Roy-model bounds on differential treatment effects

John Gardner*

September 2017

Abstract

In observational studies of treatment effects, concerns about positive selection bias are almost always motivated by the hypothesis that counterfactual outcomes and enrollment into the treatment are determined by a Roy model. I show that, if two groups enroll at different rates according to such a model, simple between-group differences in treated-untreated mean outcome differences often identify a lower bound on the group difference in the average effect of the treatment on the treated (ATT). When the ATTs are nonnegative, this difference is also a lower bound on the ATT itself for the high-treatment-rate group. I apply the results to interpret black-white differences in North-South wage differences in terms of the causal effects of the Great Migration.

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

JEL Codes: C50, C34, J71, R23.

*jrgardne@olemiss.edu, 366 Holman Hall, Department of Economics, University of Mississippi, University, MS, 38677. I thank Robert Miller, Lowell Taylor, John Conlon, several anonymous referees, and seminar participants at Carnegie Mellon University, The University of Mississippi, and the Western Economics Association annual conference for their insights.

1 Introduction

The central threat to the validity of observational studies of treatment effects is the possibility that nonrandom assignment of treatment status introduces selection bias into comparisons between treated and untreated units. A typical concern is that treated units are positively selected on outcomes—that is, would have experienced better outcomes even absent the treatment—leading observational analyses to overstate the causal effect of the treatment. Such concerns are almost always motivated, if only implicitly, by a hypothesis that outcomes and enrollment are realizations from the equilibrium of a Roy model in which individuals enroll in the treatment if they expect to benefit from it.

In this paper I show that, when two different groups enroll in a treatment at different rates according to such a model, group differences in the likelihood of enrollment can often be used to recover some of the causal content of comparisons between treated and untreated units. In particular, the difference between the high- and low-treatment-rate groups in differences in mean outcomes between treated and untreated units often identifies a lower bound on the group difference in the average effect of the treatment on the treated (ATT). When the low-rate group's ATT is nonnegative this group difference is itself a lower bound on the ATT for the high-treatment-rate group. Consequently, my results are particularly useful when interest centers on group heterogeneity in the effect of the treatment or when theory or prior evidence suggest that the treatment is, at worst, ineffective. The identification results developed below can also be viewed as providing a causal interpretation of differences in differences when, in contrast to standard research designs, it is applied across groups instead of over time.

The intuition behind the identification argument is straightforward. When enrollment follows a Roy model of positive selection, the difference in mean outcomes between the treated and untreated represents a combination of the ATT and a selection bias term that arises because treated units would have experienced better outcomes even absent the treatment, leading treated-untreated comparisons to overstate the causal effect of the treatment. If some members of two different groups enroll in the treatment, but members of one group do so with lower probability, we might expect the bias component to be larger for the low-treatment-rate group, among whom enrollment into the treatment is more selective.[1] In this case, subtracting the treated-untreated outcome difference for the low-treatment-rate group from that difference for the high-rate group (i.e., taking differences in differences) will over-control for selection bias among the high-rate group, bounding the group difference in ATTs from below.

The formal identification results in this paper provide conditions under which this intuition holds true. The lower-bound identification arguments that I present apply under functional form and distributional assumptions on the underlying Roy model that are weaker than those used

[1] To push this intuition further, suppose that individuals enroll if their counterfactual untreated outcomes exceed some threshold and that the treated population is relatively small. If the impact of a small increase in the enrollment threshold on the relative size of the treated population is negligible, this change will be spread across a relatively small treated population, increasing mean counterfactual untreated outcomes by more for the treated population than the untreated population; selection bias and enrollment will move in opposite directions.

in standard theoretical and econometric models of discrete choice, sample selection, truncation, reliability, and duration. When identification requires restrictions that are unlikely to be acceptable on the basis of theory or prior evidence, I provide heuristic tests for their consistency with observed outcomes and enrollment. Though my identification results view enrollment and outcomes through the behavioral lens of a Roy model, such a model almost always underlies concerns about positive selection into treatment; if the assumption that enrollment follows a Roy model is implausible or inappropriate, positive selection is unlikely to be a concern.

Under the least restrictive conditions, my identification results deliver lower bounds on the difference in average treatment effects between treated members of the high- and low-treatment-rate groups, making them particularly appropriate when interest centers on group heterogeneity in treatment effects. Though most studies focus on the treatment effect itself, inequality in the causal effect of a treatment is often more interesting. In the empirical portion of this paper, for example, I apply my identification results to interpret black-white differences in North-South wage differentials in terms of racial differences in the causal effect of Northward migration on wages during the African American Great Migration period. In this setting, the group difference in treatment effects conveys (at least partial) information about the wage effects of discrimination.

Furthermore, if theory or prior empirical evidence suggest that the ATT is nonnegative for both groups (or at least for the low-treatment-rate group), the group difference in these effects can also be interpreted as a lower bound on the ATT itself for the high-rate group. In most Roy models where enrollment is determined primarily by counterfactual outcomes comparisons, the equilibrium average effect of the treatment on the treated is nonnegative by definition; otherwise no one would enroll. This logic may not apply if individuals incorrectly forecast how much they will benefit from the treatment or if the enrollment decision is influenced by non-outcome factors that correlate positively with counterfactual outcomes (equivalently, if enrollment is determined with respect to different outcomes than the one under study). High-latent-skill individuals, e.g., may be more likely to sign up for a job training program that they mistakenly believe will increase their earnings, or if they enjoy receiving training, even one they know will decrease their earnings. Accordingly, conclusions about treatment effects themselves drawn using my results must be tempered by the prior degree of belief in the hypothesis that the ATT is nonnegative. The group difference in these effects is bounded regardless of this belief.

Although my results only deliver partial identification, in applications a bound may suffice to answer the question at hand. The conservative lower bounds that my results identify are likely to be the ones that are of most use when there is concern that positive selection causes outcome comparisons to overstate causal effects. Manski (1989; 1990) has shown that partial identification can permit meaningful inference about treatment effects when there is doubt that the conditions required by traditional point estimators hold. Similarly, my results may be applied when the available data contain no credible sources of variation in treatment status. Lower bounds identified by interpreting group differences in treatment rates in terms of a coherent behavioral model may be preferable to point estimates based on ad hoc instruments or otherwise-questionable sources of

quasi-experimental variation.

This paper draws from, and builds on, a large literature on problems of sample selection. Roy's "Some Thoughts on the Distribution of Earnings" was published in 1951 and has provided social scientists with a framework for thinking about self-selection ever since. Borjas (1988) uses a Roy model to characterize the selection of immigrants from different source countries, while Dahl (2002) combines a multi-market Roy model with a semiparametric implementation of Heckman's (1979) sample-selection correction to identify state-specific returns to education in the US, to give just two examples. Like these papers, mine is also closely related to the literature on the econometrics of truncation, sample selection, and regime switching (see Tobin, 1958; Heckman, 1976, 1979; Amemiya, 1984). There is also an extensive literature on the identifiability of the Roy model, broadly construed, from variation in observables (Heckman and Honoré, 1990), and hence the identification, estimation and interpretation of treatment effects from within the Roy framework (see Heckman and Vytlacil, 1999; Heckman, Urzua, and Vytlacil, 2006; D'Haultfoeuille and Maurel, 2013; Maurel and D'Haultfoeuille, 2013; D'Haultfoeuille, Maurel, and Zhang, 2014; Eisenhauer, Heckman, and Vytlacil, 2015). While these papers analyze identification and estimation of the Roy model itself (i.e., the distributions of counterfactual, or related, outcomes), the objective of this paper is to use the Roy framework to determine what differences in differences identify when applied across groups, rather than over time. Furthermore, my identification results ultimately derive from variation in (observable) treatment probabilities, rather than the exclusion-restriction or identification-at-infinity (as in Chamberlain, 1986) approaches taken elsewhere in the literature.

In Section 2 of this paper, I develop the lower-bound identification argument under a simple Roy model in which enrollment and counterfactual outcomes are linear functions of an unobserved random variable. In Section 3, I show that similar results can be obtained under less-restrictive functional form assumptions. In Section 4, I apply the identification results to interpret black-white differences in North-South wage differentials during the African American Great Migration in terms of racial differences in the causal effect of migration on wages. I conclude in Section 5.

