
224	Des Moines, IA	2.3	0.007	300	-0.78

Notes: This table shows cities ranked by residuals from a linear regression of log HHI against the covariates in Table 2.
Sources: Duke SSA/Medicare data, Haines and ICPSR (2010), United States Bureau of the Census (2008)

Table A.5: 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 2. 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.6: 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.6: 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 2.

Table A.7: Negative Selection of Southern Black Migrants into Connected Destinations, 1960 and 1970

Sample:	Men and Women			Men			Women		
Dependent variable:	Years of Schooling (1)	Log Income (2)	Log Income (3)	Years of Schooling (4)	Log Income (5)	Log Income (6)	Years of Schooling (7)	Log Income (8)	Log Income (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.
Source: Ruggles et al. (2019)

Table A.8: 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		-0.627	-0.620
	(0.018)		(0.032)	(0.036)
Share of migrants who chose location because of social interactions		-2.378	3.247	3.207
		(0.504)	(0.407)	(0.482)
State fixed effects				x
R2	0.682	0.119	0.794	0.824
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.
Source: Duke SSA/Medicare data

Table A.9: The Effect of Social Connectedness on Crime, 1970–2009, Additional Robustness Checks

	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: Including large cities with most extensive measurement error in crime							
Log HHI, Southern	-0.201	-0.122	-0.211	-0.219	-0.115	-0.078	-0.352
black migrants	(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	-0.204	-0.118	-0.211	-0.187	-0.158	-0.078	-0.129
black migrants	(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	-0.208	-0.103	-0.227	-0.216	-0.143	-0.064	-0.218
black migrants	(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	-0.221	-0.107	-0.227	-0.209	-0.143	-0.064	-0.218
black migrants	(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	-0.215	-0.091	-0.207	-0.205	-0.132	-0.067	-0.175
black migrants	(0.061)	(0.045)	(0.044)	(0.045)	(0.033)	(0.045)	(0.072)
Pseudo R2	0.822	0.871	0.946	0.913	0.952	0.944	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: 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 2. 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 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, 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 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 2. 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 Crime, 1970–2009, Robustness to Dropping Cities with Zero Crimes

	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: 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 2. 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.12: 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 2. 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.13: The Effect of Social Connectedness on Murder, 1980–1989, Possible Mechanisms, Including Crack Index

	Dependent variable: Number of murders reported to police						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
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.

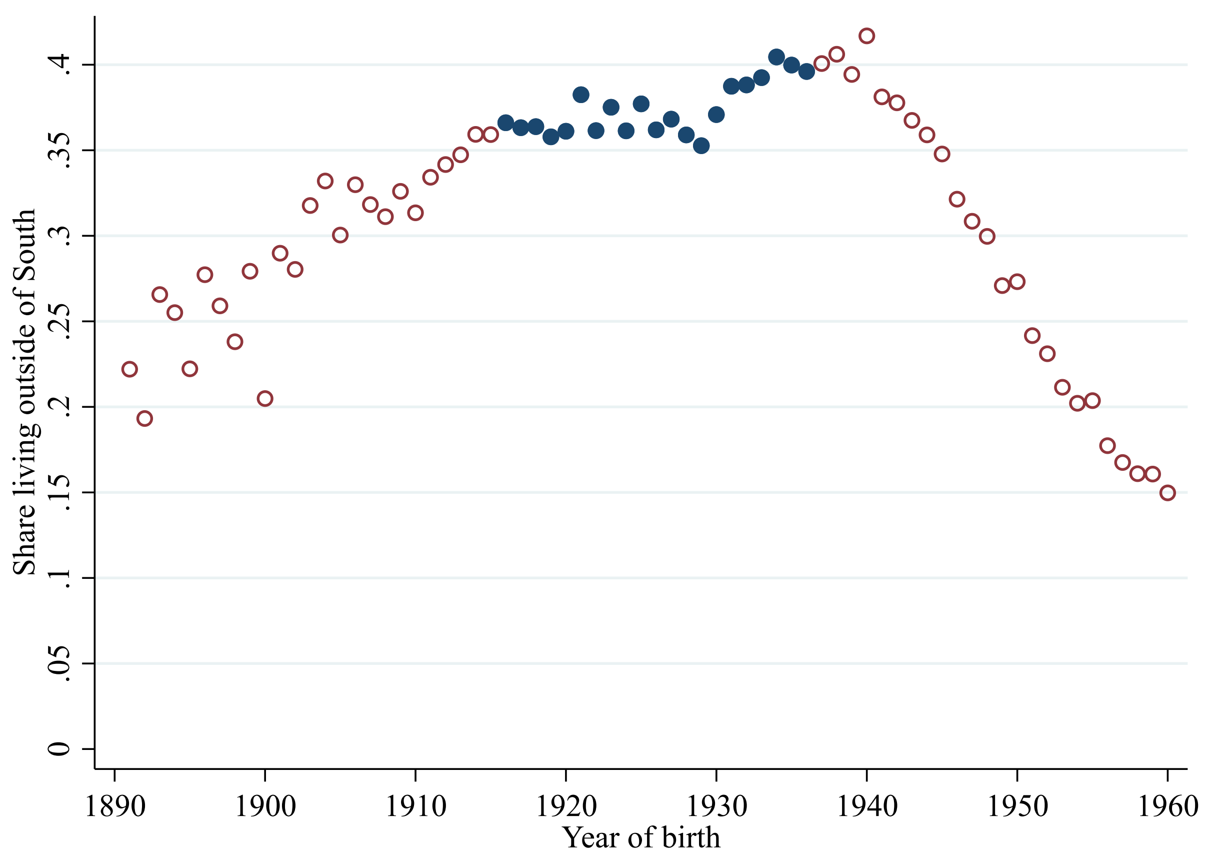
Sources: Duke SSA/Medicare data, Fryer et al. (2013), Haines and ICPSR (2010), United States Bureau of the Census (2008), United States Department of Justice, Federal Bureau of Investigation (2005)

