Structural bounds on group differences in treatment effects

John Gardner*

June 2023

Abstract

I analyze the use of group variation in the probability of receiving a treatment to recover information about the causal effects of that treatment, within the framework of a structural model of counterfactual outcomes and enrollment into the treatment. Specifically, I study the conditions under which a simple group difference in differences—that is, subtracting the difference in mean outcomes between treated and untreated units belonging to a group with a low treatment rate from that same differences for units belonging to a group with a high rate—identifies a lower bound on the difference in average treatment effects between the high- and low-rate groups. When theory of prior empirical evidence imply that the effect of the treatment is nonnegative, this group difference in differences also identifies a lower bound on the average treatment effect itself for the high-rate group. While the conditions required for the lower-bound argument are not directly verifiable, I suggest falsification tests for whether they are consistent with the data. I also prevent several examples, illustrating the applicability of the identification results and the effectiveness of the falsification tests.

Keywords: Treatment effects, causal inference, differences in differences, Roy model, partial identification, bounds.

JEL Codes: C50, C34, J71, R23.

1 Introduction

Quasiexperimental analyses of treatment effects are typically predicated on the assumption that, after conditioning on a set of observable covariates, treatment status is as good as random—that is,

*Department of Economics, University of Mississippi, University, MS. jrgardne@olemiss.edu.

independent of counterfactual outcomes (Rubin, 1974). However, there is often reason to suspect that unobserved factors affect both the decision to enroll in the treatment and the counterfactual outcomes that units would experience with and without that treatment, invalidating this conditional independence assumption. In such cases, empiricists often turn to alternative identification strategies such as instrumental variables or difference-in-difference analyses. Although these strategies can circumvent the biases that unobserved confounders may introduce into comparisons between treated and untreated units, they also place rather stringent requirements—such as the availability of an instrument that affects treatment status but not counterfactual outcomes, or the presence of a comparison group whose untreated outcomes evolve over time in parallel with those of the treatment group—on the data available to the researcher.

In this paper, I study the extent to which variation between groups in the probability of receiving a treatment can be used to identify information about the causal effect of that treatment. Specifically, I investigate the conditions under which the difference in mean outcomes between treated and untreated units belonging to a group that is relatively unlikely to receive a treatment (the low-treatment-rate group) can be used to control for the selection bias that may contaminate that same mean outcome difference for a group that is relatively likely to receive the treatment (the high-treatment-rate group).

The motivation behind this comparison is that individuals may be more likely to enroll in a treatment if they expect to benefit from that treatment, which raises the specter that comparisons between treated and untreated units are contaminated with (positive) selection bias. If enrollment into the treatment is less selective for one group (in the sense that members of that group are relatively likely to receive the treatment), then we might suspect that comparisons of mean outcomes between treated and untreated members of that group contain less selection bias than do those comparisons for the low-rate group. If this is the case, then the group difference in differences—that is, the difference between the high- and low-rate groups in mean outcome differences between treated and untreated units—will *more than control* for the selection bias present in treated-untreated comparison for the high-rate group, bounding the group difference in average treatment effects from below. The substantive identification results in this paper provide sufficient conditions for this intuition to hold true. Although the group difference in average treatment effects may be informative in itself, if theory or prior empirical evidence suggest that the effect of the treatment is, at worst, zero, a lower bound on the group difference in treatment effects is also a bound on the average treatment effect itself for members of the high-rate group.

The basic approach to identification in this paper is similar to that in Gardner (2020), which this paper complements and extends. Analyzing Black-white differences in the wage effects of northward migration during the African American Great Migration period, Gardner (2020) considers a reduced-form setting in which the treatment enrollment decision depends on an unobserved variable, and mean counterfactual outcomes are are a linear in that variable. In that setting, the group difference in treated-untreated outcome differences identifies a lower bound on the group difference in average treatment effects under restrictions on the group-specific probabilities of receiving the treatment and the shape of the distribution governing the unobserved variable that determines treatment status. By explicitly modeling the treatment enrollment decision and counterfactual outcomes, this paper obtains an analogous conclusion under relatively mild restrictions on counterfactual outcomes, the net benefit of enrolling, and notably, the distributions of the unobservables that influence counterfactual outcomes and treatment enrollment. The cost of this additional generality is that the conditions under which the lower-bound identification argument holds, which involve nonlinear functions of unobserved variables, are not directly verifiable. However, for each of these conditions, I propose falsification tests that can help assess whether the data are consistent with the requirements of the identification procedure.

The underlying approach taken to identification in this paper—writing down a model of enrollment and outcomes and taking that model seriously to identify or bound components of the model, or treatment effects implied by those components—is related to the literature on identification and estimation of Roy models (Roy, 1951; Heckman and Honoré, 1990; Heckman and Vytlacil, 1999; D'Haultfoeuille and Maurel, 2013; Maurel and D'Haultfoeuille, 2013; Eisenhauer et al., 2015). This paper is also related the literature on identification of bounds on treatment effects (Manski, 1989, 1990), although the approach taken here is comparatively parametric. de Chaisemartin and D'Haultfoeuille (2017) also study the identification of average causal effects in the absence of a group that never receives the treatment. My results, which merely bound group differences in treatment effects, are less ambitious than theirs, though they can be applied when the data do not have a time dimension.

In Section 2, I use a simplified model to illustrate the logic of the lower-bound approach to identification in a stylized setting. In Section 3, I develop a flexible extension of that simple model that can be applied in more general settings, with relatively weak restrictions on the treatment enrollment decision, counterfactual outcomes, and heterogeneity between members of different groups, and derive corresponding identification results. In Section 4, I present empirical applications to illustrate the usefulness of the identification results and the effectiveness of the falsification tests. I conclude in Section 5.

2 Motivation: Identification in a simplified setting

To motivate the lower-bound identification result developed below, and to provide some intuition for how that argument works in a more stylized setting, I first show how it can be obtained in a simplified model with limited heterogeneity between groups. To that end, suppose that there are two groups, $g \in \{l, h\}$, where members of the high-rate group h are more likely to receive the treatment than members of the low-rate group l. Also suppose that members of each group differ in their level of a binary unobserved characteristic $a \in \{\underline{a}, \bar{a}\}$, the distribution of which is independent of group membership and given by $\pi = P(a = \bar{a})$. Let $d \in \{0, 1\}$ be an indicator for treatment status, and suppose that the counterfactual outcomes that members of group g would experience given treatment status d satisfy $y_{dg}(\underline{a}) < y_{dg}(\bar{a})$. Finally, assume that the net benefit of receiving the treatment is given by $\tilde{\gamma}_g(a) = \Delta_g + \gamma 1(a = \bar{a}) - \epsilon$, where, reflecting the idea of positive selection into treatment, $\gamma > 0$, and ϵ, which represents idiosyncratic influences on the treatment decision, is drawn independently of group membership with log concave distribution function F and associated density f.[1] In this simplified model, group differences in treatment rates are entirely determined by differences in the secular cost/benefit shifter Δ_g, so that group h's higher treatment rate implies that $\Delta_h > \Delta_l$.

The difference in mean outcomes between treated and untreated members of group g can decomposed into terms reflecting the average effect of the treatment on the treated (ATT) and selection bias that arises because enrollment and counterfactual outcomes are increasing in the unobserved

[1] A function is log concave if its logarithm is concave.

characteristic a according to

$$
\begin{aligned}
E(y|1,g) - E(y|0,g) &= [P(\bar{a}|1,g)y_{1g}(\bar{a}) + P(\underline{a}|1,g)]y_{1g}(\underline{a}) - [P(\bar{a}|0,g)y_{0g}(\bar{a}) + P(\underline{a}|0,g)]y_{0g}(\underline{a}) \\
&= \underbrace{P(\bar{a}|1,g)[y_{1g}(\bar{a}) - y_{0g}(\bar{a})] + P(\underline{a}|1,g)[y_{1g}(\underline{a}) - y_{0g}(\underline{a})]}_{\text{ATT}_g} \\
&\quad + \underbrace{[P(\bar{a}|1,g) - P(\bar{a}|0,g)]\, y_{0g}(\bar{a}) + [P(\underline{a}|1,g) - P(\underline{a}|0,g)]y_{0g}(\underline{a})}_{\text{Selection bias}_g},
\end{aligned}
$$

where "1" is shorthand for the event $d = 1$ (and similarly for "0") and $P(a|d,g)$ denotes the probability of having skill type $a \in \{\underline{a}, \bar{a}\}$ conditional on treatment status d and group membership g. The difference between the high- and low-rate groups in mean outcome comparisons between treated and untreated units is $[E(y|1,h) - E(y|0,h)] - [E(y|1,l) - E(y|0,l)]$. This group difference in differences will bound the group difference in average treatment effects from below if the selection bias term is larger for the low-treatment-rate group, so that

$$
\begin{aligned}
&[P(\bar{a}|1,h) - P(\bar{a}|0,h)]\, y_{0h}(\bar{a}) + [P(\underline{a}|1,h) - P(\underline{a}|0,h)]y_{0h}(\underline{a}) \\
&\qquad \leq [P(\bar{a}|1,l) - P(\bar{a}|0,l)]\, y_{0l}(\bar{a}) + [P(\underline{a}|1,l) - P(\underline{a}|0,l)]y_{0l}(\underline{a}), \quad (1)
\end{aligned}
$$

or, since $P(\bar{a}|d,g) = 1 - P(\underline{a}|d,g)$, if

$$
[P(\bar{a}|1,h) - P(\bar{a}|0,h)][y_{0h}(\bar{a}) - y_{0h}(\underline{a})] - [P(\bar{a}|1,l) - P(\bar{a}|0,l)][y_{0l}(\bar{a}) - y_{0l}(\underline{a})] \leq 0.
$$

Since members of group h are, by definition, more likely to select into the treatment, a natural assumption is that the return to the unobserved characteristic a with respect to untreated outcomes for members of that group is smaller than for members of group l, i.e., that $y_{0h}(\bar{a}) - y_{0h}(\underline{a}) \leq y_{0l}(\bar{a}) - y_{0l}(\underline{a})$. Given the positive nature of selection into the treatment, under this ancillary assumption, the differential selection bias components also satisfy

$$
\begin{aligned}
&[P(\bar{a}|1,h) - P(\bar{a}|0,h)][y_{0h}(\bar{a}) - y_{0h}(\underline{a})] - [P(\bar{a}|1,l) - P(\bar{a}|0,l)][y_{0l}(\bar{a}) - y_{0l}(\underline{a})] \\
&\qquad \leq \{[P(\bar{a}|1,h) - P(\bar{a}|0,h)] - [P(\bar{a}|1,l) - P(\bar{a}|0,l)]\}[y_{0l}(\bar{a}) - y_{0l}(\underline{a})] \\
&\qquad\qquad \approx \frac{\partial}{\partial \Delta_g}[P(\bar{a}|1,g) - P(\bar{a}|0,g)][y_{0l}(\bar{a}) - y_{0l}(\underline{a})]d\Delta, \quad (2)
\end{aligned}
$$

where $d\Delta = \Delta_h - \Delta_l$.[2] Moreover, since $y_{0g}(\bar{a}) > y_{0g}(\underline{a})$ for $g \in \{l, h\}$ by assumption, (2), and hence (1), will hold if

$$\frac{\partial}{\partial \Delta_g}[P(\bar{a}|1,g) - P(\bar{a}|0,g)] \leq 0. \tag{3}$$

The following result establishes sufficient conditions for (3).