2 Identification under a linear Roy model of outcomes and enrollment

2.1 The model

Let $d \in \{0, 1\}$ be an indicator for whether an individual chooses to enroll in a binary treatment. In addition, suppose that individuals differ according to exogenous and observable membership in one of two groups $g \in \{l, h\}$, where members of group h are, by definition, more likely to receive the treatment (e.g., for a treatment that more men than women receive, men belong to h and women to l).

Further suppose that the counterfactual outcome y_{dg} that a member of group g would receive

upon choosing enrollment status d depends on an unobserved random variable a according to

$$y_{dg} = \beta_{dg} + \gamma_{dg} a, \qquad \gamma_{dg} \geq 0, \tag{1}$$

for $d \in \{0, 1\}$ and $g \in \{l, h\}$. In this model, a represents an unobserved factor that influences both counterfactual outcomes and enrollment. For example, the decision to enroll in a job-training program may depend on the same index a of skill that determines counterfactual earnings y_{dg} with and without completing the program.

Reflecting the hypothesis that treated individuals are positively self-selected on outcomes, suppose that individuals enroll in the treatment if their realizations of a exceeds some (group-specific) threshold, so that the enrollment decision takes the form

$$d_g(a) = 1(a \geq \hat{a}_g) \tag{2}$$

for $d \in \{0, 1\}$ and $g \in \{l, h\}$, where $1(\cdot)$ is the indicator function. An enrollment decision rule like (2) might arise if, e.g., individuals sign up for a job training program when the net earnings benefit is positive.[2] Using (1) and (2), realized outcomes can be expressed as $y = y_0(1 - d) + y_1 d$. Throughout the remainder of this paper, I use the notation "l" to denote the event "$g = l$" that an individual belongs to the low-treatment-rate group (and similarly for the high-rate group) and the notation "0" to denote the event "$d = 0$" that an individual does not receive the treatment (and similarly for the treated), so that, e.g., $E(a|l, 0)$ denotes the mean of a among untreated members of group l, etc.

Before continuing the identification argument, a few notes about the model and its interpretation are in order. First, if there is concern about selection in terms of a particular unobservable factor, a may also be viewed as a nonlinear function of that factor. In addition, when there is concern about self-selection, but not on any one articulable factor, a may be viewed as the untreated outcome itself. Second, though I abstract away from observable covariates, they can be accommodated either by repeating the identification argument within covariate strata or by applying the argument across strata under the assumption that a is a scalar index of observable covariates and unobserved factors (this approach is more conventional, though it imposes stronger restrictions), thus relaxing the restriction that a single factor explains enrollment and counterfactual outcomes.

Lastly, the above is a somewhat specialized Roy model in which treated and untreated outcomes are perfectly correlated. This purpose of this simplification is to focus on the usual "omitted-variable" econometric interpretation of selection bias, in which the same unobserved factor determines enrollment as well as treated and untreated outcomes. Though it may appear restrictive at first glance, this assumption is compatible with more complex models of enrollment behavior. One interpretation of the model is that individuals must make their enrollment decisions based on ex ante expectations of counterfactual outcomes conditional on a, and these expectations are linear, as in (1). Follow-

[2]Note that (2) allows for the possibility that non-outcome factors that are related to a influence the enrollment decision.

ing Olsen (1980) and Wooldridge (2002, Assumption 17.1), an alternative interpretation is that a represents the total utility from enrolling (which may be the product of a confluence of factors), individuals enroll if that utility exceeds some threshold, and expected counterfactual outcomes are linear in a (or at least approximately linear; if a and the y_{dg} are multivariate normal, they can be viewed as the error components of a switching or Type-5 Tobit regression model, and this linearity is automatic, see Amemiya 1984). Both interpretations allow for the possibility that counterfactual outcomes are only imperfectly correlated, as in more standard Roy models.

The purpose of the identification argument developed below is to use the difference in mean outcomes between treated and untreated members of the low-treatment-rate group l to bound the degree to which that same comparison for members of the high-rate group is contaminated with bias due to positive selection. If there are arbitrary group differences in the distributions governing the unobserved variable a, however, this approach requires detailed knowledge of the group-specific distribution functions F_g for a. A more conventional way of modeling such heterogeneity (and one that is particularly apt when both groups belong to the same parent population) is to treat the group-specific distributions as belonging to the same overarching family of distributions, if only as an approximation:

Assumption 1. *The distributions F_g, $g \in \{l, h\}$, of a belong to the same scale or location-scale family (with F_l normalized as the standard distribution).*

Under this assumption, both groups' unobservables can be modeled in terms of realizations from the low-rate-group's distribution according to $F_h(a) = F_l[(a - \mu_h)/\sigma_h]$, where μ_h and σ_h are the location and scale parameters for the high-rate group (and $\mu_l = 0$ and $\sigma_l = 1$ by the normalization).[3] Because the groups $g \in \{l, h\}$ are defined by the relative likelihood with which they enroll in the treatment, Assumption 1, along with enrollment equation (2), also implies that

$$\frac{\hat{a}_h - \mu_h}{\sigma_h} < \hat{a}_l. \tag{3}$$

In other words, group differences in treatment rates are explained by a combination of differences in preferences for enrolling and differences in the distributions of the unobserved factors that determine enrollment preferences and counterfactual outcomes.

Under Assumption 1, group-specific expected realizations of a also satisfy $E(a|a \gtreqless \hat{a}_h, h) = \mu_h + \sigma_h E[a|a \gtreqless (\hat{a}_h - \mu_h)/\sigma_h, l]$, clarifying the interpretation of an additional assumption that simplifies the analysis below:

Assumption 2. $\gamma_{0h}\sigma_h \leq \gamma_{0l}$.

This assumption, which asserts that members of the low-treatment-rate group would experience a greater increase in untreated outcomes in response to a one-standard-deviation increase in a than otherwise-similar members of the high-rate group, is consistent with (though not implied by) the

[3] If the F_g belong to a scale, rather than location-scale family (e.g., because a is distributed over $\mathbb{R}^+$ with a mean that depends on its scale parameter), this expression applies with the μ_h term omitted.

assumed group differences in treatment rates. It also places no restrictions on the relationship between the unobservable a and treated outcomes. Unlike the rest of the structure imposed so far, because Assumption 2 places specific restrictions on the parameters of the model, it cannot be viewed as a a simplifying approximation, and is unlikely to be acceptable on theoretical grounds alone. The following result shows that the ratio of observed outcome variances between untreated members of the the high- and low-rate groups can be used to assess whether the data are consistent with the assumption. The strength of the test depends on whether observations on outcomes are available before selection into treatment has taken place (either panel or repeated cross-section) and whether the densities f_g of a are log concave or log convex.[4]

Proposition 1. *Suppose that equations (1) and (2) and Assumption 1 hold. Then:*

1. *If pre-treatment-period data are available,* $Var(y|h)/Var(y|l) \gtreqless 1$ *implies that* $\gamma_{0h}\sigma_h \gtreqless \gamma_{0l}$ *(where* y *denotes observed outcomes in the pre-treatment period).*

2. *If only post-treatment-period data are available and the* f_g *are log concave,* $Var(y|h,0)/Var(y|l,0) > 1$ *implies that* $\gamma_{0h}\sigma_h > \gamma_{0l}$ *(where* y *denotes observed outcomes in the post-treatment period). Otherwise, if the* f_g *are log convex,* $Var(y|h,0)/Var(y|l,0) < 1$ *implies that* $\gamma_{0h}\sigma_h < \gamma_{0l}$*.*

All proofs are presented in Appendix B. Heuristically, the conclusion of the proposition is that the data are consistent with Assumption 2 if the untreated variance ratio is less than one, though the precise nature of the tests that it suggests depend on the structure of the available data and the distributions of the a.[5]

