Social Determinants of Inequality

Steven N. Durlauf
UW-Madison

a. Basic Ideas

Basic Ideas

1. Group Membership

- Individual beliefs, preferences, and opportunities are conditioned by group memberships.
- This dependence typically takes the form of complementarities:
- The likelihood or level of action by one person increase with respect to the behavior (or certain characteristics) of others.

Basic Ideas

2. Memberships evolve in response to those interactions.

- Groups (non-overlapping subsets of the population) stratify along characteristics which affect outcomes.
- Economic and social (typically ethnic) segregation result in neighborhoods, schools, etc.

Basic Ideas

3. Persistent intergenerational inequality and poverty result as:

- Individuals face different interactions and environments over their lives.
- Stratification of society affects both parents and children.

Basic Ideas

Social interaction models thus study the interplay of social forces which influence individual outcomes and individual decisions which determine group memberships and hence social forces.

Key Features of this Approach

1. Individual incentives and social structure meld into a more general explanation of individual behavior. From the perspective of economics, introduction of better sociology; from the perspective of sociology, better economics!
2. Approach explicitly incorporates incomplete markets and other deviations from baseline neoclassical theory of choice.
3. Aggregate behaviors such as crime or non-marital fertility rates emerge through the interactions within a heterogeneous population

Examples of Social Influences

1. Peer group effects
2. Role models
3. Social norms
4. Social learning

Phenomena where Social Interactions Plausibly Matter

1. Fertility
2. Education
3. Employment
4. Health
5. Language

Types of Groups

1. Endogenous
    - Neighborhoods
    - Firms
    - Schools
2. “Exogenous”
    1. Race/Ethnicity
    2. Gender
    3. Religion

b. Theory

Complementarity

Complementarity is a fundamental concept in the social interactions literature because it characterizes how the attributes of one individual affect the decision problems of others. The term attribute is used as a catchall for two different channels by which these interpersonal effects can manifest themselves.

One channel involves the way an individual's characteristics affects others. The other channel involves the effects that one person's choices have on others.

Complementarity

Consider a group of *I* individuals. Each individual is associated with a *K*-length vector of attributes $x_i \in R^K$. Each vector of attributes produces a payoff

$$\Phi_i(x_i, x_{-i})$$

where x_{-i} denotes the vector composed of attributes of members of the group other than *i*. For each *i, a* argument, this function is assumed to be twice differentiable.

Complementarity

Definition: Complementarity

A payoff function is said to exhibit complementarity between agent *i*'s *a*-th attribute and the agent *j*'s *b*-th attribute if

$$\frac{\partial^2 \Phi_j\left(x_j, x_{-j}\right)}{\partial x_{ia} \partial x_{jb}} \geq 0 \quad j \neq i$$

Complementarity

Complementarities thus restrict how marginal payoff changes in *i*'s attributes are affected by changes in the attributes of others.

$$\frac{\partial^2 \Phi_j\left(x_j, x_{-j}\right)}{\partial x_{ia}\partial x_{jb}} = \frac{\partial \dfrac{\partial \Phi_j\left(x_j, x_{-j}\right)}{\partial x_{jb}}}{\partial x_{ia}} .$$

Complementarity

Complementarities are distinct from positive spillovers.

Definition: Positive Spillovers

The payoff function is said to exhibit positive spillovers from attribute *a* of agent *i* to agent *j* if

$$\frac{\partial \Phi_j\left(x_j, x_{-j}\right)}{\partial x_{ia}} \geq 0, \quad j \neq i$$

Positive spillovers tell us about the payoff level changes for one agent associated with changes in the attributes of others.

Social vs. Individual Determinants of Outcomes

“Standard” Model of Individual Choice

ω_i = choice of behavior of individual *i*

Ω_i = constraint set

X_i = observable individual characteristics,

ε_i = unobservable individual characteristics (to the modeler)

Social vs. Individual Determinants of Outcomes

Algebraically, the individual choices represent solutions to:

$$\max_{\omega_i \in \Omega_i} V\left(\omega_i, X_i, \varepsilon_i\right)$$

$$\text{such that } \Omega_i = \Omega\left(X_i, \varepsilon_i\right)$$

Social Interactions Approach

$g(i) =$ group of individual i

$Y_{g(i)} =$ characteristics of $g(i)$

$\mu_i^e(\omega_{-i}) =$ subjective beliefs individual i has concerning behavior of others in his group, where

$$\omega_{-i} = (\omega_1, \ldots \omega_{i-1}, \omega_{i+1}, \ldots \omega_I)$$

Social Interactions Approach

In this case, choice is described by

$$\max_{\omega_i \in \Omega_i} V\left(\omega_i, X_i, Y_{g(i)}, \mu_i^e\left(\omega_{-i} \middle| Y_{g(i)}\right), \varepsilon_i\right)$$

$$\text{such that } \Omega_i = \Omega\left(X_i, Y_{g(i)}, \mu_i^e\left(\omega_{-i}\right), \varepsilon_i\right)$$

In words, preferences, constraints, beliefs depend on memberships.

Key Theoretical Properties

1. Multiple Equilibria
2. Social Multipliers
3. Phase Transition

Key Theoretical Properties

1. Properties are "universal," although they of course depend on parameter values.
2. These models are intrinsically nonlinear

Identification

1. How can social mechanisms be uncovered from social science data? This is the identification problem.
2. Manski (1993) launched the current literature. Blume et al. (2011) synthesizes.

In social interactions contexts, there are three types of identification problems.

1. Classical identification: Assuming one has “properly” accounted for the error structure in choice model, can different types of social interaction effects be disentangled?
2. Self-Selection: How does self-selection into neighborhoods affect standard econometric procedures and how can self-selection be accounted for.
3. Unobserved Group-Level Variables: Omitted common factors may confound social interactions.

Example

To understand the difficulties that exist in empirically identifying a causal role for groups in determining individual outcomes, it is useful to consider a specific example. Suppose that a researcher wishes to evaluate the effect of high poverty neighborhoods on teenage educational attainment, such as completion of high school. The crude fact leading one to believe such an effect is present is a bivariate relationship between high poverty neighborhoods and low educational attainment.

Possible Explanations: 1

High poverty neighborhoods are disproportionately composed of adults with low labor market aspirations (as compared to more affluent communities). If parents transmit low aspirations to their own children, and if these low aspirations adversely influence educational attainment, then poor neighborhoods will exhibit lower educational attainment than richer ones, without any causal influence from the neighborhood to the individual.

Possible Explanations: 2

Families in high poverty neighborhoods are less likely to be able to finance post-secondary education, hence the opportunities for further education generated by a high school diploma are not available to many teenagers in these neighborhoods.

Possible Explanations: 3

Teacher quality is lower in high poverty neighborhoods as better teachers should to be employed in schools in communities with lower crime rates.

Possible Explanations: 4

High poverty neighborhoods possess a relatively high concentration of individuals who, despite graduating from high school, failed to achieve success in the labor market. Hence teenagers observing the economic benefits of graduation will not observe examples where graduation had much of a payoff.

Possible Explanations: 5

Teenagers are influenced by the aspirations of role models in the community where they live. If the role models in a neighborhood have low labor market aspirations, then this will depress the educational achievements of children in the neighborhood.

Possible Explanations: 6

Teenagers in high poverty neighborhoods are, due to local public finance, higher crime, etc. provided lower quality schools than students in other communities.

Possible Explanations: 7

Teenagers are influenced by the behaviors of their peers through a "primitive" desire to conform to others. In a given community, high and low levels of educational attainment are self-reinforcing as the educational effort of a given teenager reflects his preference to seem like "one of the crowd."

Possible Explanations: Identification

Each of these explanations will produce the same correlations between low individual educational attainment and neighborhood poverty, but each is based on a different causal mechanism.

The statistical question is whether these different explanations can be disentangled in a given data set.

Possible Explanations: Identification

1. explanations 1 and 2 attribute the correlation of neighborhood poverty and low individual educational attainment to self-selection

2. explanation 3 is an example of an unobserved group level effect

3. explanations 4, 5, 6 are examples of contextual effects as the distribution of educational levels and incomes among older members of the community are affecting current behaviors;

4. explanation 6 is an example of an endogenous effect as it is based on contemporaneous interdependences in behavior.