Proposition 1. *Suppose that members of group g are treated with probability less than 1/2. Then*

$$\frac{\partial}{\partial \Delta_g}[P(\bar{a}|1,g) - P(\bar{a}|0,g)] \leq 0$$

if

$$f(\Delta_g + \gamma) - f(\Delta_g) \leq 0. \tag{4}$$

The intuition behind this result, which I prove in Appendix A, is illustrated by Figure 1. A change in Δ_g will increase the probability of enrollment among those with high a by about $f(\Delta_g + \gamma)d\Delta$, and among those with low a by about $f(\Delta_g)d\Delta$. When condition (4) holds, an increase in Δ_g will induce more low- than high-types to enroll in the treatment, so that selection bias will be decreasing in Δ_g, and hence with respect to the probability of receiving the treatment.

Proposition 1 also imbues the model with implications that are at least partially observable. Since $y_{0g}(\bar{a}) > y_{0g}(\underline{a})$ by assumption, if the conditions of the proposition are satisfied, then the relatively small differences $dP(1|\bar{a})$ between high-a members of the high- and low-rate groups will be associated with relatively large untreated outcomes $y_{0g}(\bar{a})$ for either group, while the relatively large group differences $dP(1|\underline{a})$ for low-a members will be associated with relatively small untreated outcomes $y_{0g}(\underline{a})$. In other words, the differences $dP(1|a)$ in treatment probabilities will be negatively

[2] This result relies on the fact that $P(\bar{a}|1,g) > P(\bar{a}|0,g)$, (i.e., that there is positive selection into the treatment), which is intuitive. To show this formally, note that $\partial P(\bar{a}|1,g)/\partial\gamma = (\partial/\partial\gamma)\{\pi F(\Delta+\gamma)/[\pi F(\Delta+\gamma) + (1-\pi)F(\Delta)]\}$ has the same sign as $(1-\pi)F(\Delta) > 0$ (and the opposite sign of $\partial P(\underline{a}|1,g)/\partial\gamma$).

correlated with untreated outcomes $y_{0g}(a)$ across values of a. Although it is not possible to assess the sign of the covariance between untreated outcomes and group differences in enrollment conditional on an unobserved characteristic, it may be possible to assess whether untreated outcomes are negatively correlated with treatment probabilities conditional on some observable proxy for that characteristic. This suggests an empirical test for whether selection bias, as measured by the correlation between enrollment probabilities and counterfactual outcomes across unobserved characteristics, is decreasing in the probability of receiving the treatment. I pursue this idea further in the context of a richer model of the treatment decision and counterfactual outcomes in Section 3 and put it into practice in Section 4.

3 Model and identification

3.1 The model

The simple model developed in Section 2 is intended to capture the notions that the net benefit of enrolling in the treatment is greater for members of the high-treatment-rate group, leading to group differences in enrollment rates, while both counterfactual outcomes and the probability of enrolling are increasing in a common unobserved variable, leading to positive selection bias for both groups. In this section, I develop an extended version of this model in order to capture these aspects of the decision environment in a more general setting, while keeping the model analytically tractable.

Let d be an indicator for whether an individual enrolls in the treatment. Suppose that there are two groups of units indexed by $g \in \{l, h\}$, where members of the low-treatment-rate group l are less likely to enroll than members of the high-treatment-rate group h. Also suppose that individuals differ in their realizations of an unobserved variable a that influences both their decision to enroll in the treatment and their counterfactual outcomes.

Suppose that the decision to enroll takes the form

$$d_g(a, \epsilon) = 1\left[\tilde{\gamma}(\Delta_g, a) - \epsilon \geq 0\right] \tag{5}$$

for $g \in \{l, h\}$, where Δ_g is a vector of group-specific secular cost or benefit shifters and ϵ represents the influence of idiosyncratic factors on the enrollment decision. In order to model the enrollment decision flexibly, I also assume the following.

Assumption 1. $\tilde{\gamma}(\Delta_g, a)$ *takes the form*

$$\tilde{\gamma}(\Delta_g, a) = \Delta_{1g} + \Delta_{2g}\gamma(\Delta_{3g} + a),$$

where $\gamma' \geq 0$, $\gamma'' \leq 0$ *and* γ *exhibits constant relative risk aversion.*

Besides using a standard CRRA functional form, Assumption 1 allows the effect of changes in $\Delta_g = (\Delta_{1g}, \Delta_{2g}, \Delta_{3g})$ to be constant, increasing, or decreasing in a. Note that when Δ_g is scalar, this assumption nests a set of simpler models in which $\tilde{\gamma}(\Delta_g, a)$ takes one of the forms $\{\Delta_g + \gamma(a), \Delta_g\gamma(a), \gamma(\Delta_g + a)\}$. This model allows for the possibility that the net benefit $\tilde{\gamma}$ from enrolling in the treatment is simply the difference between treated and untreated outcomes, perhaps minus a term reflecting this cost of enrollment, as well the possibility that γ reflects costs and benefits other than those associated with treatment-induced differences in outcomes.

In order to allow for the possibility that group differences in the distribution of a contribute to group differences in enrollment and outcomes, I also assume the following.

Assumption 2. *The densities* π_g, $g \in \{l, h\}$, *of* a *belong to the same scale family on* $\mathbb{R}^+$, *where* π_l *is the standard density.*

This assumption allows for heterogeneity in the group-specific distributions of a, while requiring those distributions to have the same basic shape. Note that Assumption 2 implies that $\pi_h(a) = \pi_l(a/\sigma_h)$, where σ_h is the scale parameter for group h.

Also assume that the distribution of ϵ satisfies the following conditions.

Assumption 3. *The idiosyncratic component* ϵ *of the enrollment decision rule is distributed over* $\mathbb{R}$ *independently of* a *and* g *with log concave distribution function* F *(and associated density function* f*).*

The log concavity assumption on F is typical in discrete-choice models. For example, the error distributions used in the logit and probit models satisfy this assumption. The assumption that the distribution of ϵ is identical across groups is a matter of notational convenience. It can be weakened to allow the group-specific distributions of ϵ to belong to the same location-scale family by absorbing the group-specific location and scale parameters into Δ_{1g} and Δ_{2g}.

To capture the idea that treated units are positively selected on the unobserved variable a that influences enrollment, suppose that counterfactual outcomes y_{dg}, $d \in \{0, 1\}$, $g \in \{l, h\}$, that members of the high- and low-rate groups would experience with and without the treatment are determined by

$$y_{dg} = y_{dg}(a), \quad y'_{dg}(a) \geq 0. \tag{6}$$

Note that observed outcomes can be expressed as $y_g = (1 - d)y_{0g} + dy_{1g}$. The assumption that the y_{dg} depend on a alone can be weakened, for example by letting $y_{dg} = \tilde{y}_{dg}(a) + \nu_g$, where the distribution of ν_g does not depend on treatment status.[3]

Finally, the identification argument developed below requires that counterfactual outcomes satisfy the following slope condition.

Assumption 4. *Untreated outcomes satisfy* $y'_{0h}(a)\sigma_h \leq y'_{0l}(a)$ *for all* a.

Assumption 4 essentially holds that members of the high-treatment-rate group fare worse in the untreated condition than do members of the low-rate group. This assumption (which places no restrictions on treated outcomes) is consistent with group differences in treatment rates, and may be justified in some applications by prior evidence or *a priori* theoretical considerations. For example, in the case of the Great Migration, it requires that Blacks receive lower wages than whites when living in the South. Furthermore, in Proposition 2 below, I develop a test for whether observed outcome data are consistent with this assumption on average, to within a first-order approximation.

The strength of this test depends on the nature of the available data and the shape of the π_g (which, in general, is not identified without further restrictions). Specifically, when the data have a time dimension and contain information from a pre-treatment period during which no units have been treated, it is possible to directly test whether Assumption 4 holds on average, to within a first-order approximation (under the assumption that the primitives of the model such as π_g and y_{0g} are stationary, of course). When only post-treatment data are available, this test becomes a falsification test, the implications of which depend on whether the distributions of a are log concave or log convex. In all three cases, a finding that $Var(y|h) \leq Var(y|l)$ suggests that the data are consistent with the assumption.

[3] The validity of Proposition 2 in this case requires an additional restriction on the group specific variances ν_g, for example that $Var(\nu_g)$ is the same for both groups. This assumption is consistent with the underlying idea that it is a that generates group differences in outcomes and enrollment.

Proposition 2. *Suppose that equations (5) and (6) and Assumption 2 hold. Then, by a first-order approximation:*

1. *If pre-treatment-period data are available,* $Var(y|h)/Var(y|l) = E[y'_{0h}(\sigma_h a)\sigma_h|l]/E[y'_{0l}(a)|l]$*,where* y *denotes observed outcomes in the pretreatment period.*

2. *If only the post-treatment data are available, then if the* π_g *are log concave,* $Var(y|0,h)/Var(y|0,l) > 1$ *implies that* $E[y'_{0h}(\sigma_h a)\sigma_h|l] > E[y'_{0l}(a)|l]$*, where* y *denotes observed outcomes in the post-treatment period. Otherwise, if the* π_g *are log convex,* $Var(y|0,h)/Var(y|0,l) < 1$ *implies that* $E[y'_{0h}(\sigma_h a)\sigma_h|l] < E[y'_{0l}(a)|l]$.

3.2 Identification

The logic behind the approach to identification taken in this paper is that, if selection bias is decreasing in the probability of enrollment, the group difference in differences will over-control for selection bias for the high-rate group, bounding the group difference in average treatment effects (and possibly the average treatment effect itself for the high-rate group) from below. The substantive identification results presented below elucidate the conditions under which this logic holds true. In this section, I sketch the identification results, presenting full proofs in Appendix A.