2.2 Identification

When outcomes follow (1) and enrollment follows (2), the treated-untreated mean outcome difference can be decomposed into average effect of the treatment on the treated (ATT) and selection bias components as

$$\begin{aligned} E(y|1,g) - E(y|0,g) &= [\beta_{1g} + \gamma_{1g}E(a|a \geq \hat{a}_g, g)] - [\beta_{0g} + \gamma_{0g}E(a|a < \hat{a}_g, g)] \\ &= \underbrace{[(\beta_{1g} - \beta_{0g}) + (\gamma_{1g} - \gamma_{0g})E(a|a \geq \hat{a}_g, g)]}_{\text{ATT}_g} \\ &\quad + \underbrace{\gamma_{0g}[E(a|a \geq \hat{a}_g, g) - E(a|a < \hat{a}_g, g)]}_{\text{Bias}_g}. \end{aligned} \tag{4}$$

The bias term in (4) arises because, when enrollment into the treatment follows (2), treated units have higher mean realizations of a and, consequently, would experience better outcomes even absent

[4] A function is log concave if its log is concave. Note that since log concavity is preserved under both truncation and linear transformation (Bagnoli and Bergstrom, 2005), it is possible to use data on outcomes to test whether the f_g are log concave, regardless of whether outcomes are observed before the treatment (An, 1995), and both of the f_g will have the same log concavity.

[5] This heuristic interpretation is particularly appropriate when the treatment rate is small for both groups, in which case the untreated variance more closely approximates the unconditional variance.

the treatment.[6]

Standard difference-in-differences research designs attempt to eliminate the bias term from (4) by using repeated observations on outcomes to subtract from the treated-untreated mean outcome difference the same difference recorded for a pre-treatment period in which the ATT is zero for both groups. The present model, in contrast, has no time dimension. Instead, the question is what differences in differences recovers when applied across groups, rather than over time.[7] In general, the group difference in treated-untreated mean outcome differences (differences in differences hereafter) represents the sum of group differences in ATT and bias terms, which under (1) and (2) can expressed as

$$\begin{aligned}
&[E(y|1,h) - E(y|0,h)] - [E(y|1,l) - E(y|0,l)] \\
&\quad = \underbrace{\{[(\beta_{1h} - \beta_{0h}) + (\gamma_{1h} - \gamma_{0h})E(a|a \geq \hat{a}_h, h)] - [(\beta_{1l} - \beta_{0l}) + (\gamma_{1l} - \gamma_{0l})E(a|a \geq \hat{a}_l, l)]\}}_{\text{ATT}_h - \text{ATT}_l} \\
&\qquad + \underbrace{\{\gamma_{0h}[E(a|a \geq \hat{a}_h, h) - E(a|a < \hat{a}_h, h)] - \gamma_{0l}[E(a|a \geq \hat{a}_l, l) - E(a|a < \hat{a}_l, l)]\}}_{\text{Bias}_h - \text{Bias}_l}. \quad (5)
\end{aligned}$$

When members of two different groups receive the treatment with positive probability, it is likely that the ATT is nonzero for both groups and that the group-specific bias terms differ, in which case differences in differences will not identify the ATT for either group. Following the intuition developed above, however, we might expect the treated-untreated outcome comparison for the low-treatment-rate group l to be more contaminated with selection bias than for the high-rate group h. In this case, taking differences in differences will over-control for selection bias among members of the high-rate group, bounding the group difference in ATTs from below. Moreover, if theory or prior empirical evidence suggest that the ATT is nonnegative for the low-treatment-rate group, a lower bound on the group difference in ATTs is also a conservative lower bound on the ATT itself for high-rate group.

Rearranging the differential bias term in (5) (the second term in braces) and applying Assumption 1 (to express the group-specific truncated expectations of a in terms of realizations from the low-rate group's distribution) shows that differences in differences will bound the group difference in ATTs from below if

$$\frac{\gamma_{0h}\sigma_h}{\gamma_{0l}} \leq \frac{E(a|a \geq \hat{a}_l, l) - E(a|a < \hat{a}_l, l)}{E[a|a \geq (\hat{a}_h - \mu_h)/\sigma_h, l] - E[a|a < (\hat{a}_h - \mu_h)/\sigma_h, l)]}. \quad (6)$$

If Assumption 2 holds, the left-hand side of (6) can be replaced with one. In this case, a sufficient condition for (6) is that difference between the left- and right-truncated means is increasing in the

[6]The results that follow may be interpreted in terms of proportional, rather than absolute, treatment effects by letting y represent the log of the outcome of interest. If $y_{dg} = \gamma_{dg}a$, the proportional treatment effects γ_{1g}/γ_{0g} are constant, and the results may be applied to both the absolute and proportional treatment effects, provided that the distributions of both a and $\log a$ satisfy the appropriate conditions.

[7]However, replacing group membership with time results in the fuzzy differences-in-differences estimator developed in de Chaisemartin and D'Haultfoeuille (2017), in which case the identification arguments below can be applied to bound the change in the ATT over time (I thank an anonymous referee for pointing this out).

truncation point:[8]

$$\frac{d}{d\hat{a}}\left[E(a|a \geq \hat{a}, l) - E(a|a < \hat{a}, l)\right] \geq 0 \qquad \text{for } \hat{a} \geq \frac{\hat{a}_h - \mu_h}{\sigma_h}. \tag{7}$$

To understand this condition, consider the effects of a small increase in the enrollment threshold on the mean of a among the treated and the untreated (i.e., on the left- and right-truncated means). Such an increase will cause the marginal treated unit (the unit with the lowest realization of a) to exit the treatment, increasing the mean of a among the treated. Since this unit now has the highest realization of a among the untreated, the change will also increase the mean of a among the untreated. What (7) requires is that the former effect dominates, so that the difference in mean a between the treated and untreated increases. Put differently, (7) states that selection bias is decreasing in the probability of receiving the treatment, so that the low-treatment-rate group can be used as a "quasi-control" for the high-rate group in order to recover some of the causal content of the treated-untreated mean outcome difference for the latter group.

The identification of a lower bound on group difference in ATTs then becomes a question of whether (7) is likely to hold. The following result establishes two classes of distributions for which this property holds over some subset of the support. Moreover, and as I discuss below, the distributions used in standard theoretical and econometric models fall into one of these classes.

Proposition 2. *Suppose that a is distributed over $[L, \infty]$ with density f satisfying $\lim_{a\to\infty} f = 0$. Then:*

1. *If $E(a|a \geq \hat{a})$ is convex and $E(a|a < \hat{a})$ is concave, there exists an a^* such that*

$$\frac{d}{d\hat{a}}[E(a|a \geq \hat{a}) - E(a|a < \hat{a})] \geq 0$$

   for all $\hat{a} \geq a^*$. If f is symmetric, then a^* is the mean. If the mean exceeds the median, then a^* is less than the median.

2. *If f is log convex and $f' \leq 0$ for all a,*

$$\frac{d}{d\hat{a}}[E(a|a \geq \hat{a}) - E(a|a < \hat{a})] \geq 0$$

   for all $\hat{a}$.

In Appendix A, I show that the conclusions of Proposition 2 can also be derived from more fundamental restrictions on the underlying density functions (these restrictions, which amount to requiring that the log of the density become less concave as the density itself is decreases, are closely related to the log concavity condition that Heckman and Honoré, 1990 show is pivotal for identification of the Roy model). However, the usefulness of the proposition hinges on whether it is reasonable to believe that the distributions of the unobserved determinants of counterfactual

[8] I assume throughout that means are finite and functions are differentiable.

outcomes and enrollment belong to one of its classes. Figures 1 and 2 show that the distributions used in common theoretical and econometric models of sample selection, truncation, discrete choice, duration, and reliability fall into one of these classes. The first figure shows the left- and right-truncated expectations, and their difference, for parameterizations of the normal, logistic, uniform, gamma, Weibull, exponential and logistic distributions. For each of these distributions, the left-truncated expectation is convex, the right-truncated expectation is concave, and there is a point in the support beyond which the difference in truncated means is increasing. The second figure shows the same functions for (in some cases, different) parameterizations of the gamma, Weibull, Pareto, and lognormal distributions. With the exception of the lognormal, each of these distributions belongs to the second class defined in the proposition. For each distribution, the left- and right-truncated expectations are concave and the difference in truncated means is increasing on the entire support.[9]

Thus, the conditions of Proposition 2 are much weaker than those used in standard theoretical and econometric models. If the a are drawn from any distribution that satisfies them, differences in differences will bound the group differences in ATTs from below as long as both groups receive the treatment with sufficiently low probability (i.e., if the group-specific enrollment thresholds $\hat{a}_g$ exceed the points a_g^* beyond which the differences in truncated means are increasing). In some applications, theory or prior empirical evidence may be informative about the treatment probabilities below which the lower-bound result holds (e.g., if the distributions belong to the second class in Proposition 2, the result applies at any treatment probability).