Identification – Linear Models

Within a group, assume individual behavior obeys:

$$\omega_i = \kappa + \gamma x_i + \delta \sum_j c_{ij} x_j + \phi \sum_j a_{ij} E(\omega_j | x) + \varepsilon_i$$

x_i: observable (to all group members) heterogeneity
ϵ_i: unobservable (except to *i*) heterogeneity
c_{ij}: contextual effects intensity weights; form sociomatrix *C*
a_{ij}: endogenous effects intensity weights; form sociomatrix *A*

Rows of sociomatrices sum to 1; weights are nonnegative.

Identification – Linear Models

Comment: Sociomatrices and Social Networks

- Sociomatrices are sometimes constructed from social networks.
- Social networks data identify who is connected to whom.
- Assumption is often made that individuals react to averages of others in network.

Important: Social network data does not directly measure intensities.

Identification – Linear Models

Identification Problem:

Recover $(\kappa, \gamma, \delta, \phi)$ from joint distribution of (ω, x).

What determines whether these parameters are identified?

1) Properties of unobserved heterogeneity

2) Prior knowledge on sociomatrices *C* and *A*

Identification – Linear Models

Classic identification problem

Assume ϵ independent of x.

Identification problem for this model is in fact a variant of classical identification question for linear simultaneous equations.

Link between General Linear Social Interactions Model and Literature

This general framework includes common specifications as special cases.

1: Linear-in-Means Model

The most common linear interaction model is the linear-in-means model, in which the population is partitioned into nonoverlapping groups g, no intergroup social interactions are present, and unweighted averages summarize intragroup interactions.

Identification – Linear Models

Letting n^g denote the population size of group g, these restrictions may be expressed as

$$c_{ij} = \frac{1}{n^g} \text{ if } i, j \in g,$$

$$a_{ij} = \frac{1}{n^g - 1} \text{ if } i, j \in g,$$

$$c_{ij} = a_{ij} = 0 \text{ if } i \in g, j \notin g$$

Failure of identification known as reflection problem (Manski (1993)).

Identification – Linear Models

The Manski (1993) reflection problem for nonidentification holds for the large sample limit of this model.

$$\omega_i = \frac{\gamma}{1+\phi} x_i + \frac{\delta}{(1+\phi)} \bar{x}^g + \frac{\phi}{(1+\phi)} E\left(\bar{\omega}^g \,|\, x\right) + \frac{1}{1+\phi} \varepsilon_i$$

Identification – Linear Models

2. Neighborhood Generalizations of the Linear-in-Means Model

A second class of models, typically employed when data on network structure are available, associates with each agent i a group of others to whom he is directly connected. These others define the neighborhood. The effect of the neighborhood on an individual mimics the original linear-in-means model.

$$c_{ij} = \frac{1}{n^h} \text{ if } j \in h,$$

$$a_{ij} = \frac{1}{n^h} \text{ if } j \in h,$$

$$c_{ij} = a_{ij} = \frac{1}{n^h} = 0 \text{ if } j \notin h$$

Identification – Linear Models

Result 1: (Blume, Brock, Durlauf, Jayaraman (2015))
Without additional prior information on C and A, identification fails.

This is not a surprise.

Identification – Linear Models

If one replace the expectations in the linear model with realizations, then the model is essentially the standard unidentified simultaneous equations model.

$$\omega_i =$$

$$\kappa + \gamma x_i + \delta \sum_j c_{ij} x_j + \phi \sum_j a_{ij} \omega_j$$

$$+\varepsilon_j - \phi \sum_j a_{ij} \left(E\left(\omega_j \middle| x \right) - \omega_j \right)$$

Since one would (intuitively) replace ω_j with $E(\omega_j)$ in two stage least squares (for example), identification failure is obvious.

Identification – Linear Models

Result 2: If C and A are known a priori, then identification holds generically.

Generically means "for almost all sociomatrices" in a mathematically precise sense.

Intuition: The number of unknown parameters in each individual's behavioral equation is reduced to 4 and these parameters are the same across equations.

This explains why variations of Manski's unidentified model were identified and why models in which C and A are constructed from network data are identified.

Identification – Linear Models

Result 3: Identification may hold under partial knowledge of sociomatrices

Case 1: C known, A unknown

Economics may justify prior knowledge of C.

Case 2: number of included ω_j's (nonzero weights) no larger than number of omitted x_j's.

Analogy of the classic order condition.

IMO: econometric literature has erred in not focusing on assumptions/possibilities frontier with respect to sociomatrices

Identification – Linear Models

Self-Selection

Self-selection matters because $E(\omega_i|x) \neq 0$

$$\omega_i = \kappa + \gamma x_i + \delta \sum_j c_{ij} x_j + \phi \sum_j a_{ij} E\left(\omega_j \middle| x\right) + E\left(\varepsilon_i \middle| x\right) + v_i$$

where $E(\nu_i|x) = 0$.

Strategies: instrumental variables, control functions (Heckman selection correction and generalizations)

Advantage of latter: can facilitate identification! Why? Endogeneity of social structure encodes information on social interactions.

Identification – Linear Models

Fixed Effects and Correlation in Unobserved Heterogeneity

Not amenable to behavioral modelling.

IMO harder than self selection.

Program Evaluation and Treatment Effect Approaches

Some evidence on social effects is indirect, deriving from policies which alter social structure.

Quasi-experiments that alter the location of poor families is a leading example.

The logic of these exercises is straightforward: implement a policy randomly, see what happens.

Program Evaluation and Treatment Effect Approaches

Standard problem: policy may create opportunities (use available job training, exercise a voucher), and so this needs to be accounted for in measuring effects.

- Intent To Treat (ITT) compares those eligible for program with control group

- Treatment on Treated (TOT) compares those who use program with control group

Program Evaluation and Treatment Effect Approaches

(Trivial) Observations

1. Treatment effects do not measure social interactions; they measure the consequences of the treatments.

2. Useful in process of abduction.

3. No prioritization of types of evidence appropriate.

Example: Binary Choice (Brock and Durlauf (2001))

Consider a population of individuals who are members of a common group g. Our objective is to probabilistically describe the individual choices of each i where $\omega_i \in \{-1,1\}$, and thus the collective set of choices ω.

From the perspective of theoretical modeling, it is useful to distinguish between three sorts of influences on individual choices. These influences have different implications for how one models the choice problem.

Example: Binary Choice (Brock and Durlauf (2001))

Assume utility is

$$u_i = h_i \omega_i + J\overline{\omega}_i + \epsilon_i(\omega_i)$$

h_i is the private deterministic utility associated with a choice

$\overline{\omega}_i$ is the private deterministic utility associated with a choice

$\epsilon_i(\omega_i)$ is a choice specific random utility term; iid across agents

Agents know all h_i's, but only observe their own $\epsilon_i(\omega_i)$.

Example: Binary Choice (Brock and Durlauf (2001))

If $\Pr(\epsilon_i(-\omega_i) - \epsilon_i(\omega_i) \leq z) = \frac{1}{1+\exp(-\beta \quad)}$,

Then $\Pr(\omega_i) = \frac{\exp(\beta h_i \omega_i + J m_{-i})}{\exp(\beta h_i \omega_i + J m_{-i}) + \exp(-\beta h_i \omega_i - J m_{-i})}$

And

$$E(\omega_i) =$$

$$\frac{\exp(\beta h_i \omega_i + J m_{-i}) - \exp(-\beta h_i \omega_i - J m_{-i})}{\exp(\beta h_i \omega_i + J m_{-i}) + \exp(-\beta h_i \omega_i - J m_{-i})} =$$

$$\tanh(\exp(\beta h_i \omega_i + J m_{-i}))$$

Properties of Model

1. Interplay of distribution of private incentives h_i, social interaction parameter J and degree of dispersion of unobserved heterogeneity, β.
2. Conditional on others, if J large enough, multiple equilibria.
3. Conditional on multiple equilibria at initial set of values, decreasing β sufficiently (increasing dispersion) will eliminate multiplicity.
4. Conditional on multiple equilibria at initial set of values, sufficiently large increases in h_i's will eliminate multiplicity,

When $J = 0$ and $h_i = x_i \gamma$, model reduces to textbook logit model.

Formation of Social Structure: Theory

A comprehensive theory of social influences on behavior requires understanding the determinants of social structure.

Many of the existing models of social structure can be understood as involving either groups or social networks. These terms are not exclusive, but the logic of the formation process differs across them.