The difference in mean observed outcome differences between the high- and low-rate groups is

$$[E(y|1,h) - E(y|0,h)] - [E(y|1,l) - E(y|0,h)], \tag{7}$$

where "1" is shorthand for the event $d = 1$, and similarly for "0." When counterfactual outcomes are determined as in (6), the difference in mean outcomes between treated and untreated members of group g can be decomposed into the average effect of the treatment on the treated (ATT) and a selection bias term as

$$\begin{aligned} E(y|1,g) - E(y|0,g) &= \int y_{1g}(a)p(a|1,g)da - \int y_{0g}(a)p(a|0,g)da \\ &= \underbrace{\int [y_{1g}(a) - y_{0g}(a)]\, p(a|1,g)da}_{\text{ATT}_g} + \underbrace{\int [p(a|1,g) - p(a|0,g)]\, y_{0g}(a)da}_{\text{Bias}_g}, \end{aligned} \tag{8}$$

where $p(a|g,d)$ denotes the density of a conditional on group membership g and treatment status

d.

The difference (7) between the high- and low-treatment-rate groups will bound the group differences in ATTs from below if the differential bias term $\text{Bias}_h - \text{Bias}_l$ is nonpositive, or if

$$\int [p(a|1,h) - p(a|0,h)]\, y_{0h}(a)da \leq \int [p(a|1,l) - p(a|0,l)]\, y_{0l}(a)da. \tag{9}$$

Under Assumption 4 on the relative slopes of untreated outcomes, a sufficient condition for (9) is that (see Lemma 1 of Appendix A)

$$\frac{\partial}{\partial\theta'}\left\{\int [p(a|1,g;\theta) - p(a|0,g;\theta) y_{0g}(a)da\right\} d\theta \leq 0, \tag{10}$$

between θ_l and θ_h, where $\theta = (\Delta_1, \Delta_2, \Delta_3, \sigma)$ and $d\theta = \theta_h - \theta_l$.

Under the enrollment decision rule (5) and Assumption 2 on the distributions of the a, the group-specific probabilities of enrollment can be expressed in terms of the low-rate group's distribution of a as $P(1|g) = E\{F[\tilde{\gamma}(\Delta_g, \sigma_g a)]|l\}$, so that group differences in treatment probabilities arise from differences in the distribution of a (represented by $d\sigma_g$) and different preferences for enrollment (represented by $d\Delta_g$) according to

$$P(1|h) - P(1|l) \approx \frac{\partial P(1|g)}{\partial\theta'} d\theta \equiv dP(1) > 0.$$

A change $d\theta$ in these parameters will change the likelihood of enrollment for an individual with unobservable a by $dP(1|a) \approx f[\tilde{\gamma}(\Delta, \sigma a)][\partial\tilde{\gamma}(\Delta, \sigma a)/\partial\theta']d\theta$.

Proposition 3 below shows that, under the assumptions of the model, a sufficient condition for (10), and hence for the group difference in differences to bound the group difference in ATTs from below, is that there is a group for whom the probability of enrollment is less than one half and the changes $dP(1|a)$ in the enrollment probabilities induced by the distributional and preference differences $d\theta$ between the high- and low-rate groups are negatively correlated with untreated outcomes.

Proposition 3. *Suppose that (5), (6), and Assumptions 1-4 hold. Then a sufficient condition for*

(9) is that there is a $g \in \{l, h\}$ such that $P(1|g) < 1/2$ and

$$Cov\left[f\left(\tilde{\gamma}(\Delta, \sigma a)\right)\frac{\partial\tilde{\gamma}(\Delta, \sigma a)}{\partial\theta'}d\theta, y_{0g}(\sigma_g a)\right] \leq 0 \tag{11}$$

between θ_l and θ_h.[4]

Although the proof of Proposition 3, which proceeds via a series of lemmas presented in Appendix A, is almost entirely technical, the intuition is fairly straightforward. The covariance condition (11) holds that the typical individual induced to enroll in the treatment by their membership in the high-rate group would experience a relatively low outcome absent the treatment. Selection bias, in other words, is decreasing in the probability of enrollment.

The covariance condition (11) in Proposition 3 is obviously the operative requirement for the validity of the lower-bound identification argument. Unfortunately, because it places restrictions on potentially nonlinear functions of an unobserved variable, it is neither directly verifiable nor likely to be acceptable on *a priori* grounds. A natural falsification test for whether the data are consistent with this condition is to examine the covariance

$$Cov[P(1|h, x) - P(1|l, x), y_{0g}] \tag{12}$$

for $g \in \{l, h\}$, where x is a vector of observed covariates thought to be related to both the treatment decision and counterfactual outcomes, either because they are correlated with a or because a itself is a (random) function of those covariates.[5] Note that if outcomes are only observed after enrollment decisions have been made, (12) can only be estimated with respect to the distribution of observed y among untreated units.[6] The motivation behind this approach is that, in general, a finding that untreated outcomes are negatively correlated with enrollment probabilities conditional on observables lends credibility to the hypothesis (though of course does not imply) that the unobservable

[4] In (11), expectations are taken with respect to the distribution of a among the low-rate group.

[5] An alternative version of this test replaces y_{0g} in (12) with $E(y_{0g}|x)$. The justification for this is that, by the law of total covariance, $Cov[P(1|a) - P(0, a), y_{0g}] = Cov[P(1|h, x) - P(1|l, x), E(y_{0g}|x)] + E\{Cov[P(1|a) - P(0|a), a|x]\}$. This version of the test may be easier to interpret when x is multidimensional.

[6] This problem will be less severe for the low-treatment-rate group, for whom the untreated population and overall populations are more similar. Also note that since the change in the treatment probability is likely to be small for those with large realizations of a (who are most likely to enroll regardless of group membership), the covariance among the untreated (which will exclude those with the smallest treatment probability changes but the largest untreated outcomes) is likely to overstate the covariance with respect to the full distribution of untreated outcomes.

covariance condition (11) holds (this interpretation is similar in spirit to that of Altonji, Elder, and Taber, 2005).

The identification argument developed in this paper largely abstracts away from observable covariates by bounding group differences in selection bias, rather than attempt to eliminate it entirely. One approach to incorporating covariates is to model a as some random function $a = g(x, u)$ of the covariates, in which case the identification argument applies as presented across covariate strata (this is analogous to the approach taken by popular linear sample-selection estimators). Alternatively, the model can be applied within strata defined by discrete (or binned continuous) covariates, although in that case the requirements for identification should be assessed within each stratum. I illustrate multiple approaches to incorporating observable covariates in Section 4.

4 Empirical examples

4.1 The Great Migration

Gardner (2020) uses identification results for a reduced-form model to interpret Black-white differences in North-South wage differentials during the African American Great Migration period as lower bounds on the Black-white difference in causal effects of migration from the Southern to Northern US. Although this period is named for the tremendous rates at which Southern-born blacks moved to the North, accounts of the Great Migration sometimes overlook the fact that this Southern Black diaspora was accompanied by considerable migration among whites as well (see Tolnay, 2003, for an overview of the sociodemographics of the Great Migration). In this section, I revisit that analysis to demonstrate how a structural model of migration and outcomes during this period can be combined with the results developed above to use Black-white variation in migration rates to identify bounds on the Black-white differences in the effects of northward migration from Black-white differences in regional wage differentials.

In the context of the Great Migration, the idea behind the model is that the unobserved (or incompletely observable) skills that determine wages in the South and the North are probably also related to the Southern-born individuals' migration decisions, either because the benefit to migrating is increasing in those skills, the cost of migrating is decreasing in those skills, or both. Consequently, wage differentials between migrants and non-migrants (which are widely known to be positive, both overall and by race) may be contaminated with positive selection bias. However, the greater rates at which Southern-born Blacks migrated relative to their white counterparts may suggest that migration was less selective, and that wage comparisons by migrant status contain less selection bias, for Blacks.

The data for this example are drawn from 1% Integrated Public Use Microdata Samples (IPUMS) of the 1970-1970 US Decennial Censuses (Ruggles et al., 2010). From these samples, I retain information on Southern-born Black and white men aged 16-64 and living in the US at the time of enumeration.[7] I define wages as annual income from wages and salary, and inflate all wage information to 1999 dollars using the CPI deflator variable included in the IPUMS extracts.

Recall that, insofar as the migration decision and counterfactual earnings in the North and South can be viewed through the lens of the model developed in Section 3, there are three prerequisites for the lower-bound identification procedure to apply in this setting. One of these conditions is that counterfactual outcomes satisfy the relative slope condition of Assumption 4, which in this case asserts that the wage return to the unobserved variable a that influences wages and migration status is lower in the South for Blacks than for whites. While this condition is not directly observable, the test developed in Proposition 2 can be used to assess whether it holds on average. Although in this setting wages are only observed after the migration decision is made, the second part of the proposition implies that observed outcomes are consistent with this assumption if the Black-white ratio of wage variances among those living in the South is less than one. Table 1 summarizes tests of this condition. Decade-by-decade, the standard deviation of wages for Blacks living in the South is less than that for whites living there. In every case, a normality-based test of the null that the standard deviations are the same against the alternative that the standard deviation is smaller for whites fails to reject. In addition, 95% percentile-t bootstrap confidence intervals for the ratio lie strictly to the left of one in every decade. For log wages, the story is similar for 1940 and 1950, with tests failing to reject the null of equal standard deviations. For 1960 and 1970, the standard deviations are numerically similar and the confidence intervals are tightly concentrated around one, although the fact that formal tests reject the null might warrant taking the lower-bound interpretation for log wages in these decades with a grain of salt.

[7] I define the South as Alabama, Arkansas, Delaware, Florida, Georgia, Kentucky, Louisiana, Maryland, Mississippi, North Carolina, Oklahoma, South Carolina, Tennessee, Texas, Virginia and West Virginia are Southern states, and the North as any other state in the US.

The other requirements are given by Proposition 3, which provides sufficient conditions for the group difference in differences to identify a lower bound on the group difference in ATTs. The proposition requires that there is at least one group for which the probability of migrating is less than one half, and that the covariance condition (11) is satisfied (i.e., untreated outcomes are negatively correlated with group differences in migration probabilities across values of a) for that group. As the top panel of Table 2 shows, the former condition holds for both groups. Across decades, the white migration rate is between 13 and 20%, with Black-white differences in those probabilities between 6 and 14%.