When it is unreasonable to assume such detailed knowledge of the distributions governing the a, the lower-bound procedure can still be applied if theory or prior evidence suggest that these distributions are weakly right-skewed in the (informal) sense that their means are no less than their medians (indeed, many social and economic phenomena satisfy this condition, which is weaker than the more common assumption of symmetry). In this case, differences in differences will bound the group difference in ATTs from below for treatment probabilities less than one-half (unless the distributions are symmetric, the result will apply at even greater treatment probabilities).[10] I summarize this last, most conservative, observation as a theorem.

Theorem 1. *Suppose that equations (1) and (2) and Assumptions 1 and 2 hold, the f_g belong to one of the classes defined in Proposition 2, and the group means of a are no greater than their medians. Then differences in differences identifies a lower bound on the group difference in ATTs if both groups are treated with probability less than one half. If the ATT is nonnegative for the low-treatment-rate group, this is also a lower bound on the ATT itself for the high-rate group.*

[9] The lognormal density is neither log concave nor log convex on its entire support (Bagnoli and Bergstrom, 2005), and is not everywhere decreasing. Nevertheless, the difference in truncated means for its second parameterization is everywhere increasing, showing that the conditions of Proposition 2 are sufficient but not necessary.

[10] To give a specific example, suppose that the a are drawn from a standard normal distribution, so that $E(a|a \geq \hat{a}) - E(a|a < \hat{a}) = \lambda(\hat{a}) + \lambda(-\hat{a})$, where $\lambda = \phi/\Phi$ is the ratio of the standard normal density and distribution functions. The function $\lambda(\hat{a}) + \lambda(-\hat{a})$ is strictly convex (see Heckman and Honoré, 1990, Appendix A) and, by inspection, has a critical point at zero. Thus, for treatment probabilities less than one half, the difference in truncated means is increasing in the point of truncation, and differences in differences bounds the group difference in ATTs from below.

3 Identification under a nonlinear Roy model

3.1 The model

The preceding analysis uses simplifying linearity assumptions to model the processes that determine counterfactual outcomes. Though in applications these linearity assumptions may be acceptable as first-order approximations, in this section I show that similar results can be obtained under less restrictive conditions by directly modeling the relationship between counterfactual outcomes and enrollment. The price of this additional generality is that the conditions required for the identification result to apply are more difficult to verify, though I also suggest a heuristic test for their consistency with the data.

Suppose that the enrollment decision rule takes the form

$$d_g(a, \epsilon) = 1\left[\tilde{\gamma}(\Delta_g, a) - \epsilon \geq 0\right] \tag{8}$$

for $g \in \{l, h\}$, where $\tilde{\gamma}(\Delta_g, a)$ is some function of the random variable a that determines counterfactual outcomes and a secular cost or benefit shifter Δ_g that contributes to group differences in enrollment, and ϵ represents the influence of idiosyncratic factors on the enrollment decision. This decision rule is analogous to the rule introduced in the linear Roy model in Section 2 with the exception that (8) allows for idiosyncratic factors that are unrelated to a to influence enrollment separately. This interpretation makes the following natural.

Assumption 3. *ϵ is distributed over $\mathbb{R}$ independently of a and g with log concave distribution function F.*

For example, the error distributions used in standard discrete choice models such as the logit and probit satisfy this assumption. An additional assumption simplifies the statement of the main identification result while allowing the enrollment decision to be modeled flexibly:

Assumption 4. *$\tilde{\gamma}(\Delta_g, a)$ takes one of the semi-parametric forms*

$$\tilde{\gamma}(\Delta_g, a) \in \{\Delta_g + \gamma(a), \Delta_g \gamma(a), \gamma(\Delta_g + a)\},$$

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

The first part of Assumption 4 allows the marginal utility of enrolling (and consequently, group differences in enrollment rates) to be constant, increasing, or decreasing in the cost or benefit shifters Δ_g. The requirement that γ be concave increasing reflects the notion of positive selection into enrollment and a standard hypothesis that the marginal utility of enrolling is (at least weakly) diminishing in a. As I discuss below, the assumption that γ is CRRA is a concise way to rule out pathologies in the relationship between selection bias and the probability of enrollment.[11]

[11] Assumption 4 sacrifices some generality in the interest of parsimony. In particular, if the a are distributed independently of group membership, the main identification result holds for any increasing function $\tilde{\gamma}$. Note also

Suppose that counterfactual outcomes are determined by

$$y_{dg} = y_{dg}(a), \qquad y'_{dg}(a) \geq 0, \tag{9}$$

for $d \in \{0, 1\}$ and $g \in \{l, h\}$. To accommodate nonlinear dependence of counterfactual outcomes and enrollment on the random variable a, while allowing for group heterogeneity in the distributions from which that variable is drawn, suppose also that the following holds.

Assumption 5. *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.*

Finally, suppose that counterfactual outcomes satisfy the nonlinear analog of the relative slope condition (Assumption 2) required for identification in the linear case:

Assumption 6. $y'_{0h}(a)\sigma_h \leq y'_{0l}(a)$ *for all* a.

As in the linear case, Assumption 6 is consistent with observed group differences in treatment rates, may be justified by prior evidence or *a priori* theoretical considerations, and places no restriction on treated outcomes. In addition, the untreated outcome variance ratio test developed above (Proposition 1) can also be applied when outcomes are nonlinear in a, though the test is weaker in this case, providing only approximate information about whether Assumption 6 holds on average:

Proposition 3. *Suppose that equations (8) and (9) and Assumption 5 hold. Then, by a first-order approximation:*

1. *If pre-treatment-period data are available, $Var(y|h)/Var(y|l) > 1$ implies that $E[y'_{0h}(\sigma_h a)\sigma_h|l] > E[y'_{0l}(a)|l]$ (where y denotes observed outcomes in the 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 of the identification argument in the present model is similar to in the linear case: if selection bias is decreasing in the probability of enrollment, differences in differences will over-control for selection bias among the high-rate group, bounding the group difference in ATTs from below. In what follows, I provide an intuitive sketch of the result for the nonlinear case, presenting the formal details in Appendix B in the interest of exposition.

that if $\tilde{\gamma}$ is linear in γ, the assumption that distribution F of the ϵ is group independent can be relaxed by assuming that the group-specific densities belong to the same location-scale family. Although none of the identification results presented below require that γ is concave, this is the most natural case.

Under (9), the difference in mean outcomes between treated and untreated individuals can be decomposed into ATT and selection bias terms according to

$$\begin{aligned} E(y|1,g) - E(y|0,g) &= \int y_{1g}(a)p(a|1,g)da - \int y_{0g}(a)p(a|0,g)da \\ &= \underbrace{\int \left[y_{1g}(a) - y_{0g}(a)\right] p(a|1,g)da}_{\text{ATT}_g} + \underbrace{\int \left[p(a|1,g) - p(a|0,g)\right] y_{0g}(a)da}_{\text{Bias}_g}, \end{aligned} \tag{10}$$

where $p(a|g,d)$ denotes the density of a conditional on group membership g and treatment status d. The difference in (11) between the high- and low-treatment-rate groups will bound the group differences in ATTs from below if the differential bias term is negative, or if

$$\int \left[p(a|1,h) - p(a|0,h)\right] y_{0h}(a)da \leq \int \left[p(a|1,l) - p(a|0,l)\right] y_{0l}(a)da. \tag{11}$$

As in the linear case, if the ATT term is negative for the low-rate group, this difference in difference will also bound the ATT for the high-rate group from below.