I restrict discussion to groups.

Group Formation

Marriages, neighborhoods, schools, firms all examples in which individuals are members of (within a category), nonoverlapping groups.

Distinct microeconomic foundations (of course) underlie of these, A common question for each context is the extent to which assortative matching (like matched with like) occurs in equilibrium.

Group Formation

The presence of contextual effects Y_n and endogenous effects $\mu_i^e(\omega_{n,-i})$ in individual payoff functions implies that a complete theory of neighborhood effects must account for how neighborhoods are formed in the presence of these effects. From the perspective of an abstract choice problem, individual neighborhood choices may be thought of as

$$(13) \quad \max_{n \in N} V^*(x_i, Y_n, \rho_n)$$

where $V^*(x_i, Y_n, \rho_n)$ is the expected utility associated with neighborhood n, and ρ_n denotes any additional variables that affect this payoff; typically, this will be the rental price to residence.

Group Formation

Notice that this payoff calculation requires that agents form beliefs about which equilibrium will emerge in a neighborhood when multiple equilibria exist. An equilibrium in an endogenous neighborhoods model is a configuration of agents across neighborhoods such that each agent solves (13) and the resulting values of Y_n are those that are generated by this configuration. Put differently, the neighborhood choices of individual agents help determine the effects in a neighborhood, so location decisions need to exhibit self-consistency.

Group Formation

The key theoretical feature of interest in these types of models concerns how individuals with different attributes are allocated across neighborhoods; this allocation will determine the extent to which different neighborhoods will exhibit different contextual effects; without such differences, the only way that neighborhoods may affect inequality is via multiple equilibria. The most common attribute that is studied in this literature is income, although other attributes have been considered, as is discussed below.

Group Formation

Much of the interest in neighborhood configurations, in turn, focuses on the extent to which neighborhoods are stratified by income or other attributes. Neighborhoods are said to be stratified with respect to an attribute x_i if it is the case that the supports of the intra-neighborhood distributions of x_i do not overlap except at endpoints. Stratification by income, for example, provides a basis for understanding persistence in economic status across generations: poor families are consigned to poor neighborhoods, whose effects make it more likely their children are poor, etc.

Neighborhoods

Neighborhood models can provide insights into the extent to which one observes residential income segregation. Bénabou (ReStud 1996) gives a general set of conditions for stratification.

Neighborhoods

The population consists of a continuum of agents with associated measure l who may live in one of two neighborhoods, each of size $l/2$ and denoted as A and B.

Residence in a neighborhood entails the payment of s rent $\rho_j, j = A, B$; these rents go to an absentee landlord.

There are two types of agents, denoted by whether they are associated with incomes Y^{high} or Y^{low}. Let θ denote the percentage of Y^{high} agents.

Neighborhoods

Payoff to a given neighborhood:

$$V(Y_i, \overline{Y_n}, \rho_n)$$

When do neighborhoods stratify by income? Consider how individual agents are willing to trade off the rental price ρ_n against $\overline{Y_n}$. Specifically, for each individual, one may define the function R$(Y_i, \overline{Y_n}, \rho_n)$ which characterizes the marginal tradeoff between ρ_n and $\overline{Y_n}$ that leaves utility unchanged at a given initial V_0 .

Neighborhoods

This tradeoff implicitly describes how different agents will react to changes in relative neighborhood rental prices.

This function may be derived by differentiating the payoff function with respect to ρ_n and $\overline{Y_n}$ when utility is held constant, i.e.

$$V_{\bar{Y}_n}\left(Y_i,\bar{Y}_n,\rho_n\right)d\bar{Y}+V_{\rho_n}\left(Y_i,\bar{Y}_n,\rho_n\right)d\rho_n=0\Rightarrow$$

$$R\left(Y_i,\bar{Y}_n,\rho_n\right)=\frac{d\rho_n}{d\bar{Y}_n}=-\frac{V_{\bar{Y}_n}\left(Y_i,\bar{Y}_n,\rho_n\right)}{V_{\rho_n}\left(Y_i,\bar{Y}_n,\rho_n\right)}$$

Neighborhoods

The properties of $R(Y_i, \overline{Y_n}, \rho_n)$ determine whether neighborhoods are stratified in equilibrium.

Proposition (Bénabou (1996))

If $R(Y_i, \overline{Y_n}, \rho_n)$ is increasing in $\overline{Y_n}$, then the only stable equilibrium configuration of families is one that is stratified.

This is intuitive: the more affluent need to be more willing to pay for high income neighbors than the less affluent for stratification to hold. What might lead to such a willingness to pay difference?

Matching

Complementarity and optimal matching (Becker (1973))

Consider a population of *N* men and *N* women.

Suppose that the product of a marriage between man *u* and woman *v* depends on scalar characteristics m_u and w_v of the man and woman respectively.

Suppose that the product of a given match is $\Phi(m, w)$ and that this function is increasing in both arguments.

Matching

Proposition: Optimality of assortative matching in the Becker marriage model

If $\frac{\partial^2 \Phi(m,w)}{\partial m \partial w} \geq 0$ then assortative matching maximizes the sum of products across marriages.

Matching

Proof:

For assortative matching to be *in*efficient, there must exist pairs $(\underline{m}, \overline{w})$ and $(\overline{m}, \underline{w})$ such that $\overline{m} > \underline{m}$, $\overline{w} > \underline{w}$, and

$$\Phi(\underline{m},\bar{w}) + \Phi(\bar{m},\underline{w}) > \Phi(\bar{m},\bar{w}) + \Phi(\underline{m},\underline{w})$$

But this can be rewritten as

$$\Phi(\underline{m},\bar{w}) - \Phi(\underline{m},\underline{w}) - \Phi(\bar{m},\bar{w}) + \Phi(\bar{m},\underline{w}) = -\int_{\underline{m}}^{\bar{m}} \frac{\partial^2 \Phi(m,w)}{\partial m \partial w} > 0$$

The last inequality is inconsistent with complementarity.■

Matching

For firms, one might simply say that one has *NK* agents with scalar skills characteristics a_i who must be organized into *K*-tuples, each of which produces some payoff.

In this case, equating supermodularity with the efficiency of assortative matching also requires *permutation invariance* (Durlauf and Seshadri (2003)): if *a* is a *K*-tuple of characteristics and a' is a permutation of a, then

$$\Phi(a) = \Phi(a') \; .$$

Matching

Equity/Efficiency Tradeoffs

Assuming that returns to a group are not redistributed, then the maximization of average output also maximizes the gap between the highest and lowest groups output.

In my opinion, a great example of an equity/efficiency tradeoff.

Matching

A Caveat

Assortative matching can be dynamically inefficient even if every static function of interest exhibits complementarities.

This following numerical example, taken from Durlauf and Seshadri (in progress) illustrates general ideas.

Matching

Consider 4 agents who are tracked over 3 periods.

Each agent is associated with a period-specific characteristic ω_{it}; for concreteness assume that it is educational attainment. The distribution of period-0 values is 10, 10, 20, 20.

In periods 0 and 1, the agents are placed in two person groups, Think of these as classrooms. Pairings can differ between periods 0 and 1.

Matching

The value of ω_{it+1} is determined by ω_{it} and $\omega_{i't}$, the value for the agent with whom he is paired. The policymaker chooses the pairings.

The objective of the policymaker is to maximize $\overline{\omega}_2$, i.e. the average characteristic in period 2.

Matching

Suppose that one step ahead transformation function for agent characteristics is

$$\phi\left(\omega_{it+1}\middle|\omega_{it},\omega_{i't}\right)=f_1\left(\omega_{it}\right)+f_2\left(\omega_{it},\omega_{i't}\right)$$

$$f_1\left(\omega_{it}\right)=0 \text{ if } \omega_{it}\leq 9; .9\omega_{it} \text{ if } 9<\omega_{it}\leq 10; \omega_{it} \text{ if } 10<\omega$$

$$f_2\left(\omega_{it},\omega_{i't}\right)=\max\left\{\varepsilon\left(\omega_{i't}-10\right)\omega_{it},0\right\}+\eta\omega_{it}\omega_{i't}$$

This function exhibits strict increasing differences (I do not use the term complementarities because the function is not differentiable everywhere.)