The covariance condition is the most difficult to assess, since it depends on the covariance between nonlinear functions of an unobserved variable. Since wages and migration are thought to depend on an unobserved skill index in this case, to provide evidence on the plausibility of the covariance condition, I examine the decade-specific empirical correlations between wages and racial differences in migration probabilities by educational attainment (measured in years), which is likely correlated with overall skills. Specifically, for each decade, I compute race-specific probability of migrating for each level of educational attainment, then report the race-specific correlation between Black-white differences in the probability of migrating conditional on education and wages for those remaining in the South. The results of this exercise are summarized in the left panel of Table 3. Across decades and races, and regardless of whether wages are measured in levels of in logs, the estimated correlations are all positive. Each of the decade- and race-specific null hypotheses of zero correlation is rejected by both a parametric test based on a joint normality assumption and by a test based on a 95% nonparametric bootstrap confidence interval.

These positive correlations make it difficult to justify applying the identification argument in this setting. However, as Figure 2 shows, racial differences in selective migration during this period were not so simple, at least with respect to education. The figure, which plots Black-white differences in education-specific migration probabilities against education, reveals an interesting and robust pattern. In each decade, Black-white differences in migration probabilities are increasing in education for lower levels of education, and decreasing in education for higher levels. This suggests that, for relatively educated Southerners, racial differences in selective migration are consistent with the intuition used to motivate the lower-bound approach—less skilled Blacks were more willing to migrate than their white counterparts—while for less-educated Southerners, selection into migration followed the opposite pattern.

This finding implies that the requirements of the identification argument may be more likely to hold among relatively educated Southerners. Accordingly, I refine the sample to include only observations with above race-and-decade-specific median educational attainment. Before revisiting the covariance condition, I verify that the other requirements for identification hold within this subsample. The lower panel of Table 1 shows that Black-white relative wages in the South, as measured by their relative variances, are consistent with this condition according to both parametric and bootstrap tests, as are relative log wages in every decade but 1960 (and even in that decade, the point estimate is not that far from one). The lower panel of Table 2 shows that group-specific migration rates in this sample are also consistent with the identification argument, with between 13 and 20% of Southern-born whites migrating, and an additional 6 to 14% of Blacks doing so.

The corresponding covariance tests are presented in the right panel of Table 3. Here, the covariance between wages in the South and Black-white differences in education-specific migration rates is positive (and statistically significant) for Blacks in 1940 and 1950, negative and statistically significant in 1970, and close to zero (and statistically insignificant) in 1970. For whites, these covariances are negative and statistically insignificant in every decade. Note that, because it is only a sufficient condition, the lower-bound identification argument applies if the covariance condition is satisfied for either group.

Among relatively educated Southerners, therefore, the data appear to be consistent with the requirements for identification. Accordingly, Table 4 presents estimates of decade-specific group difference-in-difference regressions of wages and log wages on indicators for being Black, migrating north, and an interaction between these indicators, for Southern-born Black and white men with educational attainment above their respective decade-and-race-specific medians. The estimates show that, across all decades, Southern-born white men who migrated north earned more than those who remained in the South (by \$3,700-7,900, depending on the decade), and Black men who migrated earned more still in most decades (an additional \$800-\$1,700, although the estimate for 1940 is a negative and statistically significant \$1,200).[8] In log terms, white migrants earned between 17 and 24% more than non-migrants, while the premium for Black migrants was between 23 and 39%

[8]The finding of a negative "premium" for Blacks in 1940 is consistent with the notion that Black migration was less selective with respect to wages.

larger.

Since, among relatively educated Southerns, the evidence in Tables 1, 2, and 3 is consistent with the requirements for identification, the group difference-in-differences results can be interpreted as lower-bounds on Black-white differences in the causal effects of migration on wages. The wage results from Table 4 imply that northward migration increased Blacks' wages by at least \$800-\$1,740, depending on the decade (the negative estimate for 1940 is consistent with a negative, zero, or positive racial difference in that decade), and the log wage results imply that migration increased Blacks' wages by 23-39% more than they did whites' wages. As I note in the introduction, when it is thought that the effect of the treatment is, at worst, zero, a lower bound on the group difference in ATTs is also a lower bound on the ATT itself for the high-rate group. Based on the sizes of the raw wage differentials, combined with what we know about the North relative to the South during this period, there is a strong case to be made that this logic applies to the causal effect of migration on wages during the Great Migration. We might therefore also conclude that northward migration increased Blacks' wages themselves by at least \$800-\$1,740 or 33-39%, with the understanding that these bounds are likely very conservative.

4.2 The National Supported Work Demonstration

To further illustrate the applicability and effectiveness of this paper's results, I use them in a setting for which there is a reliable experimental benchmark against which structural bounds can be compared: the well-known National Supported Work Demonstration (NSW) program (see Lalonde, 1986, for a nice review of the program and its administration). In particular, I use the Dehejia and Wahba (1999; 2002) data on experimental treatment and control units from the original program as well as their observational data combining treated units from the NSW with control units drawn from the Panel Study of Income Dynamics to study group heterogeneity in the causal effects of the program.[9]

I analyze differences in the effect of the program on earnings in 1978 (i.e., after the administration of the treatment to treated units) between Black and white participants. The top panel of Table 5 presents race-specific estimates of the effect of the program on earnings, obtained from the

[9]Rajeev Dehejia has generously made these data available at `http://users.nber.org/~rdehejia/data/.nswdata3.html`.

experimental sample. The estimated treatment effects are positive for both groups, and statistically significant for Blacks. The Black-white difference in estimated treatment effects is about \$1,200, although this difference is not statistically significant.

Before drawing comparisons between treated and untreated units in the observational sample, I first estimate propensity scores predicted from a logit regression of treatment status on age, education in years, earnings in the pre-treatment years 1974 and 1975, indicators for having less than a high-school diploma, being Black, being Hispanic, and being married. I then apply the Crump et al. (2009) rule of thumb, retaining only observations for which the estimated propensity score is between .1 and .9. The simple justification for this in the current context is that the model is more likely to be an accurate descriptor of the enrollment decision and counterfactual outcomes for individuals with similar likelihoods of participating in the program, at least according to their observed characteristics. After making this sample selection, the high-rate group consists of Blacks, among whom 49% received the treatment, compared to 40% among whites. Within this selected sample, I estimate a group difference-in-differences regression relating post-period earnings to both treatment and race. As the bottom panel of Table 5 shows, the estimated earnings difference between treatment and control units in this observational sample is statistically insignificant for both groups, as is the Black-white difference of about -\$1,700. in these differences.

In the case of the Great Migration, large positive regional wage differentials for each group raise obvious concerns about positive selection into migration. The observational results for the NSW program might lead to concerns about selection bias of two different natures (both consistent with the experimental estimates after accounting for sampling error). Since the racial difference in treated-untreated comparisons is negative, one concern might be that whites are positively selected into the program, while Blacks are negatively selected. If this were the case, the Black-white difference in treated-untreated mean outcome comparisons would necessarily represent a lower bound on the Black-white difference in average treatment effects. Another possibility, however, is that both Blacks and whites are positively selected, but enrollment is more selective for whites, leading to a negative estimated group difference in differences. I apply the identification results in this application to address this second possibility.

To assess the applicability of the lower-bound identification result in this setting, I begin by testing whether the relative variances of earnings are consistent with the slope condition required by Assumption 4. Note that, in the case of the NSW, earnings are observed in 1974 and 1975, prior to the administration of the program. Thus, by the first part of Proposition 2, this slope condition can be tested directly (at least to within a first-order approximation). As Table 6 shows, the sample standard deviations of earnings for Blacks exceed those for whites in both pre-treatment years, although for 1974, neither parametric nor bootstrap tests reject the null hypothesis of equality. The variance ratio for 1974 (which is consistent with the slope condition) is probably more representative of the long-run relative variation in earnings absent program participation, since program participation and the variance of earnings are probably highly correlated with employment shocks in 1975, the last year before the administration of the program to treated units.

To assess the covariance condition, I re-estimate the propensity score within the selected sample. Using the estimated parameters from this logit, I then predict the difference in the (counterfactual) conditional probabilities with which each individual would enroll in the program if they were Black or if they were white, holding the rest of their covariates constant at their observed values. I also estimate race-specific linear regressions of earnings in both 1974 and 1975 on the same covariates (except for an indicator for being Black) used to estimate the propensity score, which I use to obtain estimates of expected earnings conditional on those covariates. Finally, I compute the correlation between these differential conditional probabilities and expected conditional earnings for both races and both pre-treatment years.[10] As the results in Table 6 show, these correlations are negative and statistically significant for whites in both 1974 and 1975. For Blacks, they are positive and statistically significant when expected 1974 earnings are used, and not statistically different from zero when 1975 earnings are used.

The data are therefore consistent with the required treatment rate, slope, and covariance conditions, suggesting that the lower-bound identification result applies. Under that model of Section 3, the group difference in differences estimate of -$1,700 reported in Table 5 can be interpreted as a lower bound on the Black-white difference in the average effect of the program on earnings. The lower-bound procedure does not produce an "interesting" result here, in the sense that neither a negative nor a statistically insignificant lower bound reveal much about the group difference in ATTs. However, since in the experimental sample we cannot reject the null of no differential effect,

[10] Although I use expected wages conditional on covariates to illustrate that variation of the covariance test, I obtain similar results when I use observed wages instead.

this is the appropriate outcome of the procedure—a researcher applying the lower-bound result in this setting would conclude, correctly, that the observational results are consistent with zero and positive Black-white differences in the average effects of the program on earnings.

5 Conclusion

Concerns about selection bias in observational studies are almost always motivated, if only implicitly, by a structural model in which the same factors that determine counterfactual outcomes also influence the decision to enroll in the treatment. In this paper, I use a flexible specification of such a model to study the conditions under which group differences in enrollment rates can be combined with group differences in mean-outcome comparisons to recover information about the causal effect of the treatment. I present sufficient conditions for the group difference in differences—that is, the difference between the high- and low-treatment-rate groups in mean outcome comparisons between treated and untreated units—to identify a lower bound on the difference in the average effect of the treatment between the high- and low-rate groups. It is not difficult to conceive of applications in which group differences in average treatment effects are interesting, informative and relevant for policy. In some cases, moreover, the most reasonable null hypotheses is that the effect of the treatment is, at worst, zero. In such cases, a lower bound on the group difference in average treatment effects is also a conservative lower bound on the average treatment effect itself for the high-treatment-rate group.