Under Assumption 5 and enrollment decision rule (8), 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 distributional differences (σ_g) and different preferences for enrollment (Δ_g) according to

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

where $\theta = (\Delta, \sigma)$ and $d\theta = \theta_h - \theta_l$. A perturbation $d\theta$ of these parameters will change the likelihood of enrollment for an individual with unobservables a by $dP(1|a) = f[\tilde{\gamma}(\Delta, \sigma a)][\partial \tilde{\gamma}(\Delta, \sigma a)/\partial \theta']d\theta$. If these changes are negatively correlated with untreated outcomes—that is, if untreated outcomes are relatively low for the the typical individual induced to enroll by this perturbation—selection bias will be decreasing in the treatment probability. The following result formalizes this intuition and provides a sufficient condition for differences in differences to identify a lower bound.

Proposition 4. *Suppose that equations (8) and (9) and Assumptions 3-6 hold. Then if both groups are treated with probability less than one half, a sufficient condition for (11) is that there is a* $g \in \{l, h\}$ *such that*

$$Cov\left[f\left(\tilde{\gamma}(\Delta, \sigma a)\right) \frac{\partial \tilde{\gamma}(\Delta, \sigma a)}{\partial \theta'} d\theta, y_{0g}(\sigma_g a)\right] \leq 0 \tag{12}$$

between θ_l *and* θ_h.[12]

The proof of Proposition 4 proceeds by showing first that (11) holds if it also holds when both groups face the same untreated outcome functions and second that the derivative of the selection bias term holding untreated outcomes constant is bounded from above by the covariance defined

[12] In (12), expectations are taken with respect to the distribution of a among the low-rate group.

in (12). Note that (12) is sufficient, but not necessary. In particular, it is possible that changes in the likelihood of enrollment due to simultaneous changes in Δ and σ fluctuate across the support of a in such a way that they are are uncorrelated with untreated outcomes even though there is less selection bias for the high-treatment-rate group. The assumption that the function γ exhibit constant relative risk aversion restricts the relative magnitudes of the first and second derivatives of γ through which the effects of changes to Δ and σ operate, eliminating this possibility.[13]

In Proposition 4, the operative requirement for identification is the covariance condition (12), which asserts that group differences $P(1|a,h)-P(1|a,l)$ in the conditional probability of enrollment are negatively correlated with untreated outcomes for at least one of the groups. Because it depends on nonlinear functions of an unobserved random variable, this condition is neither directly verifiable nor likely to be acceptable on *a priori* grounds. However, if an observable proxy q for a is available, data on outcomes and enrollment can be used to estimate the sample analog of

$$Cov\left[P(1|h,q)-P(1|l,q),E(y_{0g}|q)\right]. \tag{13}$$

Since, by the law of total covariance, the unconditional covariance between group differences in enrollment probabilities and untreated outcomes can be written as the sum of (13) and the average covariance within q strata, evidence that (13) is nonpositive would add considerable credibility to the hypothesis that (12) itself holds.[14] Natural candidates for proxies with which this heuristic test can be implemented are observable covariates (which are often included in empirical treatment-effect models to proxy for unobservable determinants of enrollment and outcomes rather than because they are of direct interest).[15] Though it may seem odd to use this approach (which is similar in spirit to that of Altonji, Elder, and Taber, 2005) if a proxy is readily available, treatment effects that are point identified by controlling for a proxy may overstate or understate both the ATT and the group difference in ATTs, while it is clear what the Roy-model bound approach identifies.[16]

The other requirements of the proposition are relatively mild. In particular, the proposition itself places no restriction on the shapes of the densities π_g for a, requires only that the distribution of ϵ is log concave, applies for any increasing counterfactual outcome functions, and allows any enrollment rule belonging to the large, common, and flexible class of concave increasing CRRA functions. The

[13]There is an interesting symmetry between this condition and the sufficient condition (derived in Appendix A) for the convexity of the truncated expectation that $\log f$ exhibit declining absolute risk aversion, which is implied by constant relative risk aversion.

[14]In addition, if outcomes are only observed after enrollment decisions have been made, (13) can only be estimated with respect to the distribution of q among the untreated. 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 (and therefore high enrollment likelihoods), 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 unconditional covariance.

[15]Though this approach precludes the application of the identification argument within q strata, the argument can be applied across covariate strata by relaxing the assumption that a and ϵ are independent as long the conditional density of ϵ is log concave for each a.

[16]In the terminology of Wooldridge (2002, Chapter 4), a variable q that is correlated with a is an imperfect proxy; though a true proxy can be used to point-identify the ATT, such a variable must satisfy considerably more exacting requirements.

following theorem summarize the main result of this section.

Theorem 2. *Suppose that conditions of Proposition 4 hold. Then differences in differences identifies a lower bound on the group difference in ATTs if both groups are treated with probability less than one half. If the ATT is nonnegative for the low-treatment-rate group, this is also a lower bound on the ATT itself for the high-rate group.*

4 Application: The wages of the Great Migrants

4.1 The Great Migration

The Great Migration refers to a period of US history spanning roughly 1915-1970 during which a millions of Southern-born blacks left the South in favor of cities in the North. This episode is believed to have had profound social and economic consequences, both for the migrants themselves and for the areas to which they moved (Tolnay, 2003 provides a detailed review). In particular, it is widely acknowledged (see, e.g., Smith and Welch, 1989; Donohue and Heckman, 1991) that Northward migration played an important role in the relative economic progress experienced by blacks during the 20th century. Accounts of the magnitude of this migration often overlook the fact—which I document below—that many whites also moved North, albeit at lower rates. On average between 1940 and 1970, for example, 12% of Southern-born white men, compared to 23% of Southern-born black men, migrated to the North.

Smith and Welch (1989) and Donohue and Heckman (1991) decompose changes in black-white wage gaps into components explained by racial differences in residential location, education and other factors. An important aim of these studies is to understand the contribution of migration to black-white relative wage gains during the 20th century. This estimand can easily be recovered from black-white differences in North-South wage differentials. However, since unobserved characteristics may have contributed to the decision to migrate, and since they might have done so differently for blacks and whites, decompositions of regional wage gaps into parts explained or not by differential migration rates are necessarily descriptive and may not be informative about the causal effects of migration on wages. As Smith and Welch (1989) note in their survey of the determinants of black economic progress,

> Even among men who have the same amount of education and job experience, large geographic wage differentials prevail among regions. Identifying their underlying causes is a complex empirical problem. Some of these wage disparities reflect cost-of-living differences between regions, or compensating payments for the relative attractiveness or undesirability of locational attributes (e.g. climate, crime, and density). Given the magnitude of the regional wage differentials we estimate, it is also likely that they proxy for unobserved indices of skill. Finally, the large black-white gap in the South may well reflect the historically more intense racial discrimination there.

and

> If they proxy for unobserved skill differences, cross-sectional wage differentials would not represent the wage gain an individual would receive by moving from the South to the North.