Matching

Proposition:

If $\epsilon < 0.03$, then for η sufficiently small, then $\overline{\omega}_2$ is maximized by negative assortative matching in period 0 and assortative matching in period 1.

Matching

What is the general idea from the example?

Assortative matching is efficient when one wants to maximizes the average of something.

For this problem, the period 0 rule should not maximize $\overline{\omega}_1$; it should choose the feasible distribution of ω_{i1}'s which is best for maximization of $\overline{\omega}_2$. This distribution depends on higher moments of the period one distribution than $\overline{\omega}_1$.

The shift from negative assortative matching to assortative matching in the efficient sorting rule has "real world" analogs, e.g. mixed high schools and stratified colleges.

Matching

Equilibrium

It is one thing to ask how agents should be configured by a social planner who maximizes the sum of payoffs across groups. A distinct question is how agents will organize themselves in a decentralized environment. In the marriage case, Becker shows that the efficient (in terms of aggregate output) equilibrium in terms of male/female matches will occur when marriages are voluntary choices, so long as marital partners can choose how to divide the output of the marriage. This division of marital output is the analogy to market prices that would apply to labor market models in which workers are sorted to firms. Similarly, one can show that wages can support the efficient allocation of workers when increasing returns are absent.

Matching

Unequal Division can Promote Equality

Second, suppose that marital output cannot be arbitrarily divided; assume for simplicity that the output is nonrival so that both marriage partners receive it. (Parents will understand). Further, rule out transfers between the partners.

The ruling out of transfers is important as it means, in essence, that neither member of the marriage can undo the nonrival payoff of the marriage.

Matching

Under these assumptions, assortative matching will still occur, even when it is socially inefficient. The assumption that $\Phi(m, w)$ is increasing in both arguments is sufficient to ensure that the highest m_i will match with the highest w_j, etc.

This indicates how positive spillovers can create incentives for segregation by characteristics even when the segregation is socially inefficient. Durlauf and Seshadri (2003) hint at this possibility; it is systematically and much more deeply addressed in Gall, Legros, and Newman (2015). It is worth thinking about the possibility of inefficient segregation in various contexts. Residential neighborhoods and college student bodies are possible examples beyond marriage.

c. Neighborhoods and Schools

Wodtke, Harding, and Elwert (2015)

Neighborhood Effect Heterogeneity by Family Income and Developmental Period (AJS, forthcoming 2016)

Research question: How does timing of exposure to disadvantaged neighborhoods during childhood versus adolescence affect high school graduation? And how do these effects vary across income levels?

- Estimates structural nested mean model via two-stage regression with residuals
- Uses PSID data on 6,137 children from childhood through adolescence

Wodtke, Harding, and Elwert (2015)

- Neighborhood effects are detrimental to education attainment, and much more so for poor relative to non-poor families
- For poor blacks, high school graduation rates are 25 p.p. lower for students living in the most disadvantaged neighborhoods compared to least disadvantaged neighborhoods. For poor whites it is 8 p.p.
- For non-poor blacks and non-poor whites, it is 8 p.p. and 3 p.p. respectively.

Wodtke, Elwert, and Harding

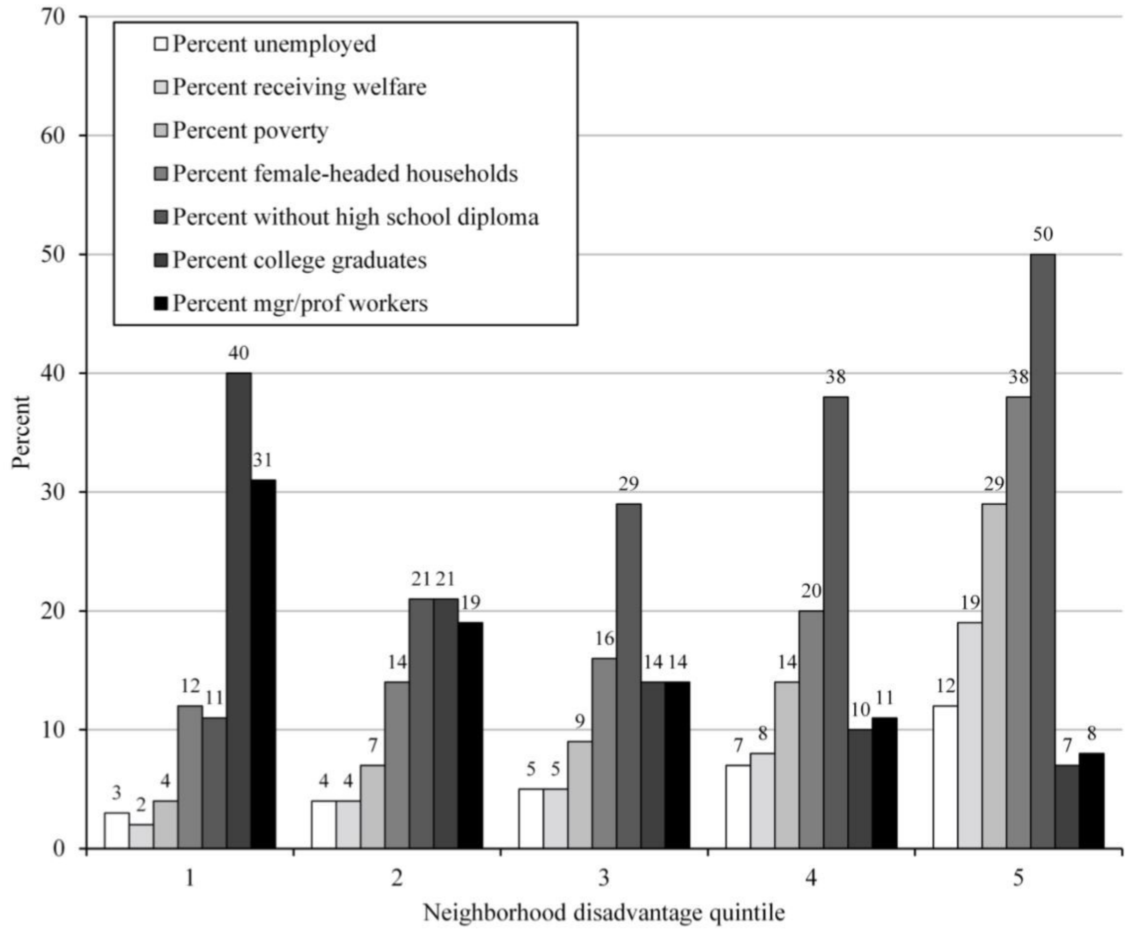


Fig. A1.— Neighborhood socioeconomic characteristics by disadvantage index quintile