Table A.14: 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)	All victims	1.00	-0.286	1.00	-0.287
			(0.092)		(0.073)
	Circumstance				
(2)	Gang and drug activity	0.11	-0.546	0.09	-0.814
			(0.209)		(0.199)
(3)	Felony	0.13	-0.300	0.22	-0.347
			(0.135)		(0.088)
(4)	Argument	0.32	-0.204	0.30	-0.188
			(0.083)		(0.091)
(5)	Other	0.12	-0.147	0.15	-0.150
			(0.103)		(0.062)
(6)	Unknown	0.32	-0.390	0.22	-0.310
			(0.185)		(0.115)
	Weapon				
(7)	Gun	0.70	-0.359	0.54	-0.454
			(0.127)		(0.110)
(8)	Other	0.26	-0.132	0.40	-0.136
			(0.051)		(0.052)
(9)	Unknown	0.04	-0.219	0.04	-0.198
			(0.156)		(0.107)
	Age of victim				
(10)	0-9	0.04	-0.156	0.04	-0.237
			(0.094)		(0.082)
(11)	10-17	0.07	-0.394	0.06	-0.430
			(0.150)		(0.140)
(12)	18-25	0.33	-0.317	0.25	-0.392
			(0.111)		(0.104)
(13)	26-35	0.29	-0.255	0.24	-0.363
			(0.087)		(0.085)
(14)	36+	0.27	-0.291	0.37	-0.192
			(0.087)		(0.062)
	Relationship between victim and offender				
(15)	Romantic partner	0.08	-0.128	0.09	-0.143
			(0.065)		(0.059)
(16)	Family	0.06	-0.211	0.07	-0.075
			(0.078)		(0.074)
(17)	Known, not family	0.31	-0.178	0.28	-0.176
			(0.072)		(0.081)
(18)	Stranger	0.12	-0.179	0.20	-0.363
			(0.133)		(0.103)
(19)	Unknown	0.44	-0.475	0.34	-0.494
			(0.187)		(0.125)