The identification results developed above provide a framework for understanding racial differences in North-South wage differentials in terms of the causal impact of migration on wages. In fact, because they recover bounds on group-differences in average treatment effects, they are particularly well-suited to the analysis of the impacts of the Great Migration. As I note above, a limitation of my identification results is that, unless there is reason to believe that the ATT is nonnegative, differences in differences only delivers a bound on the group difference in ATTs, rather than the ATT itself. In some applications this group difference is interesting in its own right; The Great Migration is one such application. As Smith and Welch (1989) argue, equilibrium North-South wage differentials may partially reflect regional variation in amenity values and the cost of living (in addition to productivity effects), so that differencing the nominal wage effect of migration for whites from that for blacks removes the component of that effect explained by these factors (at least insofar as this component is similar for both groups). In the context of the Great Migration, then, group heterogeneity in treatment effects is suggestive about the extent to which black migrants earned more in the North because its denizens were less discriminatory than their Southern counterparts.[17]

In addition, under the hypothesis that migration did not decrease wages for white migrants (which is likely given the more industrial nature of the Great Migration North as well as the sheer sizes, documented below, of the flow of white migrants and the regional wage gaps they faced), the identification results also provide a lower bound on the average effect of Northward migration on black migrants' wages. Owing to a paucity of sources of credibly exogenous variation in migration, the magnitude of this causal effect was an open question for decades until, after constructing a new panel dataset from linked historical Census data and imputed wages, Collins and Wanamaker (2014) provided evidence on within-individual North-South wage differences for blacks. The congruity between their findings and the ones that I present below underscore the usefulness of the identification results presented in this paper (especially given the arduous nature of their data collection effort). At the same time, Collins and Wanamaker (2014) only examine blacks, and their fixed-effects identification strategy is limited in that it can only account for permanent unobserved heterogeneity (and because a lack of wage data requires them to use imputed wages). This paper builds on the empirical evidence that they provide; to my knowledge, this is the first paper to provide rigorous evidence about racial differences in the causal effect of the Great Migration on migrants' wages.

[17] Not all of the racial difference in treatment effects can be attributed to discrimination, however, since (ii) the distribution of earnings potential may have differed between blacks and whites and (ii) region-specific racial differences in wages may have partially reflect geographic differences in local housing and amenity prices coupled with racial differences in residential location choices.

4.2 Descriptive evidence

Figure 3 plots Northward migration rates by year of birth for black and white men born in the Southern US and, to avoid age effects, at least 30 years old at the time of enumeration. The data for this graph, and all further results in this section, are based on 1% Integrated Public Use Microdata Samples (IPUMS) of the 1940-1970 US Censuses (Ruggles, Alexander, Genadek, Goeken, Schroeder, and Sobek, 2010), from which I include only Southern-born black and white men.[18] For birth years prior to 1880, the sample sizes are small and the estimated migration rates are imprecise for men of both races. Past 1880, as the figure shows, black migration rates dominate white rates at all birth years, and exhibit a steeper trend. For example, a white man born in the South around 1940 had about a 20% chance of migrating to the North, while his black counterpart had about a 35% chance. To examine these differences in migration rates in greater detail, I present in Table 1 linear models of the probability of migrating between 1940 and 1970. Pooling across all four decades of Census data, the average probability of migrating for whites was about 12%; for blacks, it was twice as high at about 23%. The decade-specific regressions show that the white migration rate increased from about 13% in 1940 to 19% by 1970 and that the black-white difference in migration rates increased during this period as well. Within decades, the pattern of black-white differences in migration rates are similar.

Table 2 presents regression estimates of the black-white difference in North-South (annual) wage and log wage differentials over the same periods.[19] The top panel of the table presents the results for wages in levels. On average between 1940 and 1970, the estimated black-white difference in North-South wage gaps was about $3,300. This average is driven by decade-specific gaps that increase from about $500 to $4,300 between 1940 and 1970. For whites, the North-South wage gap was $5,600 on average, and increased similarly during the four-decade period. The bottom panel shows the log wage results. On average between 1940 and 1970, the North-South wage gap was 42 log points higher for blacks than for whites, among whom the gap was 32 log points. These results are also similar within decades: between 1940 and 1970, the racial difference in North-South wage gaps was between about 30 and 40 log points, while the North-South wage gap for whites decreased from about 37 to 24 log points during this period. Regardless of how wages are measured, these difference-in-differences estimates show that the North-South wage differential was substantially higher for blacks than for whites, within and across decades.

[18] I classify states as Southern using the Census Bureau's definition of the South: Alabama, Arkansas, Delaware, Florida, Georgia, Kentucky, Louisiana, Maryland, Mississippi, North Carolina, Oklahoma, South Carolina, Tennessee, Texas, Virginia and West Virginia are Southern states. I define the North as any other state in the US. Although the Great Migration started well-before 1940, that decade is the first for which individual-level wage data are available.

[19] The wage measure consists of all income from wages and salary in the year before enumeration (this variable is named INCWAGE in the IPUMS dataset). The self-employed are included but business and farm income are not. All wages are inflated to 1999 dollars using the CPI weights supplied with the IPUMS. To make the wage and log wage regressions comparable, I restrict the sample to include only those reporting nonzero wages. In addition, to focus on men working-age men, I restrict the sample to those aged 16-64.

4.3 Applying the identification results

If, as Smith and Welch (1989) caution, the North-South wage differentials detailed in Table 2 represent a combination of the causal effect of migration on migrants' wages and selection bias arising because those with greater skill find migration less costly, more beneficial, or both, and thus are more likely to migrate—that is, if equilibrium wages and migration behavior are determined by a Roy model—the identification results developed in this paper may be applicable. Under the linear Roy model, black-white differences in North-South wage differentials can be interpreted as a lower-bound on the black-white difference in the average effect of migration on migrants' wages, provided that (i) the distributions governing the unobservable determinants of enrollment (that is, skill, a nonlinear transformation of skill, or a composite of idiosyncratic factors that correlate with counterfactual wages and skill) belong to one of the classes defined in Proposition 2, (ii) the treatment probability is less than one-half for both groups, and (iii) the slopes of the counterfactual outcome functions satisfy Assumption 2.

As discussed above, the distributional requirements for identification are met by those employed in standard economic and econometric models. Figure 3 and Table 1 show that blacks migrated with higher probability than whites and that the migration probability was less than one-half for both groups. Although wage data are available only after the beginning of the Great Migration, Proposition 1 implies that the data are consistent with Assumption 2 if the variance of wages among non-migrant blacks is no larger than among non-migrant whites. Table 3 summarizes tests of this condition. Across and within decades, the sample standard deviation of wages (measured in levels) is smaller for blacks than for whites and the null hypothesis that the standard deviations are equal can be rejected. For 1960 and 1970, the standard deviation of black log wages is statistically different from that for white log wages, though the ratios of the sample standard deviations are very close to one; for 1940 and 1950, and for the pooled period, the tests give no indication that blacks' log wages are more variable than whites'.

Even if outcomes and enrollment are not well-approximated by a linear Roy model, the framework of the nonlinear Roy model may still apply. In addition to relatively weak distributional and functional-form restrictions, identification under this model requires that untreated outcomes satisfy Assumption 6 (the nonlinear analog of the relative slope assumption) and covariance condition (12). In light of Proposition 3, the variance tests discussed above suggest that the data are consistent with the first of these requirements.

To implement the heuristic test proposed in Section 3 for whether the data support the covariance condition, I assign observations to education×age×birthplace×year cells, compute within-cell black-white differences in migration probabilities and mean wages among those working in the South, and estimate the covariance between cell-specific wages and differential migration rates. The justification for this proxy strategy is that these covariates may correlate with unobserved skill (and other unobserved determinants of counterfactual wages), but covariate-specific treatment effect estimates would be uninteresting and redundant if skill or earnings potential were directly observable. Table 4 presents the estimated covariances between black-white differences in migration probabilities and

the wages of black and white men working in the South, by decade. Regardless of whether wages are measured in logs or in levels, or for blacks or whites, the estimated covariances are either negative or very close to zero. To give a sense of the statistical significance of these estimates, the table also reports 95% nonparametric bootstrap confidence intervals. For only two of the sixteen estimates can the null hypothesis that the covariance exceeds zero be rejected, and even these estimates are close to zero (considering the scales of log and absolute wages). These results are consistent with the hypothesis that the covariance condition holds, so that the black-white differences in selection bias terms is bounded above by zero.

The data are therefore broadly consistent with the conditions required for differences in differences to identify a lower bound on the black-white difference in ATTs under any of the Roy models developed above. Consequently, the estimates presented in Table 2 imply that, on average between 1940 and 1970, whatever the proportional change that a white Southerner would have experienced as a consequence of migrating to the North, a black Southerner would have experienced an increase in wages that was at least 40% greater. Similarly, whatever the absolute effect of migrating on the white Southerner's wages was, the wages of his black counterpart would have increased by at least an additional $3,300. Furthermore, since it is unlikely that migration decreased the typical white migrant's wage, the difference-in-differences estimates in Table 2 can also be interpreted as lower bounds on the ATT itself for black migrants. The implied bounds agree well with the first-differenced estimate, reported by Collins and Wanamaker (2014), of 63 log points for black men who migrated between 1910 and 1930, suggesting that the difference in differences recovers a (comfortingly) conservative lower bound on the treatment effect.