Black students

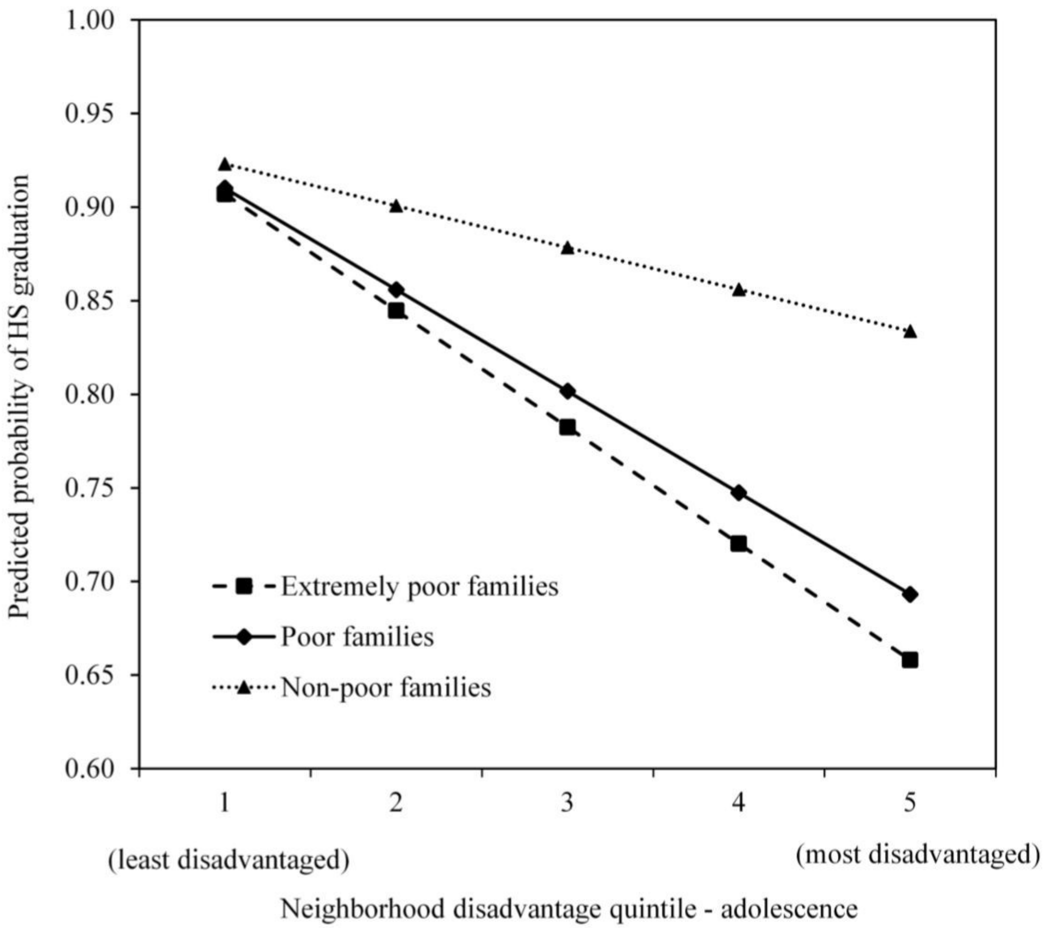


FIG. 4.—Predicted probability of high school graduation by adolescent exposure to neighborhood disadvantage and family poverty history, black respondents. Probabilities are computed with childhood treatment set to residence in a third-quintile, or “middle class,” neighborhood.

White students

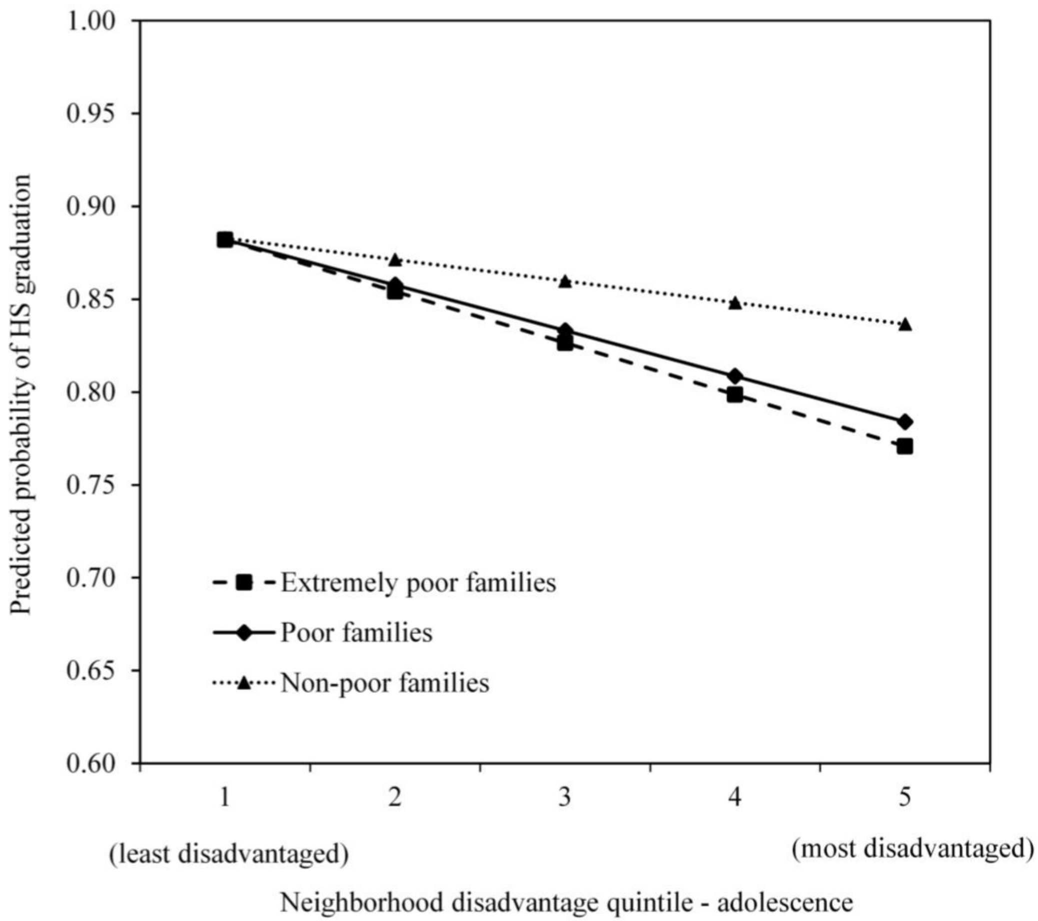


FIG. 5.—Predicted probability of high school graduation by adolescent exposure to neighborhood disadvantage and family poverty history, white respondents. Probabilities are computed with childhood treatment set to residence in a third-quintile, or "middle class," neighborhood.

d. Identity

Ainsworth-Darnell and Downey

Assessing the Oppositional Culture Explanation for Racial/ethnic Differences in School Performance (Am. Soc. Rev., 1998)

Variable name	Description	Metric	Mean	S.D.	Alpha
Good student	Do you think that other students see you as a good student?	0 = Not at all; 2 = Very much.	1.18	.60	—
Popularity among Peers					
Popularity	Do you think that other students see you (a) as popular; (b) as socially active; (c) as part of the leading crowd?	0 = Not at all to all three questions; 6 = Very much to all three questions.	2.88	1.36	.74

Note: High School Sophomores from the National Education Longitudinal Study, 1990. Valid cases range from 11,937 to 16,972. “Somewhat” is the omitted category between “Not at all” and “very much.”

Ainsworth-Darnell and Downey

Independent Variable	Popularity Model 1	Model 2	Model 3	Model 4	Model 5
African American	.010 (.029)	.042 (.032)	.079* (.032)	.024 (.031)	–.027 (.036)
Asian American	–.177* (.074)	–.164* (.076)	–.170* (.076)	–.235** (.075)	–.167 (.093)
Family income (in $1,000s)	—	—	.020*** (.003)	.018*** (.003)	.018*** (.003)
Parental occupational prestige	—	—	.094*** (.016)	.083*** (.016)	.083*** (.016)
Parental education	—	—	.006 (.011)	–.010 (.011)	–.009 (.011)
Does youth think that other students see him/her as a good student?					
Very good	—	—	—	.373*** (.026)	.340*** (.029)
Not at all good	—	—	—	–.542*** (.039)	–.534*** (.040)
African American × "Very" good student	—	—	—	—	.227*** (.067)
Asian American × "Very" good student	—	—	—	—	–.154 (.156)
African American × "Not at all" good student	—	—	—	—	–.089 (.137)
Asian American × "Not at all" good student	—	—	—	—	–.368 (.446)
R^2	.000	.004	.011	.040	.040

Note: Numbers in parentheses are standard errors. White is the omitted category for race; "somewhat good" is the omitted category for good student. Models 2 through 5 also include the following control variables: sex, region, number of siblings, parents' age, family structure, and urban/suburban/rural location of school.
*p<.05; **p < .01; ***p< .001 (two-tailed tests)

Fryer and Torelli

An Empirical Analysis of Acting White (J. Pub. Econ., 2010)

Data from Add Health survey (a nationally representative sample of US students entering grades 7 through 12 in the 1994–1995 academic year):

- Grades: Most recent GPA on a 4.0 scale (A=4, D or Lower=1), averaged across four subjects.
- Friends: Each respondent nominates up to five friends of each sex who are themselves surveyed

Fryer and Torelli

Social status: the (unique) index that satisfies monotonicity, homogeneity, and linearity (Echenique and Fryer, 2005)

- Monotonicity: an increase in individual v's same-race connections implies an increase in v's social status s_v.
- Homogeneity (a normalization): when all individuals of a given race have exactly d same-race friends, they all have $s = d$
- Linearity: individual v's social status s_v should be the average among v's same-race friends, relative to the average social status of the individuals in v's connected component (all v's friends, friends of friends, etc.)

Qualitatively, linearity implies that v's social status depends on the social status of her friends, and a decrease in the social status of one of v's friends will affect v less if v is in a high social status component.