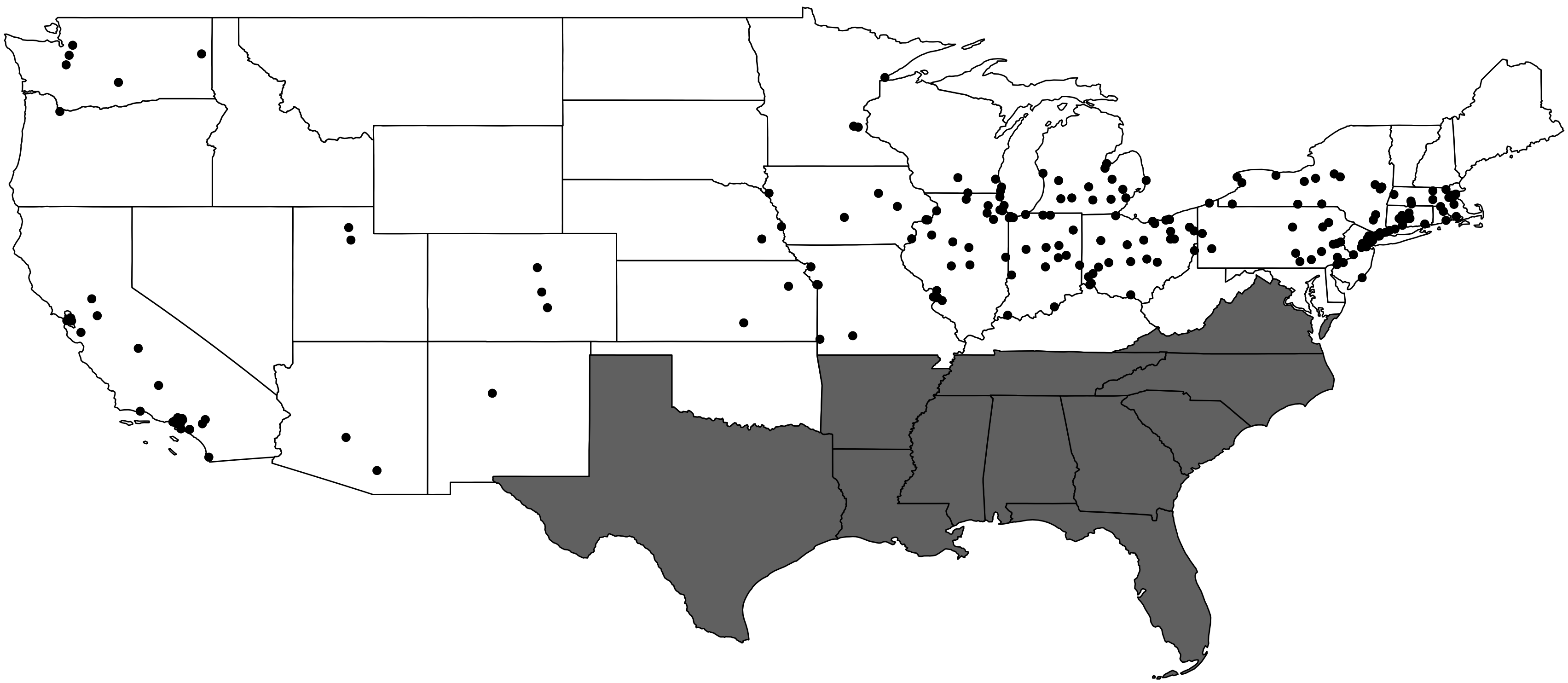
Notes: Table displays estimates of equation (1), using the same specification as Table 2. The dependent variable is the number of murders, by the indicated characteristic or circumstance. Standard errors, clustered at the city level, are in parentheses. Sources: Duke SSA/Medicare data, Haines and ICPSR (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