The theoretical results developed in this paper permit identification of a lower bound under three key conditions. The first of these places restrictions on group differences in the returns to the unobservable determinants of counterfactual outcomes. While this condition is not directly observable, I present a simple test for whether it holds on average, to within a first-order approximation. The second condition, which holds that there is at least one group for whom the probability of receiving the treatment is less than one half, is directly observable. The third condition requires that group differences in the probability of receiving the treatment conditional on the unobserved determinants of outcomes are negatively correlated with untreated outcomes. Although this condition cannot be directly verified, and is unlikely to be acceptable *a priori*, I suggest testing the plausibility of this condition by examining whether it holds with respect to group differences in treatment probabilities conditional observed correlates of outcomes and enrollment. When these conditions are satisfied, the identification results imply that a simple group difference in differences recovers a lower bound on the group difference in average treatment effects. Although the conditions required for identification are exacting and not directly testable, they may permit some inference about the causal effect of the treatment when there is concern about selection bias in cases where methods predicated on selection on observables, or those that require instrumental variables or panel data, cannot be applied.

I apply the identification results in two empirical examples. In the first, I interpret Black-white differences in differences in wages between migrants and non-migrants from the South to the North during the African American Great Migration as lower bounds on racial differences in the effect of northward migration on wages. Although there is reason to believe that not all of the requirements for identification are met in this case, I present evidence that the data are consistent with those conditions among relatively educated Southerners. I also apply the results to study racial differences in the effect of participation in the National Supported Work Demonstration program, comparing observational results to experimental results obtained using data on applications whose treatment status was determined by randomization. In this application, the data are consistent with the requirements of the identification argument, and the resulting lower bound is also consistent with the experimental difference.

Appendix A: Proofs

Proof of Proposition 1. Dropping the g subscripts for simplicity and applying Bayes' rule,

$$P(\bar{a}|1) = \frac{\pi F(\Delta+\gamma)}{\pi F(\Delta+\gamma) + (1-\pi)F(\Delta)}$$

and

$$P(\bar{a}|0) = \frac{\pi\left[1 - F(\Delta+\gamma)\right]}{\pi\left[1 - F(\Delta+\gamma)\right] + (1-\pi)\left[1 - F(\Delta)\right]}.$$

Differentiating,

$$\frac{\partial P(\bar{a}|1)}{\partial \Delta} = \frac{\pi(1-\pi)}{P(1)^2}\left[F(\Delta)f(\Delta+\gamma) - F(\Delta+\gamma)f(\Delta)\right] \equiv \frac{\pi(1-\pi)}{P(1)^2}c_1$$

and

$$\frac{\partial P(\bar{a}|0)}{\partial \Delta} = \frac{\pi(1-\pi)}{P(0)^2}\left\{[1-F(\Delta+\gamma)]\, f(\Delta) - [1-F(\Delta) f(\Delta+\gamma)]\right\} \equiv \frac{\pi(1-\pi)}{P(0)^2} c_0,$$

where $P(1) = 1 - P(0) = \pi F(\Delta+\gamma) + (1-\pi)F(\Delta)$.

Notice that $\partial P(\bar{a}|1)/\partial \Delta$ (and hence c_1) have the same sign as

$$\frac{f(\Delta+\gamma)}{F(\Delta+\gamma)} - \frac{f(\Delta)}{F(\Delta)} < 0,$$

where the inequality is a consequence of the log concavity of $F(\epsilon)$, which implies that $(f/F)' < 0$.

Since $P(0) < P(1)$ and $c_1 < 0$,

$$\begin{aligned}
\frac{\partial P(\bar{a}|1)}{\partial \Delta} - \frac{\partial P(\bar{a}|0)}{\partial \Delta} &= \pi(1-\pi)\left[\frac{c_1}{P(1)^2} - \frac{c_0}{P(0)^2}\right] \\
&< \frac{\pi(1-\pi)}{P(0)^2}(c_1 - c_0) \\
&= \frac{\pi(1-\pi)}{P(0)^2}\left[f(\Delta+\gamma) - f(\Delta)\right]. \qquad \square
\end{aligned}$$

Proof of Proposition 2. To prove the first part, normalize $\sigma_l = 1$ and note that, by first-order approximations about μ_l,

$$Var[y_{0g}(a)|g] = Var[y_{0g}(\sigma_g a)|l] \approx [y'_{0g}(\sigma_g \mu_l)]^2 \sigma_g^2 \approx \{E[y'_{0g}(\sigma_g a)|l]\sigma_g\}^2.$$

To prove the second part, use similar approximations to write

$$\begin{aligned}
Var\left[y_{0g}(a)|g, \tilde{\gamma}(\Delta, a) < \epsilon\right] &= Var\left[y_{0g}(\sigma_g a)|l, \tilde{\gamma}(\Delta, \sigma_g a) < \epsilon\right] \\
&\approx [y'_{0g}(\mu_g)]^2 \sigma_g^2 Var\left(a|l, \tilde{\gamma}(\Delta, \sigma_g a) < \epsilon\right).
\end{aligned}$$

By the law of total variance,

$$Var(a|l, \tilde{\gamma}(\Delta, \sigma a) < \epsilon) = E\left[Var\left(a|l, \epsilon, a < \frac{\tilde{\gamma}^{-1}(\Delta, \epsilon)}{\sigma}\right)\right] + Var\left[E\left(a|l, \epsilon, a < \frac{\tilde{\gamma}^{-1}(\Delta, \epsilon)}{\sigma}\right)\right], \tag{13}$$

where $\tilde{\gamma}^{-1}$ is defined for fixed Δ. By a first-order expansion about $\mu_\epsilon = E(\epsilon)$, the change in the first term in (13) is approximately

$$\frac{\partial}{\partial(\tilde{\gamma}^{-1}/\sigma)} Var\left(a|l, a < \frac{\tilde{\gamma}^{-1}(\Delta, \mu_\epsilon)}{\sigma}\right)\left(\frac{\partial\tilde{\gamma}^{-1}(\Delta, \mu_\epsilon)/\sigma}{\partial\Delta} d\Delta - \frac{\tilde{\gamma}^{-1}(\Delta, \mu_\epsilon)}{\sigma^2} d\sigma\right). \tag{14}$$

By Proposition 1 of Heckman and Honoré (1990), π log concave implies that the leading term in (14) is positive and π log convex implies that it is negative. Furthermore, a higher treatment rate implies that (using another first-order approximation)

$$\begin{aligned} dP(0) &= dP\left(a < \frac{\tilde{\gamma}^{-1}(\Delta, \epsilon)}{\sigma}\right) \\ &= \int \pi\left(\frac{\tilde{\gamma}^{-1}(\Delta, \epsilon)}{\sigma}\right)\left[\left(\frac{\partial\tilde{\gamma}^{-1}(\Delta, \epsilon)/\sigma}{\partial\Delta}\right) d\Delta - \frac{\tilde{\gamma}^{-1}(\Delta, \epsilon)}{\sigma^2} d\sigma\right] f(\epsilon) d\epsilon \\ &\approx \pi\left(\frac{\tilde{\gamma}^{-1}(\Delta, \mu_\epsilon)}{\sigma}\right)\left[\left(\frac{\partial\tilde{\gamma}^{-1}(\Delta, \mu_\epsilon)/\sigma}{\partial\Delta}\right) d\Delta - \frac{\tilde{\gamma}^{-1}(\Delta, \mu_\epsilon)}{\sigma^2} d\sigma\right] < 0. \end{aligned}$$

Thus the change in the first term in (13) is approximately negative if π is log concave and positive if π is log convex.

The second term in (13) can be expressed

$$Var\left[E\left(a|l, \epsilon, a < \frac{\tilde{\gamma}^{-1}(\Delta, \epsilon)}{\sigma}\right)\right] = E\left[(\mu_{0\epsilon} - \mu_0)^2\right]$$

where $\mu_{0\epsilon} = E(a|\epsilon, a < \tilde{\gamma}^{-1}/\sigma)$ and $\mu_0 = E_\epsilon(\mu_{0\epsilon}) = E_{a\epsilon}(a|a < \tilde{\gamma}^{-1}/\sigma)$. As long differentiation and integration can be interchanged, the derivative of this term with respect to a vector of parameters θ can be expressed

$$2E\left[(\mu_{0\epsilon} - \mu_0)\frac{\partial}{\partial\theta}(\mu_{0\epsilon} - \mu_0)\right] = 2\left[E\left(\mu_{0\epsilon}\frac{\partial\mu_{0\epsilon}}{\partial\theta}\right) - \mu_0\frac{\partial\mu_0}{\partial\theta}\right].$$

Since, by a first-order approximation about μ_ϵ,

$$E\left(\mu_{0\epsilon}\frac{\partial\mu_{0\epsilon}}{\partial\theta}\right) \approx \mu_0\frac{\partial\mu_0}{\partial\theta},$$

the change in the second term in (13) is approximately zero.

Thus, since

$$\frac{Var(y|0,h)}{Var(y|0,l)} \approx \frac{[y_{0h}'(\mu_h)]^2\sigma_h^2 Var\left(a|l,\tilde{\gamma}(\Delta_h,\sigma_h a)<\epsilon\right)}{[y_{0l}'(\mu_l)]^2 Var\left(a|l,\tilde{\gamma}(\Delta_l,a)<\epsilon\right)} \approx \left(\frac{E[y_{0h}'(\sigma_h a)|l]\sigma_h}{E[y_{0l}'(a)|l]}\right)^2 \frac{Var\left(a|l,\tilde{\gamma}(\Delta_h,\sigma_h a)<\epsilon\right)}{Var\left(a|l,\tilde{\gamma}(\Delta_l,a)<\epsilon\right)},$$

the result follows. □

The proof of Proposition 3 makes use of the following lemmas.

Lemma 1. *Suppose that $y_{0h}'(\sigma_h a)\sigma_h \le y_{0l}'(a)$ and define $p(a|1;\theta) = p\left(a|l,\epsilon<\tilde{\gamma}(\Delta,\sigma a)\right)$. Then*

$$\int \left[p(a|h,1)-p(a|h,0)\right] y_{0h}(\sigma_h a)da \le \int \left[p(a|l,1)-p(a|l,0)\right] y_{0l}(a)da$$

if there is a $g \in \{l,h\}$ such that

$$\frac{\partial}{\partial\theta'}\left\{\int \left[p(a|1;\theta)-p(a|0;\theta)\right] y_{0g}(\sigma_g a)da\right\} d\theta \le 0$$

between θ_l and θ_h.