5 Conclusion

A frequent concern in observational studies of treatment effects is that observed outcome differences between treated and untreated individuals are contaminated with selection bias arising because those who enroll in the treatment would have experienced better outcomes regardless of whether they were treated. Such concerns are usually motivated, if only implicitly, by a Roy model in which individuals enroll if they expect to benefit from the treatment. The results in this paper demonstrate that, in many such circumstances, group differences in treated-untreated mean outcome differences identify lower bounds on group differences in the average effect of the treatment on the treated. These identification results hold under distributional and functional form assumptions that are substantially more general than those maintained by sample selection and other econometric models that are routinely used in practice. The other conditions required for identification of a lower bound can be tested using data on enrollment and outcomes.

In many applications, group heterogeneity in treatment effects is of direct, if not primary, interest. In addition, under the hypothesis that the treatment is, at worst, ineffective, my results imply that differences in differences also identify a lower bound on the average effect of the treatment itself for treated members of the group that enrolls with higher probability. This hypothesis may

be justified by theoretical reasoning or previous empirical evidence—it is implied in equilibrium by most Roy models of enrollment. Although treatment effect bounds are less informative than point estimates, they may suffice to answer the research question at hand; they are preferable in any case to point estimates based on questionably exogenous sources of variation in treatment.

I apply these identification results to interpret black-white differences in North-South wage differentials in terms of the causal effect of migration on wages during the Great Migration, finding that Northward migration increased blacks' wages by at least 40%, or $3,300, more than whites' wages. Additionally, since the nature of the South during this period and the sheer sizes of white migrant flows and wage differentials suggest that migration did not decrease wages for whites, this finding also implies that migration increased blacks' wages by at least 40%. The estimates reported by Collins and Wanamaker (2014) confirm this result and imply that the bound is conservative.

A The concavity of truncated expectations

A comparison of Figures 1 and 2 shows that distributions with log concave densities tend to belong to the first class of distributions defined in Proposition 2 while those with log convex densities tend to belong the the second class. For example, the normal, logistic, and exponential densities are log concave, as are the gamma and Weibull densities when their shape parameters exceed one. Figure 1 shows that these distributions have convex (concave) right- (left-) truncated expectations. Similarly, the Pareto density is log convex, as are the gamma and Weibull when their shape parameters are less than one. Figure 2 shows that these distributions have concave left- and right-truncated expectations. The uniform and lognormal densities lie somewhere between these extremes; the uniform density is both log concave and log convex while the lognormal density switches from log concave to log convex.

Proposition 5 shows that this pattern is not coincidental; while the log concavity of the density is not sufficient for the convexity of the truncated expectation, these properties are closely related.

Proposition 5. *Suppose that a is distributed over $[L, H]$ with density f and*

$$\frac{d}{da}\left|\frac{[\log f(a)]''}{\left\{[\log f(a)]'\right\}^2}\right| \lesseqgtr 0 \quad \textit{when} \quad f'(a) \lesseqgtr 0. \tag{14}$$

Then:

1. *If f is log concave and $\lim_{a\to L} f = \lim_{a\to H} f = 0$, $E(a|a \geq \hat{a})$ is convex and $E(a|a < \hat{a})$ is concave.*

2. *If f is log convex and $\lim_{a\to H} f = 0$, $E(a|a \geq \hat{a})$ is concave.*

Noting its similarity to the measure of absolute risk aversion, what condition (14) requires is that the log of the density become less concave as the density itself decreases.[20] To illustrate the

[20]For an increasing utility function u, $-u''/u'$ will be decreasing if $-u''/(u')^2$ is (though the risk aversion measure

proposition, consider first the standard normal and logistic distributions, both of which have log concave densities and, as Figure 1 shows, convex left- and concave right-truncated expectations. The expression $|(\log f)''/(\log f)'^2|$ evaluates to $1/a^2$ for the normal density and $2\exp(a)/[1-\exp(a)]^2$ for the logistic density; both of these functions are decreasing on $a > 0$. For the log convex Pareto density with shape parameter β, this expression is $1/(\beta+1)$, which does not depend on a; the left-truncated expectation is linear.

The proof of Proposition 5 relies on an extension of the Prékopa-Borell theorem (Prékopa, 1971, 1973; Borell, 1975) due to Mares and Swinkels (2014).[21] Define the local ρ-concavity of $g(c)$ at c by

$$\rho_g(c) = 1 - \frac{g(c)g''(c)}{[g'(c)^2]}.$$

The justification for this definition is that if the local ρ-concavity of $g(c)$ at c is t, then g^t/t is linear at c. In showing that the local ρ-concavity of g can be used to bound the local ρ-concavity of the function $\bar{G}(c) = \int_c^1 g(t)dt$, Mares and Swinkels (2014, Lemma 3) provide the following lemma for an arbitrary, positive function g on the unit interval.

Lemma 1 (Mares and Swinkels, 2014)**.** *If $g(0) = 0$ and ρ_g is monotone on some interval $[0, \hat{c}]$, then $\rho_{\int_0^c g(s)ds}$ and $\rho_g(c)$ share the same monotonicity on $[0, \hat{c}]$. If $g(1) = 0$ and ρ_g is monotone on $[\hat{c}, 1]$, then $\rho_{\int_c^1 g(s)ds}$ and ρ_g share the same monotonicity on $[\hat{c}, 1]$.*

Proof of Proposition 5. For the log concave case, I prove the result for $E(a|a \geq \hat{a})$. The convexity of $-E(a|a < \hat{a})$ follows by analogy. First, note that, since $E(a|a \geq \hat{a}) - \hat{a} = [\int_{\hat{a}}^H 1 - F(t)dt]/[1 - F(\hat{a})]$ (this follows from integration by parts, see Bagnoli and Bergstrom, 2005), we can write

$$\frac{d}{d\hat{a}} E(a|a \geq \hat{a}) = \frac{f(\hat{a})}{1-F(\hat{a})} \frac{\int_{\hat{a}}^H 1 - F(t)dt}{1-F(\hat{a})}.$$

Since

$$\rho_{\int 1-F}(\hat{a}) = 1 - \frac{f(\hat{a}) \int_{\hat{a}}^H 1 - F(t)dt}{[1-F(\hat{a})]^2},$$

$E(a|a \geq \hat{a})$ convex is equivalent to $\rho'_{\int_{\hat{a}}^H 1-F}(\hat{a}) \leq 0$. By Lemma 1, $\rho'_{1-F}(\hat{a}) \leq 0$ implies $\rho'_{\int_{\hat{a}}^H 1-F}(\hat{a}) \leq 0$. Because log concave densities are unimodal (see An, 1995), $\rho_{1-F}(\hat{a})' \leq 0$ whenever $\hat{a}$ is less than or equal to the mode of a, since

$$\rho_{1-F}(\hat{a}) = 1 - \frac{[-f'(\hat{a})][1-F(\hat{a})]}{[f(\hat{a})]^2} = 1 + \frac{f'(\hat{a})[1-F(\hat{a})]}{[f(\hat{a})]^2}$$

and, when $f' > 0$, f'/f and $(1-F)/f$ are positive and, by log concavity, they are always decreasing.

has to be renormalized when applied to log densities, which are not monotone increasing). The proof relies on the concept, due to Mares and Swinkels (2014), of local ρ-concavity, which those authors show is closely related to risk aversion. Note that there is no condition for the right-truncated expectation in the log convex case because the curvature of this expectation is determined by the log concavity of the distribution function, and many log convex densities have log concave distribution functions.

[21] See Caplin and Nalebuff (1991) for an introduction to ρ-concavity and the Prékopa-Borell theorem.