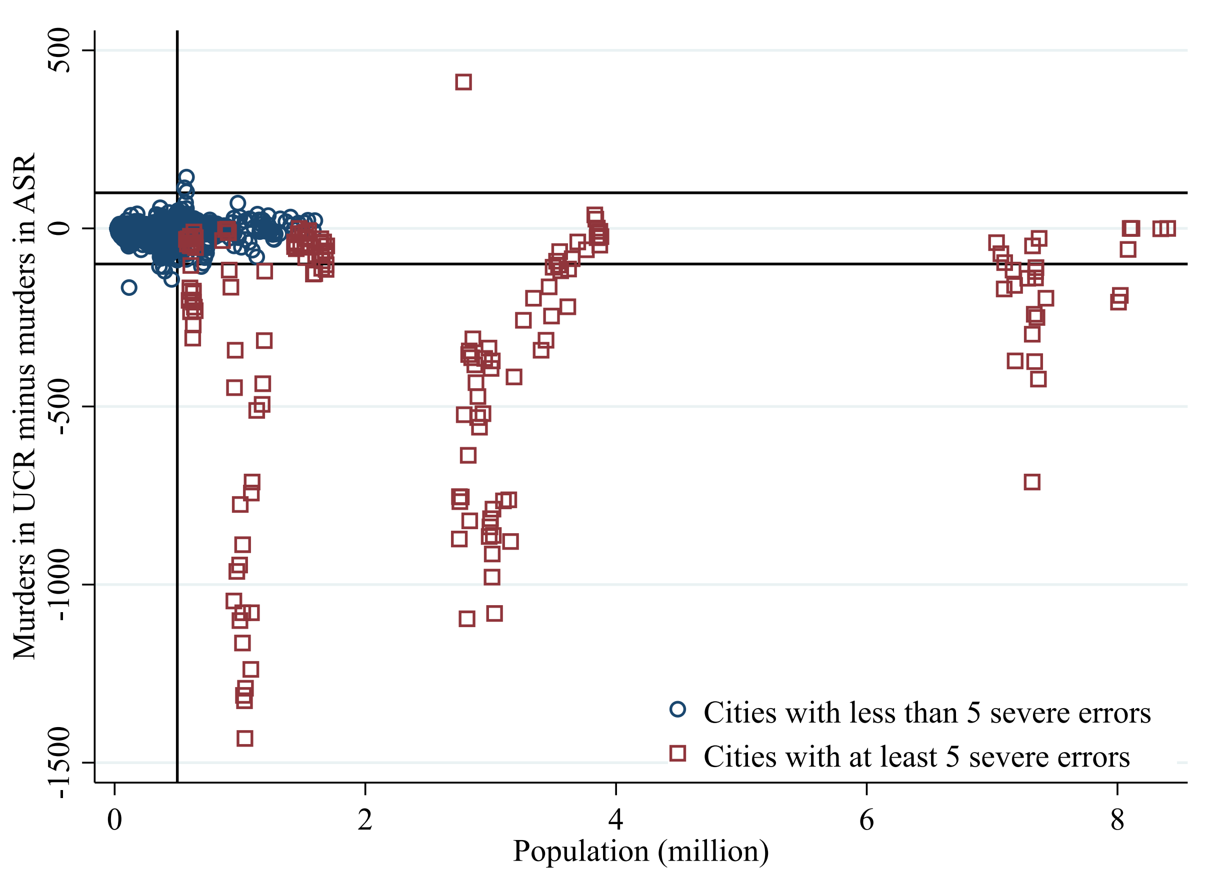
Notes: Sample contains African Americans from the eleven former Confederacy states. For individuals born from 1891–1900, we measure their location using the 1940 Census. For individuals born from 1901–1910, we use the 1950 Census, and so forth. The shaded circles correspond to individuals born from 1916–1936,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