Proof. If

$$\int \left[p(a|1;\theta)-p(a|0;\theta)\right] y_{0h}(\sigma_h a)da - \int \left[p(a|1;\theta)-p(a|0;\theta)\right] y_{0l}(a)da$$
$$= \int \left[p(a|1;\theta)-p(a|0;\theta)\right]\left[y_{0h}(\sigma_h a)-y_{0l}(a)\right] da \le 0$$

then the conclusion follows.[11] Note that

$$\begin{aligned} p(a|1;\theta)-p(a|0;\theta) &= \pi(a)\left[\frac{F\left(\tilde{\gamma}(\Delta,\sigma a)\right)}{P(1|\theta)} - \frac{1-F\left(\tilde{\gamma}(\Delta,\sigma a)\right)}{P(0|\theta)}\right] \\ &= \pi(a)\left[\frac{F\left(\tilde{\gamma}(\Delta,\sigma a)\right)}{P(0|\theta)P(1|\theta)} - \frac{1}{P(0|\theta)}\right], \end{aligned}$$

is negative when $a < a^*$ where a^* satisfies $F\left(\tilde{\gamma}(\Delta,\sigma a^*)\right) = P(1|\theta)$ and positive otherwise.

[11] To be explicit, if $\rho_g = p(a|1;\theta_g) - p(a|0;\theta_g)$ then we have $\int \rho_h y_{0h} - \int \rho_l y_{0l} = \int(\rho_h-\rho_l)y_l + \int \rho_h(y_h - y_l)$, which since, as the proof shows, the second term on the right is nonpositive, implies that $\int \rho_h y_{0h} - \int \rho_l y_{0l} \le \int(\rho_h-\rho_l)y_{0l} = (\partial/\partial\theta')\left(\int \rho_l y_{0l}\right)|_{\theta=\theta^*}d\theta$ or $\int \rho_h y_{0h} - \int \rho_l y_{0l} \le \int(\rho_h-\rho_l)y_{0h} = -(\partial/\partial\theta')\left(\int \rho_h y_{0h}\right)|_{\theta=\theta^*}(-d\theta)$ where θ^* lies on the segment between θ_l and θ_h. The lemma will hold approximately if the differential is nonpositive at either of the θ_g. Note that the derivative in the statement of the lemma is taken with respect to $\theta = (\Delta,\sigma)$, with σ_g in $y_{0g}(\sigma_g a)$ fixed.

Thus since $y'_{0h}(\sigma_h a)\sigma_h - y'_{0l}(a) < 0$,

$$\begin{aligned}\int [p(a|1;\theta) - p(a|0;\theta)]\,[y_{0h}(\sigma_h a) - y_{0l}(a)]\,da &< \int [p(a|1;\theta) - p(a|0;\theta)]\,[y_{0h}(\sigma a^*) - y_{0l}(a^*)]\,da \\ &= [y_{0h}(\sigma_h a^*) - y_{0l}(a^*)] \int [p(a|1;\theta) - p(a|0;\theta)]\,da \\ &= 0. \qquad \square\end{aligned}$$

Lemma 2. *Suppose that* $P(1|g) < 1/2$, $y'_0 > 0$, *and that there exists a* $d \in \{0,1\}$ *and an* a^* *such that* $[\partial p(a|d;\theta)/\partial\theta']d\theta$ *is positive on* $a < a^*$ *and negative on* $a > a^*$. *Then*

$$\frac{\partial}{\partial\theta'}\left\{\int [p(a|1;\theta) - p(a|0;\theta)]\,y_{0g}(\sigma_g a)da\right\}d\theta \leq 0$$

if

$$Cov\left[f\left(\tilde{\gamma}(\theta,a)\right)\frac{\partial\tilde{\gamma}(\theta,a)}{\partial\theta'}d\theta, y_{0g}(\sigma_g a)\right] \leq 0.$$

Proof. As long as differentiation and integration can be interchanged, the derivative in question will be nonpositive if the derivative of the integrand is nonpositive.[12] Then, suppressing the dependence of $p(a|d)$ on θ, we can write

$$\frac{\partial p(a|1)}{\partial\theta'}d\theta = \frac{\partial}{\partial\theta'}\frac{\pi(a)F\left(\tilde{\gamma}(\theta,a)\right)}{\int \pi(a)F\left(\tilde{\gamma}(\theta,a)\right)da}d\theta = \frac{c_1(a)}{P(1)^2}$$

and

$$\frac{\partial p(a|0)}{\partial\theta'}d\theta = \frac{\partial}{\partial\theta'}\frac{\pi(a)\left[1 - F\left(\tilde{\gamma}(\theta,a)\right)\right]}{\int \pi(a)\left[1 - F(\tilde{\gamma}(\theta,a)\right]da}d\theta = \frac{c_0(a)}{P(0)^2}$$

[12] This requires that $|(\partial/\partial\theta')[p(a|1;\theta) - p(a|0;\theta)]d\theta| \leq g(a)$ where g is some integrable function. As is shown in the following,

$$\int\left|\frac{\partial}{\partial\theta'}[p(a|1;\theta) - p(a|0;\theta)]d\theta\right|da \leq \left|E\left[f\left(\tilde{\gamma}(\theta,a)\right)\frac{\partial\tilde{\gamma}(\theta,a)}{\partial\theta'}d\theta y_0(\sigma_g a)\right]\right| + \left|E\left[y_0(\sigma_g a)\right]E\left[f\left(\tilde{\gamma}(\theta,a)\right)\frac{\partial\tilde{\gamma}(\theta,a)}{\partial\theta'}d\theta\right]\right|,$$

thus a sufficient condition is that the expectations on the right-hand side of the inequality are finite.

where

$$c_1(a) \equiv \pi(a) f\left(\tilde{\gamma}(\theta, a)\right) \frac{\partial \tilde{\gamma}(\theta, a)}{\partial \theta'} d\theta \int \pi(a) F(\tilde{\gamma}(\theta, a)) da \\ - \pi(a) F\left(\tilde{\gamma}(\theta, a)\right) \int \pi(a) f\left(\tilde{\gamma}(\theta, a)\right) \frac{\partial \tilde{\gamma}(\theta, a)}{\partial \theta'} d\theta da,$$

and

$$c_0(a) \equiv -\left[\int \pi(a) \left[1 - F\left(\tilde{\gamma}(\theta, a)\right)\right] da\right] \pi(a) f\left(\tilde{\gamma}(\theta, a)\right) \frac{\partial \tilde{\gamma}(\theta, a)}{\partial \theta'} d\theta \\ + \pi(a) \left[1 - F\left(\tilde{\gamma}(\theta, a)\right)\right] \int \pi(a) f\left(\tilde{\gamma}(\theta, a)\right) \frac{\partial \tilde{\gamma}(\theta, a)}{\partial \theta'} d\theta da.$$

Suppose that c_1 is positive on $a < a^*$ and negative on $a > a^*$. Since y_0 is increasing,

$$\int c_1(a) y_0(\sigma_g a) da \leq \int_0^{a^*} c_1(a) y_0(\sigma_g a^*) da + \int_{a^*}^{\infty} c_1(a) y_0(\sigma_g a^*) da = y_0(\sigma_g a^*) \int c_1(a) da = 0.$$

Similarly, if c_0 is positive on $a < a^*$ and negative otherwise,

$$\int c_0(a) y_0(\sigma_g a) da \leq 0.$$

Since $P(1) < P(0)$ and either $\int c_1 y_0 \leq 0$ or $\int c_0 y_0 \leq 0$, we have either

$$\frac{\partial}{\partial \theta'} \left\{ \int \left[p(a|1) - p(a|0)\right] y_0(\sigma_g a) da \right\} d\theta = \int \left[\frac{c_1(a)}{P(1)^2} - \frac{c_0(a)}{P(0)^2}\right] y_0(\sigma_g a) da \\ < \frac{1}{P(0)^2} \int \left[c_1(a) - c_0(a)\right] y_0(\sigma_g a) da$$

or

$$\frac{\partial}{\partial \theta'} \left\{ \int \left[p(a|1) - p(a|0)\right] y_0(\sigma_g a) da \right\} d\theta = \int \left[\frac{c_1(a)}{P(1)^2} - \frac{c_0(a)}{P(0)^2}\right] y_0(\sigma_g a) da \\ < \frac{1}{P(1)^2} \int \left[c_1(a) - c_0(a)\right] y_0(\sigma_g a) da.$$

To complete the proof, note that

$$\begin{aligned}\int [c_1(a) - c_0(a)]\, y_0(\sigma_g a) &= \int \left[\pi(a) f\left(\tilde{\gamma}(\theta,a)\right) \frac{\partial \tilde{\gamma}(\theta,a)}{\partial \theta'} d\theta da - \pi(a) \int \pi(a) f\left(\tilde{\gamma}(\theta,a)\right) \frac{\partial \tilde{\gamma}(\theta,a)}{\partial \theta'} d\theta da\right] y_0(\sigma_g a) da \\ &= E\left[f\left(\tilde{\gamma}(\theta,a)\right) \frac{\partial \tilde{\gamma}(\theta,a)}{\partial \theta'} d\theta y_0(\sigma_g a)\right] - E\left[y_0(\sigma_g a)\right] E\left[f\left(\tilde{\gamma}(\theta,a)\right) \frac{\partial \tilde{\gamma}(\theta,a)}{\partial \theta'} d\theta\right] \\ &= Cov\left[f\left(\tilde{\gamma}(\theta,a)\right) \frac{\partial \tilde{\gamma}(\theta,a)}{\partial \theta'} d\theta, y_0(\sigma_g a)\right]. \qquad \square\end{aligned}$$

Lemma 3. *Suppose that* $[\partial \tilde{\gamma}(\theta,a)/\partial\theta']d\theta$ *is (weakly) monotone increasing or (weakly) monotone decreasing in* a*,* $[\partial P(1|\theta)/\partial\theta']d\theta > 0$*, and* f *is log concave. Then there exists a* $d \in \{0,1\}$ *and an* a^* *such that* $[\partial p(a|d;\theta)/\partial\theta']d\theta$ *is positive on* $a < a^*$ *and negative on* $a > a^*$*.*

Proof. Suppose that $[\partial \tilde{\gamma}(\theta,a)/\partial\theta']d\theta$ is weakly decreasing. Note that, by assumption,

$$\frac{\partial P(1)}{\partial \theta'} d\theta = E\left[f\left(\tilde{\gamma}(\theta,a)\right) \frac{\partial \tilde{\gamma}(\theta,a)}{\partial \theta'} d\theta\right] > 0$$

and that $[\partial p(a|1)/\partial\theta']d\theta$ (which is proportional to the function c_1 defined in Lemma 2) has the same sign as

$$\frac{f\left(\tilde{\gamma}(\theta,a)\right)}{F\left(\tilde{\gamma}(\theta,a)\right)} \frac{\partial \tilde{\gamma}(\theta,a)}{\partial \theta'} d\theta - \frac{E\left[f\left(\tilde{\gamma}(\theta,a)\right) \frac{\partial \tilde{\gamma}(\theta,a)}{\partial \theta'} d\theta\right]}{E\left[F\left(\tilde{\gamma}(\theta,a)\right)\right]}. \tag{15}$$

Since f log concave implies that f/F is monotone decreasing, and since $[\partial \tilde{\gamma}(\theta,a)/\partial\theta']d\theta$ is decreasing by assumption, the first term in (15) is monotone decreasing whenever $[\partial \tilde{\gamma}(\theta,a)/\partial\theta']d\theta > 0$. The second term is a positive constant. Hence there is an a^* such that (15) is positive on $a < a^*$ and negative on $a > a^*$.