Fryer and Torelli

An Empirical Analysis of Acting White (J. Pub. Econ., 2010)

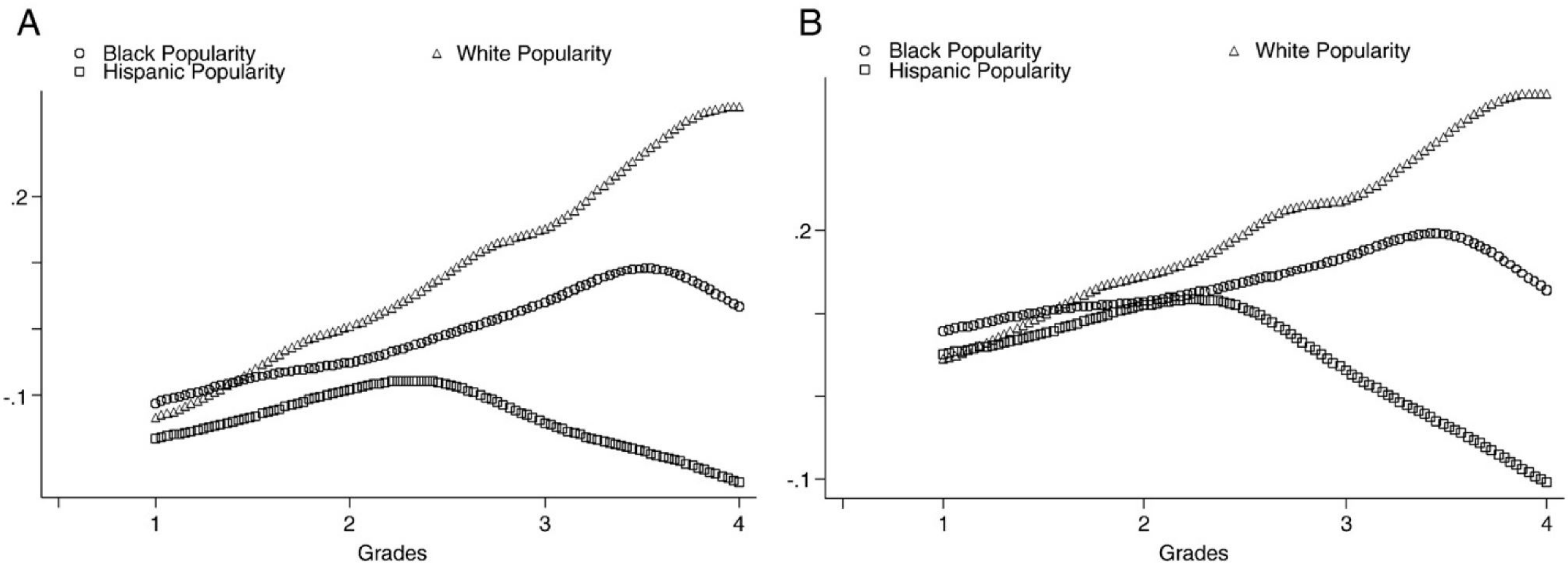


Fig. 1. A: Spectral popularity and grades by race, raw data. B: Spectral popularity and grades by race, with controls.

Akerlof and Kranton

- Members of a social group (exogenously determined) may have negative associations with certain activities
- Undertaking such an activity may result in a utility penalty from “loss of identity”
- Members of such a group may thus avoid activities despite apparent economic benefits

Akerlof and Kranton

Implications:

- Can rationalize group differences in e.g. antisocial behavior, educational attainment
- Residential job training programs may be more successful (Stanley, Katz, and Krueger1998) because they allow participants to avoid interaction with their original group
- May explain success of programs which separate students adhering to different identities (e.g. Central Park East Secondary School in East Harlem)
- "…Legal equality may not be enough to eliminate racial disparities"

d. Discrimination and Stigma

Discrimination

Why (Many) Economists are Skeptical that Discrimination is a First Order Determinant of Black/White Differences (following Heckman (2008))

1. Strongest empirical evidence is at micro level; equilibrium implications unclear. (Becker discrimination model).

2. Identification problems with respect to taste-based versus statistical discrimination.

3. Many measures of discrimination involve interpretation of ethnicity-indexed differences in residuals.

4. Educational gaps seem first-order in explaining gaps.

Pager

The Mark of a Criminal Record (2003 AJS)

Audit methodology: matched pairs of individuals (called testers) to apply for real job openings in order to see whether employers respond differently to applicants on the basis of selected characteristics

Two pairs of testers

- Black pair audited 200 employers, white pair audited 150 employers
- All testers claimed to have high school diploma
- Each tester alternated between no criminal record and a felony drug conviction (possession with intent to distribute, cocaine), 18 months (served) prison time
- Testers assigned favorable work histories with steady work experience in entry-level jobs and nearly continual employment (until incarceration). In the job prior to incarceration (and, for the control group, prior to the last short-term job), testers reported having worked their way from an entry-level position to a supervisory role