Figure A.2: Cities in our Main Sample



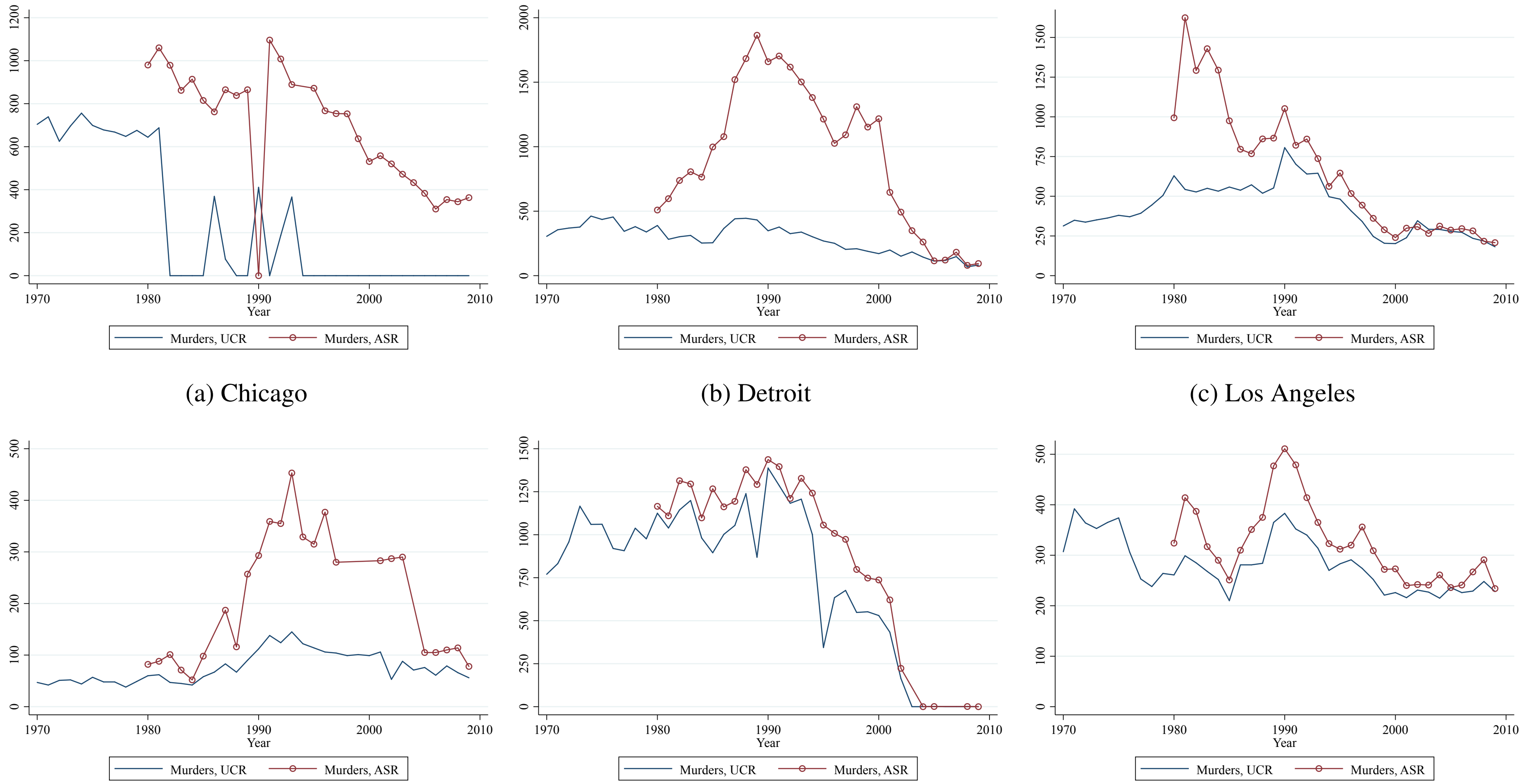
Notes: Figure displays the 224 cities in our main analysis sample. Former Confederacy states, which are excluded from our sample, are in gray.

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.
Sources: 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