Now suppose that $[\partial \tilde{\gamma}(\theta,a)/\partial\theta']d\theta$ is weakly increasing and note that $[\partial p(a|0)/\partial\theta']d\theta$ has the same sign as

$$\frac{E\left[f\left(\tilde{\gamma}(\theta,a)\right) \frac{\partial \tilde{\gamma}(\theta,a)}{\partial \theta'} d\theta\right]}{E\left\{\left[1 - F\left(\tilde{\gamma}(\theta,a)\right)\right]\right\}} - \frac{f\left(\tilde{\gamma}(\theta,a)\right)}{1 - F\left(\tilde{\gamma}(\theta,a)\right)} \frac{\partial \tilde{\gamma}(\theta,a)}{\partial \theta'} d\theta. \tag{16}$$

The first term in (16) is a positive constant. Since f log concave implies that $f/(1-F)$ is monotone increasing, the second term in (16) is monotone increasing whenever $[\partial \tilde{\gamma}(\theta,a)/\partial\theta']d\theta > 0$. Therefore, there is an a^* such that (16) is positive on $a < a^*$ and negative on $a > a^*$.[13] $\square$

[13] Note that (16) positive for all a implies $[\partial p(a|0)/\partial\theta']d\theta$ is always positive, in contradiction to $\int c_0 = 0$.

Proof of Proposition 3. If $\tilde{\gamma}(\Delta, \sigma a) = \Delta + \gamma(\sigma a)$ then

$$\frac{\partial \tilde{\gamma}(\Delta, \sigma a)}{\partial \theta'} d\theta = d\Delta + \gamma'(\sigma a) a d\sigma.$$

The change in this expression with respect to a is $(\gamma'' \sigma a + \gamma')\, d\sigma \gtreqless 0$ as $-(\gamma'' \sigma a / \gamma') d\sigma \lesseqgtr d\sigma$. Since γ is CRRA, $(\partial \tilde{\gamma} / \partial \theta') d\theta$ is (weakly) monotone increasing or decreasing.

If $\tilde{\gamma}(\Delta, \sigma a) = \Delta \gamma(\sigma a)$ then

$$\frac{\partial \tilde{\gamma}(\Delta, \sigma a)}{\partial \theta'} d\theta = \gamma(\sigma a) d\Delta + \Delta \gamma'(\sigma a) a d\sigma.$$

The change in this expression is $\gamma' \cdot (\sigma d\Delta + \Delta d\sigma) + \Delta \gamma'' \sigma a d\sigma \gtreqless 0$ as $-\Delta d\sigma (\gamma'' \sigma a / \gamma') \lesseqgtr (\sigma d\Delta + \Delta d\sigma)$. Since γ is CRRA, $(\partial \tilde{\gamma} / \partial \theta') d\theta$ is (weakly) monotone increasing or decreasing.

In either case, the proposition follows from Lemmas 1, 2, and 3.

If $\tilde{\gamma}(\Delta, \sigma a) = \gamma(\Delta + \sigma a)$ then

$$\frac{\partial \tilde{\gamma}(\Delta, \sigma a)}{\partial \theta'} d\theta = \gamma'(\Delta + \sigma a)(d\Delta + a d\sigma).$$

There are two cases to consider. If $d\sigma < 0$ then we must have that $d\Delta > 0$ (otherwise $dP(1)$ would be negative). Since $\gamma' \geq 0$, $(\partial \tilde{\gamma} / \partial \theta') d\theta$, and hence (15) and $[\partial p(a|1)/\partial \theta'] d\theta$, are positive when $a < -d\Delta / d\sigma$ and negative otherwise.

If $d\sigma > 0$, note that as in the proof of Lemma 3, $[\partial p(a|0)/\partial \theta'] d\theta$ has the same sign as

$$\frac{E\left[f\left(\gamma(\Delta + \sigma a)\right) \frac{\partial \gamma(\Delta + \sigma a)}{\partial \theta'} d\theta\right]}{E\left\{\left[1 - F\left(\gamma(\Delta + \sigma a)\right)\right]\right\}} - \frac{f\left(\gamma(\Delta + \sigma a)\right)}{1 - F\left(\gamma(\Delta + \sigma a)\right)} \gamma'(\Delta + \sigma a)(d\Delta + a d\sigma). \tag{17}$$

The first term is a positive constant. Since f is log concave, as long as $\lim_{\epsilon \to \infty} f = 0$, $1 - F$ is log concave as well, and since log concavity is preserved by both linear and concave transformations, the function $1 - F[\gamma(\Delta + \sigma a)]$ is also log concave (Bagnoli and Bergstrom, 2005). Therefore,

$$\frac{-d\{1 - F[\gamma(\Delta + \sigma a)]\}/da}{1 - F[\gamma(\Delta + \sigma a)]} = \frac{f[\gamma(\Delta + \sigma a)]}{1 - F[\gamma(\Delta + \sigma a)]} \gamma'(\Delta + \sigma a) \sigma$$

is monotone increasing. Hence (17) will be monotone decreasing as soon as $a > -d\Delta / d\sigma$, so there must be an a^* such that (17) is positive on $a < a^*$ and negative on $a > a^*$. The conclusion then follows from Lemmas 1 and 2. This case does not use the assumption that γ is CRRA. $\square$

References

Altonji, J. G., T. E. Elder, and C. R. Taber (2005). Selection on Observed and Unobserved Variables: Assessing the Effectiveness of Catholic Schools. *Journal of Political Economy 113*(1), 151–184.

Bagnoli, M. and T. Bergstrom (2005). Log-Concave Probability and its Applications. *Economic Theory 26*(2), 445–469.

Crump, R. K., J. V. Hotz, G. W. Imbens, and O. A. Mitnik (2009). Dealing with Limited Overlap in Estimation of Average Treatment Effects. *Biometrika 96*(1), 187–199.

de Chaisemartin, C. and X. D'Haultfoeuille (2017). Fuzzy Differences in Differences. *Review of Economic Studies*, Forthcoming.

Dehejia, R. H. and S. Wahba (1999). Causal Effects in Nonexperimental Studies: Reevaluating the Evaluation of Training Programs. *Journal of the American Statistical Association 94*(448), 1053–1062.

Dehejia, R. H. and S. Wahba (2002). Propensity-Score Matching Methods for Nonexperimental Causal Studies. *Review of Economics and Statistics 84*, 151–161.

D'Haultfoeuille, X. and A. Maurel (2013). Another Look at the Identification at Infinity of Sample Selection Models. *Econometric Theory 29*(1), 213–224.

Eisenhauer, P., J. Heckman, and E. Vytlacil (2015). Generalized Roy Model and the Cost-Benefit Analysis of Social Programs. *Journal of Political Economy 123*(2), 413–443.

Gardner, J. (2020). Roy-Model Bounds on the Wage Effects of the Great Migration. *Econometrics Journal 39*(10), 1014–1041.

Heckman, J. J. and B. E. Honoré (1990). The Empirical Content of the Roy Model. *Econometrica 58*(5), 1121–1149.

Heckman, J. J. and E. J. Vytlacil (1999). Local Instrumental Variables and Latent Variable Models for Identifying and Bounding Treatment Effects. *Proceedings of the National Academy of Sciences of the United States of America 96*(8), 4730–4734.

Lalonde, R. J. (1986). Evaluating the Econometric Evaluations of Training Programs with Experimental Data. *The American Economic Review 76*(4), 604–620.

Manski, C. F. (1989). Anatomy of the Selection Problem. *The Journal of Human Resources 24*(3), 343.

Manski, C. F. (1990). Nonparametric Bounds on Treatment Effects. *The American Economic Review 80*(2), 319–323.

Maurel, A. and X. D'Haultfoeuille (2013). Inference on an Extended Roy Model, with an Application to Schooling Decisions in France. *Journal of Econometrics 174*(2), 95–106.

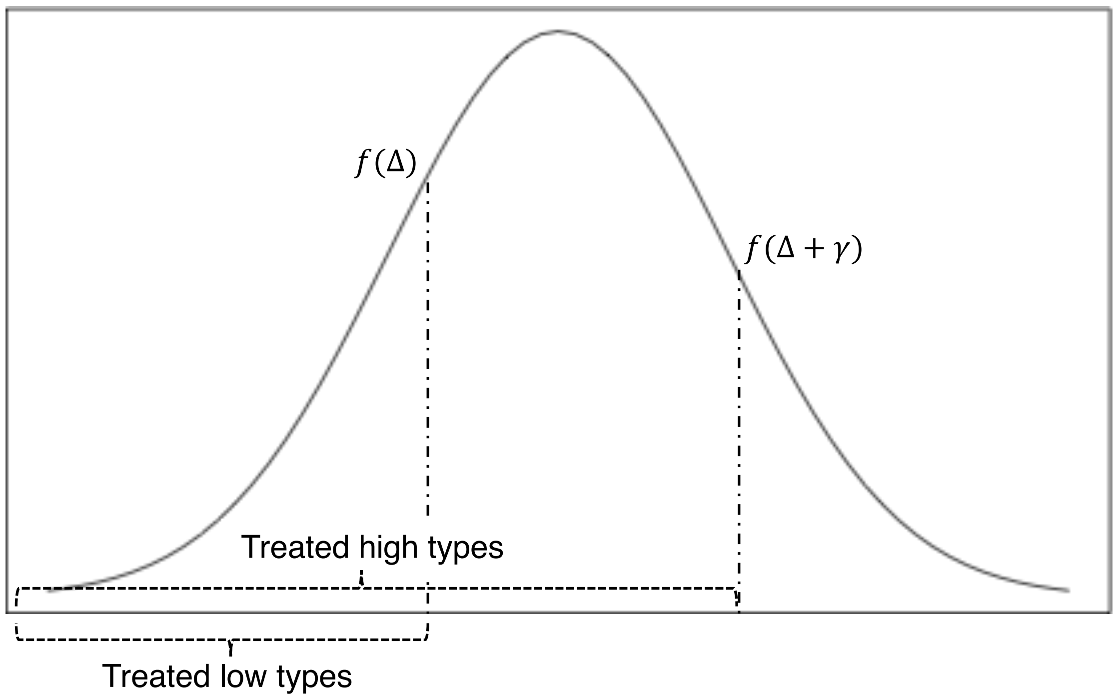
Roy, A. (1951). Some Thoughts on the Distribution of Earnings. *Oxford Economic Papers 3*(2), 135–146.