Pager

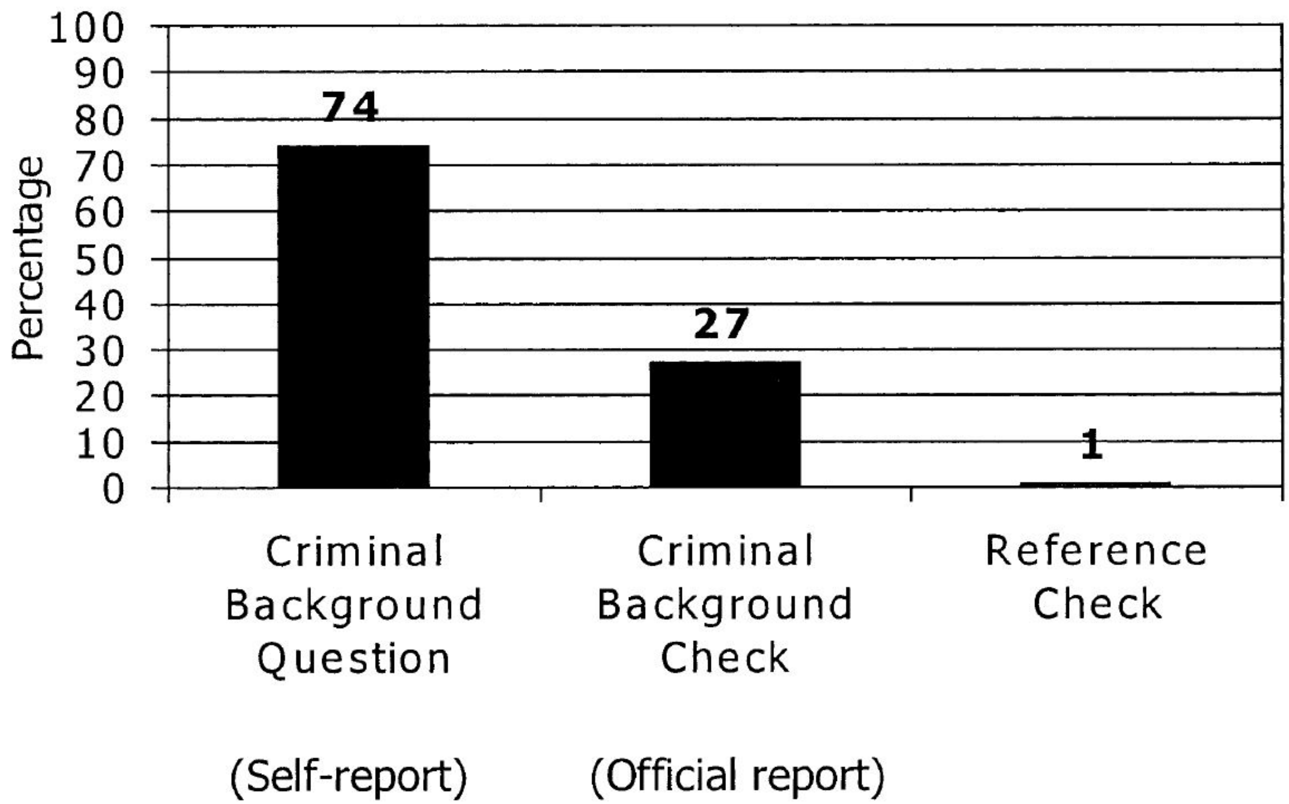


FIG. 4.—Background checks

Pager

Greater disparity for applications that asked about criminal history

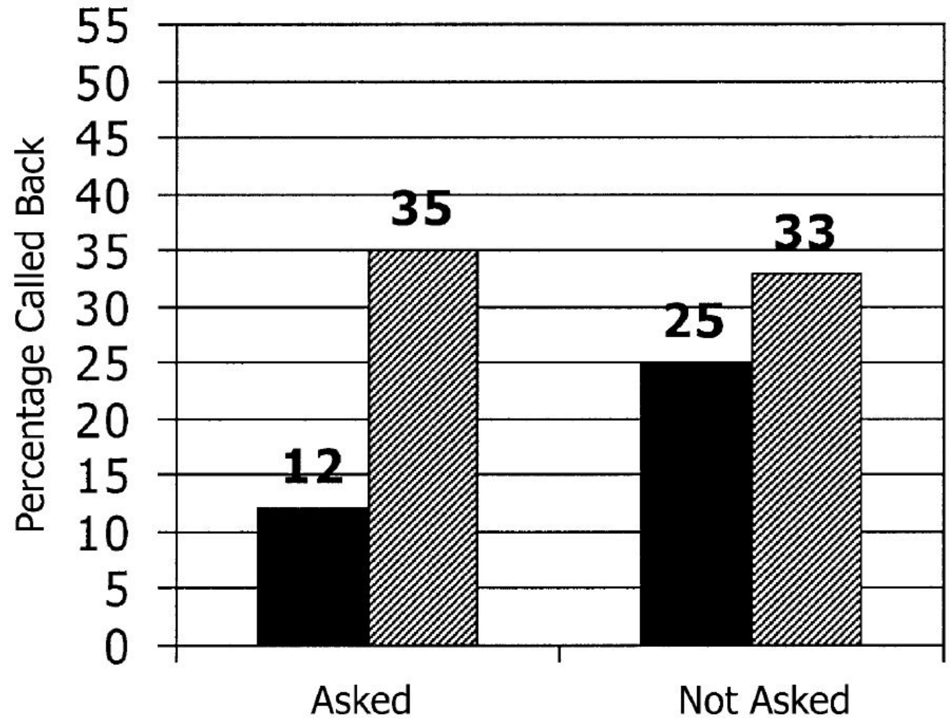


FIG. A1.—Differences by whether criminal history information was solicited: black bars represent criminal record; striped bars represent no criminal record.

Pager

Greater disparity for applications submitted without personal contact

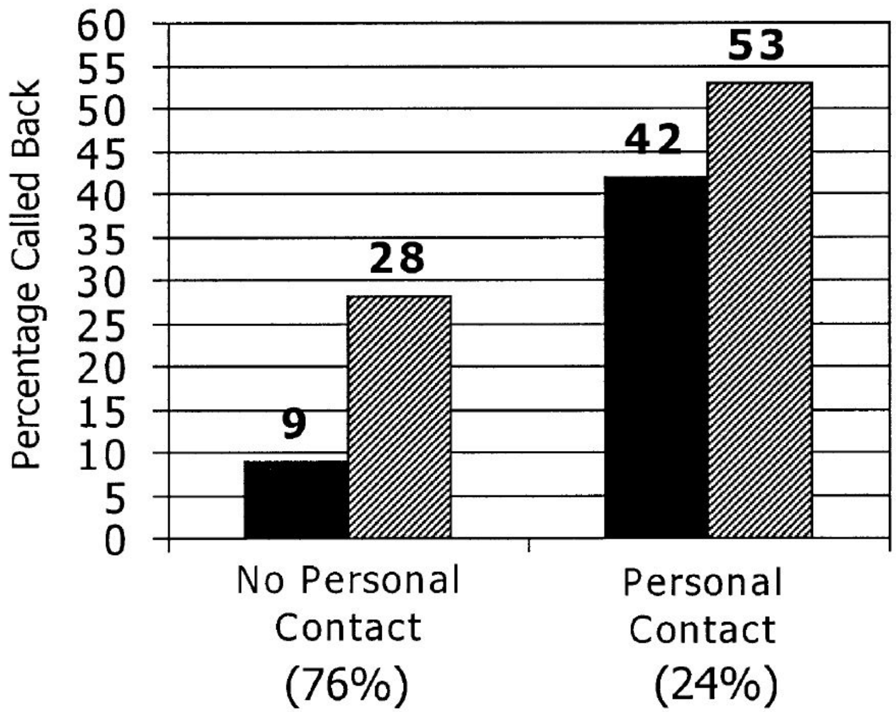


FIG. A2.—The effect of personal contact: black bars represent criminal record; striped bars represent no criminal record.

Discrimination

Why Has Black-White Skill Convergence Stopped? (Neal, 2006)

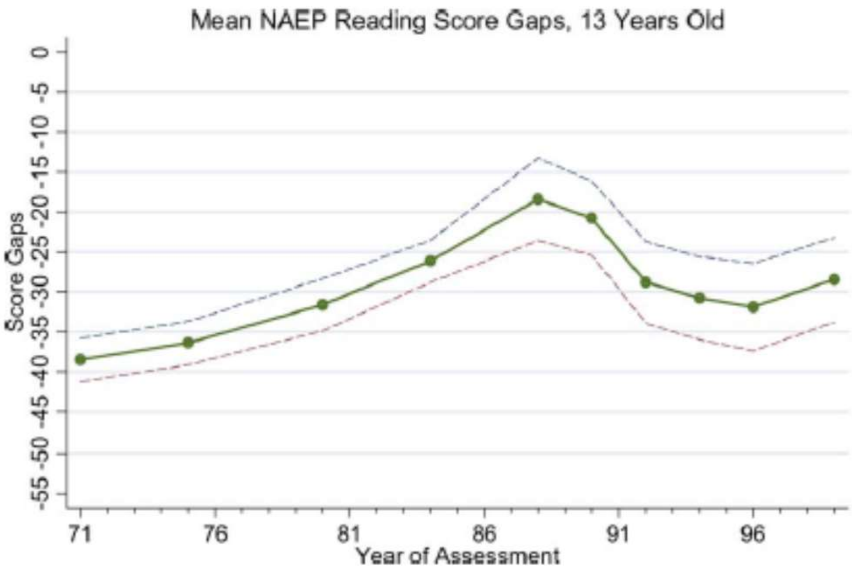
1. Gap in reading and math scores narrowed to 20-25 p.p. in the late 80s

2. Gap has stagnated or even widened slightly from the 80s to the late 90s

3. Discrimination models suggest that returns to skills for whites should be higher than blacks

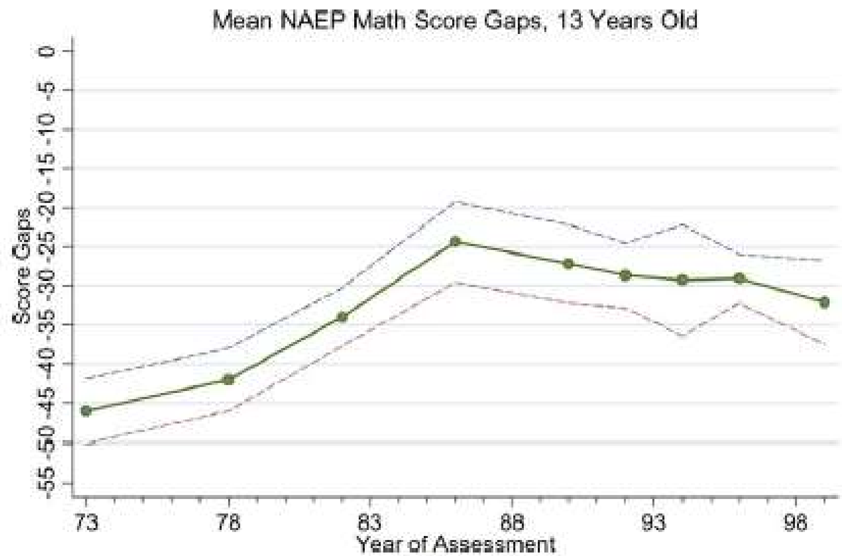
4. However, skill-returns are higher for blacks than whites

Neal (2006)