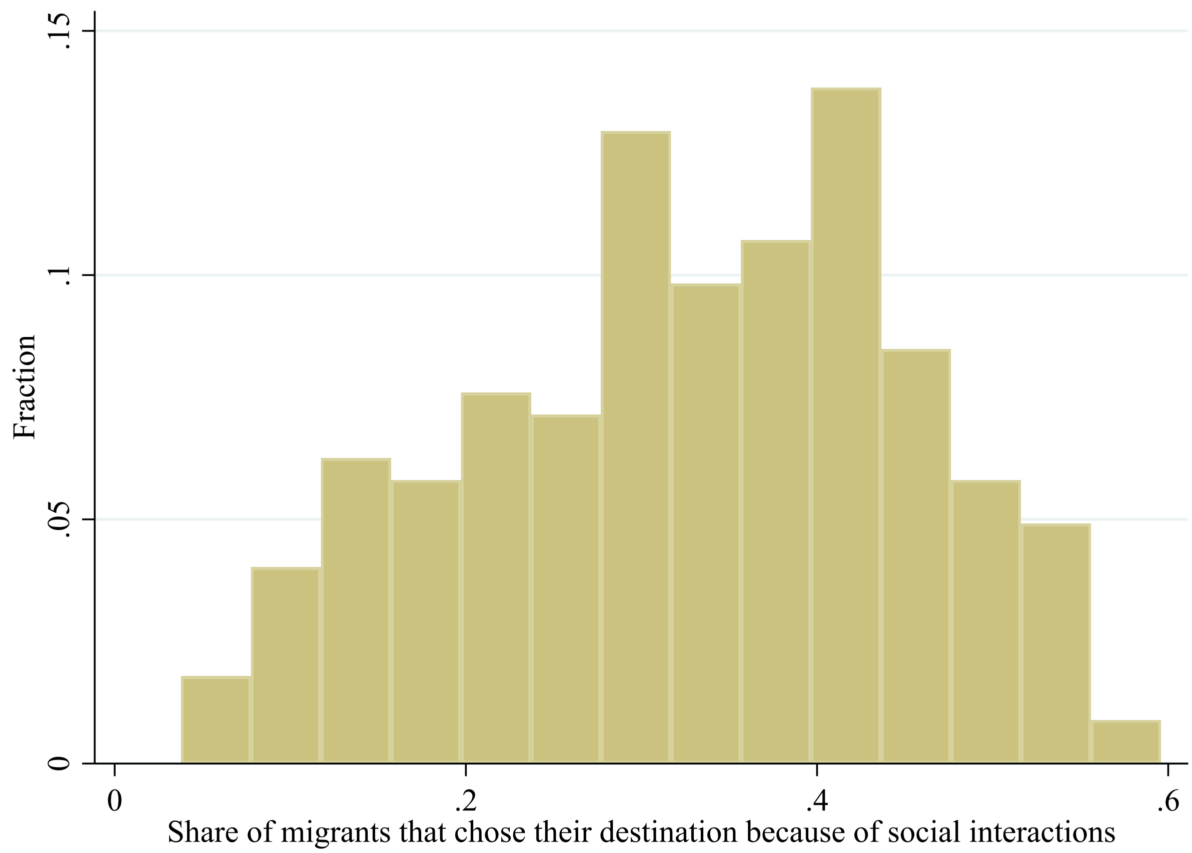
(d) Milwaukee

(e) New York

(f) Philadelphia

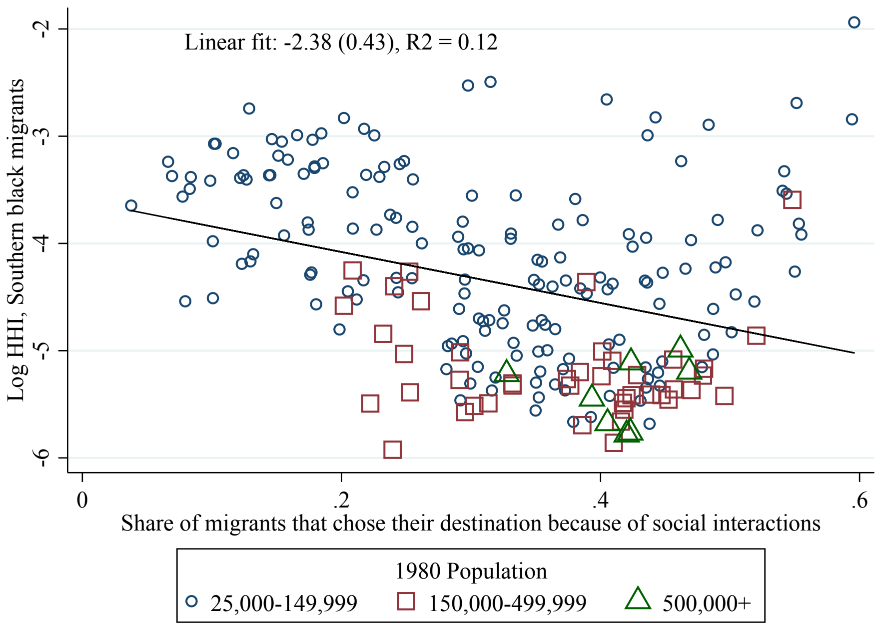
Notes: ASR data are first available in 1980. The cities in Appendix Figure A.4 are those for which the absolute value of the difference in murders between UCR and ASR data is at least 100 for at least five years.
Sources: United States Department of Justice, Federal Bureau of Investigation (2005, 2009)