Rubin, D. B. (1974). Estimating Causal Effects of Treatments in Randomized and Nonrandomized Studies. *Journal of Educational Psychology 66*(5), 688–701.

Ruggles, S., J. T. Alexander, K. Genadek, R. Goeken, M. B. Schroeder, and M. Sobek (2010). Integrated Public Use Microdata Series: Version 5.0 [machine-readable database]. *Minneapolis: University of Minnesota*.

Tolnay, S. E. (2003). The African American "Great Migration" and Beyond. *Annual Review of Sociology 29*(1), 209–232.

Figure 1: Illustration of Proposition 1



If $f(\Delta + \gamma) < f(\Delta)$, an increase in Δ will induce relatively more low-a types to enroll in the treatment.

Table 1: Variance tests

	Full sample			
	Wages			
	1940	1950	1960	1970
White SD	9,667	12,148	17,974	24,562
Black SD	4,275	6,818	9,394	13,901
P-value (Black/white)	1	1	1	1
Bootstrap CI	[0.43, 0.46]	[0.54, 0.58]	[0.51, 0.54]	[0.55, 0.59]
	Log wages			
	1940	1950	1960	1970
White SD	1.04	1.00	.95	.92
Black SD	0.89	1.00	.99	.96
P-value (Black/white)	1	.74	0	0
Bootstrap CI	[0.84, .0.87]	[0.96, 1.02]	[1.02, 1.06]	[1.02, 1.06]
	Above-median education			
	Wages			
	1940	1950	1960	1970
White SD	11,550	13,862	20,625	32,521
Black SD	4,741	7,468	10,266	15,946
P-value (Black/white)	1	1	1	1
Bootstrap CI	[.39. .43]	[.51, .57]	[.48, .51]	[.47, .52]
	Log wages			
	1940	1950	1960	1970
White SD	0.97	0.97	0.94	1.10
Black SD	0.90	0.97	1.01	.98
P-value (Black/white)	1	.39	0	1
Bootstrap CI	[.89, .95]	[.96, 1.05]	[1.05, 1.10]	[.87, .92]

Notes: P-values are for normality-based tests that $H_0 : SD(\text{White})/SD(\text{Black}) = 1$ against $H_1 : SD(\text{Black})/SD(\text{White}) > 1$. Numbers in brackets are 95% percentile-t based nonparametric bootstrap confidence intervals based on 999 replications.

Table 2: Migration rates

	1940	1950	1960	1970
	Full sample			
Black-white	0.0612***	0.127***	0.137***	0.137***
	(0.00789)	(0.0103)	(0.00897)	(0.00833)
White	0.134***	0.170***	0.198***	0.191***
	(0.00774)	(0.00982)	(0.0109)	(0.0104)
N	123,927	42,766	152,930	164,396
	Above-median education			
	1940	1950	1960	1970
Black-white	0.144***	0.199***	0.205***	0.175***
	(0.0107)	(0.0132)	(0.0112)	(0.0103)
White	0.147***	0.194***	0.200***	0.204***
	(0.00941)	(0.0123)	(0.0118)	(0.0121)
N	55,590	19,625	69,628	50,809

Notes: Sample consists of Southern-born men greater aged 16-64, or those with above decade-and-race-specific educational attainment. Standard errors clustered on state-year of birth. *** p<0.01, ** p<0.05, * p<0.1.

Table 3: Covariance tests

	Full sample		Above-median education	
	Black wages	White wages	Black wages	White wages
1940	.21	.15	0.12	-.12
	(0.00)	(0.00)	(0.00)	(0.00)
	[.19, .22]	[.11, .19]	[.10, .14]	[-.17, -.05]
1950	.17	.11	.06	-.05
	(0.00)	(0.00)	(0.00)	(0.00)
	[.15, .19]	[.06, .15]	[.03, .10]	[.10, .00]
1960	.16	.10	-.08	-.17
	(0.00)	(0.00)	(0.00)	(0.00)
	[.14, .17]	[.06, .14]	[-.12, -.03]	[-.19, -.13]
1970	.17	.13	.01	-.18
	(0.00)	(0.00)	(.31)	(0.00)
	[.14, .19]	[.06, .18]	[-.09, .10]	[-.26, .00]
	Black log wages	White log wages	Black log wages	White log wages
1940	.21	.21	.12	-.14
	(0.00)	(0.00)	(0.00)	(0.00)
	[.19, .23]	[.16, .25]	[.10, .14]	[-.19, -.05]
1950	.17	.13	.06	-.026
	(0.00)	(0.00)	(0.00)	(0.00)
	[.14, .19]	[.08, .16]	[.01, .10]	[-.08, .04]
1960	.13	.08	-.07	-.19
	(0.00)	(0.00)	(0.00)	(0.00)
	[.11, .15]	[.04, .12]	[-.11, -.03]	[-.21, -.14]
1970	.13	.08	0.03	-.19
	(0.00)	(0.00)	(0.00)	(0.00)
	[.10, .15]	[.02, .13]	[-.06, .11]	[-.30, .03]

Notes—Table entries represent correlations between the black-white difference in migration probabilities and average wages (in levels and logs) among black and white men living in the South, across education×age×birthplace×birth-year cells by year. Numbers in parentheses are p-values from a normality-based test of the null of no correlation. Numbers in brackets are 95% percentile-t based nonparametric bootstrap confidence intervals based on 999 replications.

Figure 2: Black-white differences in migration rates by education and decade

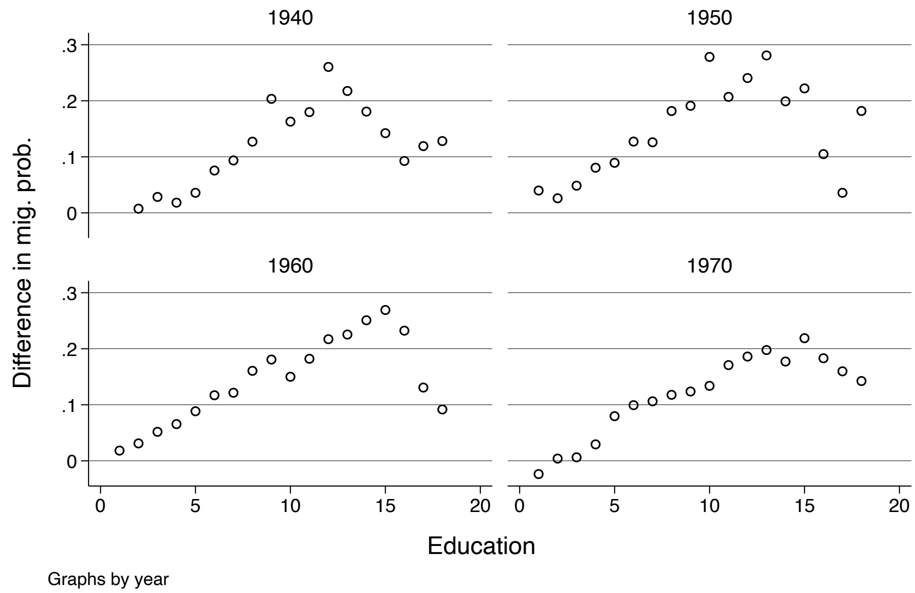


Table 4: Difference-in-difference regressions

	Wage, above-median education			
	1940	1950	1960	1970
Black	-5,491***	-6,888***	-11,279***	-15,338***
	(178.1)	(254.3)	(254.6)	(675.4)
North	4,598***	3,690***	5,978***	7,900***
	(401.3)	(494.0)	(643.6)	(1,304)
Black×North	-1,216***	1,740***	781.1*	1,574
	(345.1)	(473.3)	(461.6)	(979.2)
N	55,590	19,625	69,628	50,809

	Log wage, above-median education			
	1940	1950	1960	1970
Black	-0.843***	-0.654***	-0.714***	-0.494***
	(0.0181)	(0.0279)	(0.0197)	(0.0299)
North	0.241***	0.172***	0.173***	0.219***
	(0.0355)	(0.0408)	(0.0420)	(0.0586)
Black×North	0.320***	0.389***	0.324***	0.233***
	(0.0311)	(0.0430)	(0.0285)	(0.0381)
N	38,600	15,048	57,746	44,373

Notes: Sample consists of Southern-born men greater aged 16-64 with above decade-and-race-specific median educational attainment. Wage is defined as all income from wages in the year before enumeration. Standard errors clustered on state-year of birth. *** p<0.01, ** p<0.05, * p<0.1.

Table 5: Experimental and observational group comparisons for the NSW

	Experimental	
	White	Black
Treated−control	802.80	2028.67
	(1,331.86)	(706.10)
Black–white difference	1,225.87	
	(1703.48)	
N	74	371
	Observational	
	White	Black
Treated−control	1,856.23	156.43
	(1809.45)	(951.91)
Black-white difference	-1,699.81	
	(2125.07)	
N	62	245

Notes: "Experimental" denotes estimates obtained from the sample of National Supported Work Demonstration experimental treatment and control observations. "Observational" denotes a sample that combines treated units from the NSW data with control units from the Panel Study of Income Dynamics.

Table 6: Identification requirements for the NSW

	White	Black
P(Treat)	.40	.49
SD earnings 1974	3,558	4,202
P-value (Black/white)	.06	
Bootstrap CI	[0.90, 1.70]	
SD earnings (1975)	2,088	2,992
P-value (Black/white)	0.00	
Bootstrap CI	[1.16, 1.91]	
Corr[ΔP(Treat), Earnings 74]	-.84	.31
	(0.00)	(0.00)
	[-.89, -.73]	[.16, .44]
Corr[ΔP(Treat), Earnings 75]	-.77	.05
	(0.00)	(0.44)
	[-.85, -.68]	[-.15, .23]

Notes: Numbers in parentheses are p-values for SD ratio from a normality-based test that $H_0 : SD(\text{High-rate})/SD(\text{Low-rate}) = 1$ against $H_1 : SD(\text{Black})/SD(\text{White}) > 1$. Numbers in brackets are 95% percentile-t based nonparametric bootstrap confidence intervals based on 999 replications. $\Delta P(\text{Treat}|X)$ denotes $\hat{P}(\text{Treat}|X, \text{Black}) - \hat{P}(\text{Treat}|X, \text{White})$. P-values for correlations in parentheses.