Mean NAEP Reading Score Gaps, 13 Years Old
Score Gaps
0
-5
-10
-15
-20
-25
-30
-35
-40
-45
-50
-55
71
76
81
86
91
96
Year of Assessment

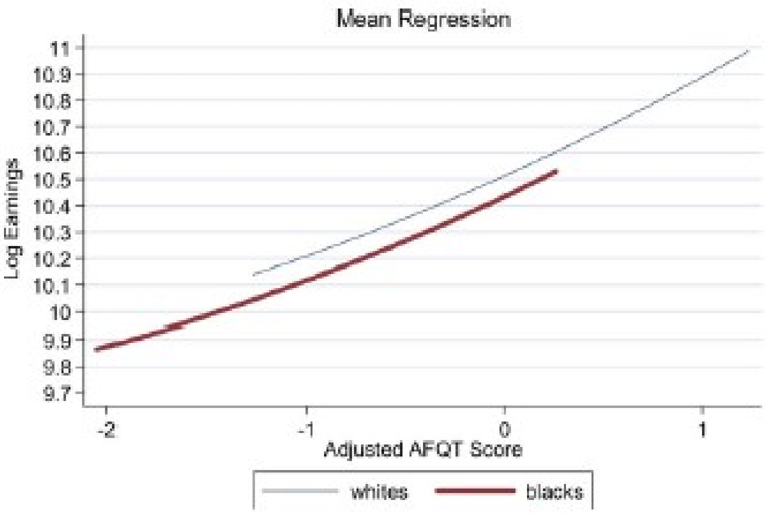
(a)



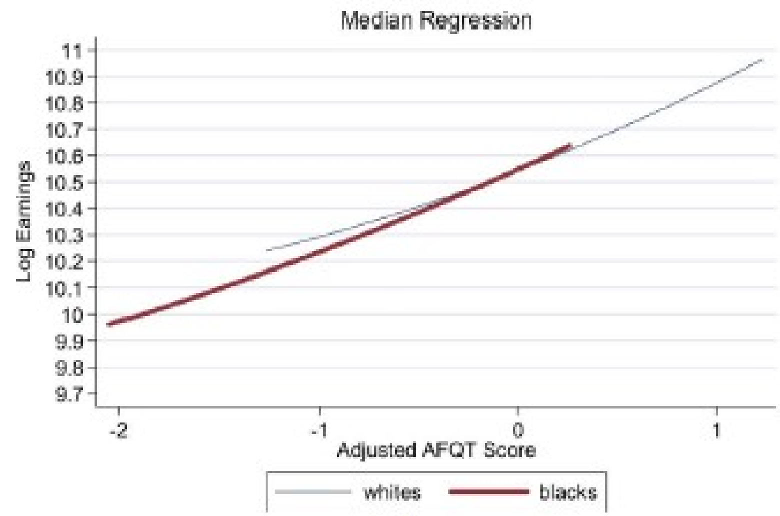
Mean NAEP Math Score Gaps, 13 Years Old
Score Gaps
0
-5
-10
-15
-20
-25
-30
-35
-40
-45
-50
-55
73
78
83
88
93
98
Year of Assessment

(b)

Neal (2006)


Mean Regression
11
10.9
10.8
10.7
10.6
10.5
10.4
10.3
10.2
10.1
10
9.9
9.8
9.7
Log Earnings
-2
-1
0
1
Adjusted AFQT Score
whites
blacks

(a)


Median Regression
11
10.9
10.8
10.7
10.6
10.5
10.4
10.3
10.2
10.1
10
9.9
9.8
9.7
Log Earnings
-2
-1
0
1
Adjusted AFQT Score
whites
blacks

(b)

Discrimination

Racial Inequality in the 21st Century (Fryer, 2011)

1. Declining significance of discrimination in labor markets
2. After accounting for skills (AFQT), the male black-white wage gap falls by 72% for the 1979 NLSY cohort and 39% for the 1997 NLSY cohort
3. Similar declines in the importance of skills on racial wage gaps for females

Fryer (2011)

Table 1 The importance of educational achievement on racial differences in labor market outcomes (NLSY79).

	Wage				Unemployment			
	Men		Women		Men		Women	
Black	−0.394	−0.109	−0.131	0.127	2.312	1.332	3.779	2.901
	(0.043)	(0.046)	(0.043)	(0.046)	(0.642)	(0.384)	(1.160)	(1.042)
Hispanic	−0.148	0.039	−0.060	0.161	2.170	1.529	2.759	2.181
	(0.049)	(0.047)	(0.051)	(0.051)	(0.691)	(0.485)	(0.973)	(0.871)
Age	0.027	0.012	−0.011	0.016	1.191	1.202	0.956	0.941
	(0.023)	(0.022)	(0.024)	(0.022)	(0.175)	(0.178)	(0.131)	(0.133)
AFQT		0.270		0.288		0.561		0.735
		(0.021)		(0.023)		(0.082)		(0.123)
$AFQT^2$		0.039		−0.009		1.005		1.276
		(0.019)		(0.020)		(0.151)		(0.161)
Obs.	1167	1167	1044	1044	1315	1315	1229	1229
R^2	0.068	0.206	0.009	0.135	0.022	0.050	0.040	0.058
% Reduction		72		197		75		32

Fryer (2011)

Table 2 The importance of educational achievement on racial differences in labor market outcomes (NLSY97).

	Wage				**Unemployment**			
	Men		**Women**		**Men**		**Women**	
Black	−0.179	−0.109	−0.153	−0.044	2.848	2.085	2.596	1.759
	(0.023)	(0.024)	(0.020)	(0.021)	(0.377)	(0.298)	(0.380)	(0.278)
Hispanic	−0.065	−0.014	−0.057	0.035	1.250	0.994	1.507	1.065
	(0.023)	(0.024)	(0.023)	(0.023)	(0.205)	(0.170)	(0.267)	(0.202)
Mixed race	0.007	0.009	−0.090	−0.057	3.268	3.216	1.317	1.278
	(0.143)	(0.145)	(0.072)	(0.065)	(1.661)	(1.618)	(0.975)	(0.911)
Age	0.064	0.062	0.039	0.039	0.934	0.937	1.084	1.081
	(0.006)	(0.006)	(0.006)	(0.006)	(0.038)	(0.038)	(0.048)	(0.048)
AFQT		0.089		0.148		0.664		0.595
		(0.011)		(0.012)		(0.049)		(0.052)
$AFQT^2$		−0.022		−0.035		1.248		1.140
		(0.012)		(0.012)		(0.095)		(0.107)
Obs.	3278	3278	3204	3204	3294	3294	3053	3053
R^2	0.047	0.065	0.029	0.081	0.032	0.051	0.026	0.049
% Reduction		39		71		41		52

Fryer (2011)

Table 4 The importance of educational achievement on racial differences in labor market outcomes (C&B 76).

	Men		Women	
Black	−0.273	−0.152	0.186	0.286
	(0.042)	(0.047)	(0.035)	(0.031)
Hispanic	−0.038	−0.007	0.005	0.059
	(0.081)	(0.077)	(0.094)	(0.088)
Other race	0.153	0.147	0.271	0.270
	(0.066)	(0.062)	(0.048)	(0.049)
SAT		0.003		0.001
		(0.001)		(0.001)
SAT^2		−0.000		−0.000
		(0.000)		(0.000)
Obs.	11,088	11,088	8976	8976
R^2	0.007	0.015	0.004	0.012
% Reduction		44		53

The dependent variable is the log of annual income. Annual income is reported as a series of ranges; each individual is assigned the midpoint of their reported income range as their annual income. Income data were collected for either 1994 or 1995. Individuals who report earning less than $1000 annually or who were students at the time of data collection are excluded from these regressions. Those individuals with missing SAT scores are included in the sample and a dummy variable is included in the regressions that include SAT variables to indicate that a person did not have a valid AFQT score. All regressions use institution weights and standard errors are clustered at the institution level. Standard errors are in parentheses.

Discrimination

Open Research Questions (IMO)

1. Discrimination may manifest itself via costs to actions, not via wages (for example).

Common to see unobserved cost differences explains educational choice differences.

Discrimination

2. Stigma vs. statistical discrimination (Loury).

My interpretation:

Standard statistical discrimination models ask for circumstances under which beliefs of whites about blacks, given data, are confirmed by data.

A distinct question is whether data are rich enough to swamp priors. In such a world; segregation inhibits.