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

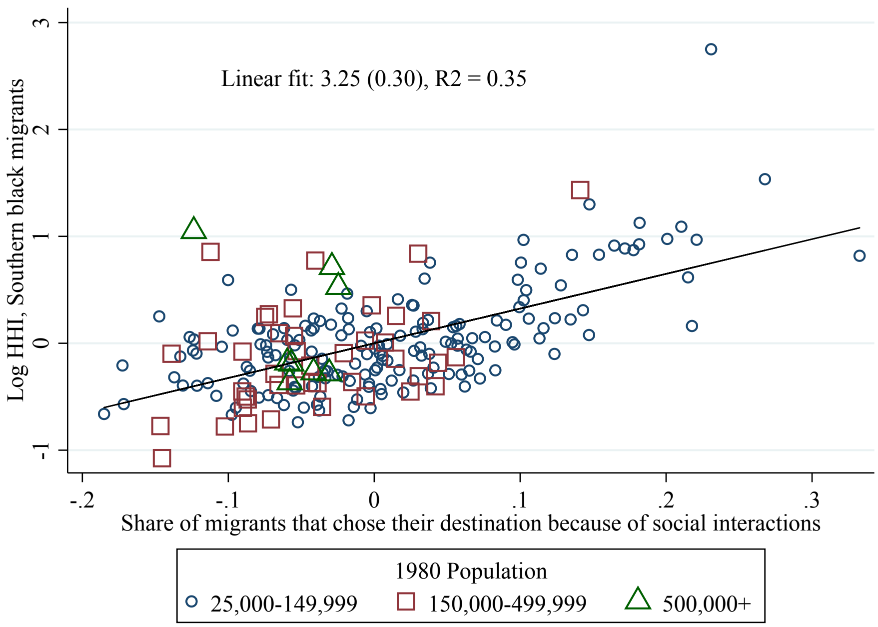


Notes: We estimate the share of migrants that chose their destination because of social interactions using a structural model, as described in the text.
Source: Duke SSA/Medicare data

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



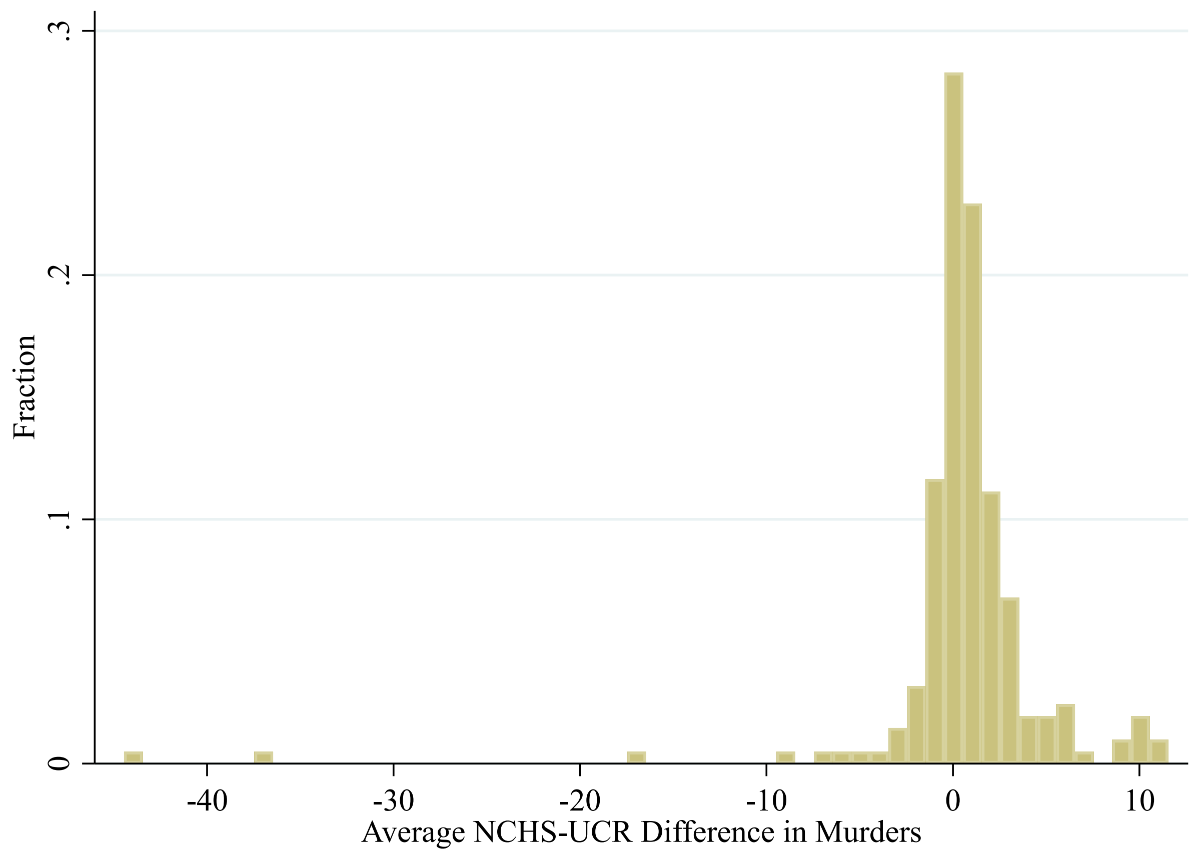
(a) Unconditional



(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. Sources: 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)