When $\hat{a}$ exceeds the mode, so that $f' < 0$, we can apply Lemma 1 once again in order to infer the sign of $\rho'_{1-F}(\hat{a})$ from that of $\rho'_f(\hat{a})$. Noting that, since f is log concave, it can be written $f(\hat{a}) = \exp[h(\hat{a})]$ where h is a concave function,

$$\rho_f(\hat{a}) = 1 - \frac{f''(\hat{a})f(\hat{a})}{[f'(\hat{a})]^2} = 1 - \frac{\exp[h(\hat{a})]\left\{\exp[h(\hat{a})]h'(\hat{a})^2 + \exp[h(\hat{a})]h''(\hat{a})\right\}}{\{\exp[h(\hat{a})]h'(\hat{a})\}^2} = -\frac{h''(\hat{a})}{[h'(\hat{a})]^2}.$$

Since log concavity implies $h'' < 0$, under the conditions of the proposition, $-h''/(h')^2$ is positive and (weakly) decreasing, so $\rho_f(\hat{a})' \leq 0$, implying that $\rho_{1-F}(\hat{a})' \leq 0$ and hence $\rho_{\int_{\hat{a}}^H 1-F}(\hat{a})' \leq 0$, establishing the result.

For the log convex case, note that if f is log convex then f'/f is increasing and, since $f(H) = 0$ implies that $1 - F$ is also log convex (see Theorem 2 of Bagnoli and Bergstrom, 2005), $(1 - F)/f$ is increasing as well. Thus, $\rho_{1-F}(\hat{a})$, and consequently $\rho_{\int_{\hat{a}}^H 1-F}(\hat{a})$ are positive and increasing when $f' > 0$. When $f' < 0$, by the conditions of the proposition, we have $h''/(h')^2$ positive and decreasing, so that $\rho_f(\hat{a})$, $\rho_{1-F}(\hat{a})$ and hence $\rho_{\int_{\hat{a}}^H 1-F}(\hat{a})$ are increasing, implying that $E(a|a \geq \hat{a})$ is concave. □

B Proofs

B.1 Section 2

Proof of Proposition 1. The first part of the proposition follows directly from (1). To prove the second part, note that by Corollary 5 of Bagnoli and Bergstrom (2005), log concavity and convexity are preserved by linear transformations, so if f_l is log concave then so is f_h. Further, by Proposition 1 of Heckman and Honoré (1990), if a has a log concave (convex) density then $Var(a|a \leq \hat{a})$ is increasing (decreasing) in $\hat{a}$. Since

$$\frac{Var(y|h,0)}{Var(y|l,0)} = \frac{(\gamma_h\sigma_h)^2}{\gamma_l^2}\frac{Var[a|l, a < (\hat{a}_h - \mu_h)/\sigma]}{Var(a|l, a < \hat{a}_l)}$$

and $(\hat{a}_h - \mu_h)/\sigma_h < \hat{a}_l$, the result follows. □

Proof of Proposition 2. To prove the first part, note first that since $E(a|a \geq \hat{a}) - E(a| < \hat{a})$ is convex by assumption, if this difference is increasing at L, it is increasing on the entire support. Otherwise, suppose that $\lim_{\hat{a}\to\infty} dE(a|a < \hat{a})/d\hat{a} > 0$. Then, since a has infinite support, there exists an $\hat{a}$ such that $E(a|a \leq \hat{a}) > E(a)$, a contradiction. Thus $\lim_{\hat{a}\to\infty} dE(a|a < \hat{a})/d\hat{a} = 0$, and since $dE(a|a \geq \hat{a})/d\hat{a} \geq 0$, there is a unique a^* at which $d[E(a|a \geq \hat{a}) - E(a|a < \hat{a})]/d\hat{a} = 0$ and the difference in truncated means is minimized.

Next, write

$$\frac{d}{d\hat{a}}[E(a|a \geq \hat{a}) - E(a|a < \hat{a})] = \frac{f(\hat{a})}{1 - F(\hat{a})}\left(\frac{\int_{\hat{a}}^{\infty} tf(t)dt}{1 - F(\hat{a})} - \hat{a}\right) - \frac{f(\hat{a})}{F(\hat{a})}\left(\hat{a} - \frac{\int_L^{\hat{a}} f(t)dt}{F(\hat{a})}\right).$$

At the median, $\tilde{a}$, of a, this expression becomes $4f(\tilde{a})\left[E(a)-\tilde{a}\right]$. Thus, for f symmetric, $a^*=E(a)=\tilde{a}$. Instead, if $E(a)>\tilde{a}$, $d[E(a|a\geq\hat{a})-E(a|a<\hat{a})]/d\hat{a}>0$ at $\tilde{a}$, so $a^*\leq\tilde{a}$.

To prove the second part, note that $f(a)$ log convex with $\lim_{a\to\infty}f=0$ implies that $1-F$ is log convex and that $f'\leq 0$ for all a implies that F is log concave (Bagnoli and Bergstrom, 2005). But $1-F$ log convex implies $dE(a|a\geq\hat{a})/d\hat{a}\geq 1$ while F log concave implies $dE(a|a<\hat{a})/d\hat{a}\leq 1$ (see, e.g., Heckman and Honoré, 1990). Thus $d[E(a|a\geq\hat{a})-E(a|a<\hat{a})]/d\hat{a}\geq 0$ for all $\hat{a}$. □

B.2 Section 3

Proof of Proposition 3. 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,

$$Var(a|l,\tilde{\gamma}(\Delta,\sigma a)<\epsilon)=E\left[Var\left(a|l,\epsilon,a<\frac{\tilde{\gamma}^{-1}(\Delta,\epsilon)}{\sigma}\right)\right]+Var\left[E\left(a|l,\epsilon,a<\frac{\tilde{\gamma}^{-1}(\Delta,\epsilon)}{\sigma}\right)\right],\tag{15}$$

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 (15) is approximately

$$\frac{\partial}{\partial(\tilde{\gamma}^{-1}/\sigma)}Var\left(a|l,a<\frac{\tilde{\gamma}^{-1}(\Delta,\mu_\epsilon)}{\sigma}\right)\left(\frac{\partial\tilde{\gamma}^{-1}(\Delta,\mu_\epsilon)/\sigma}{\partial\Delta}d\Delta-\frac{\tilde{\gamma}^{-1}(\Delta,\mu_\epsilon)}{\sigma^2}d\sigma\right).\tag{16}$$

By Proposition 1 of Heckman and Honoré (1990), π log concave implies that the leading term in (16) 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 (15) is approximately negative if π is log concave and positive if π is log convex.

The second term in (15) can be expressed

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

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

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

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 (15) is approximately zero.

Thus, since

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

the result follows. □

The proof of Proposition 4 makes use of the following lemmas.

Lemma 2. *Suppose that* $y'_{0h}(\sigma_h a)\sigma_h\leq y'_{0l}(a)$ *and define* $p(a|1;\theta)=p\left(a|l,\epsilon<\tilde{\gamma}(\Delta,\sigma a)\right)$. *Then*

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

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

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

between θ_l *and* θ_h.

Proof. If

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

then the conclusion follows.[22] 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.

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

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

□

Lemma 3. *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*

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

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. 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

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

[22] To be explicit, if $\rho_g = p(a|1;\theta_g) - p(a|0;\theta_g)$ then we have $\int \rho_h y_{0h} - \int \rho_l y_{0l} \leq \int(\rho_h - \rho_l)y_{0l} = (\partial/\partial\theta')\left(\int \rho_l y_{0l}\right)|_{\theta=\theta^*}d\theta \leq 0$ or $\int \rho_h y_{0h} - \int \rho_l y_{0l} \leq \int(\rho_h - \rho_l)y_{0h} = -(\partial/\partial\theta')\left(\int \rho_h y_{0h}\right)|_{\theta=\theta^*}(-d\theta) \leq 0$ 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.

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 negative on $a < a^*$ and positive 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 negative on $a < a^*$ and positive 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

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

or

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

To complete the proof, note that

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

□

Lemma 4. *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 3) has the same sign as

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

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 (17) 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 (17) 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

$$\frac{E\left[f\left(\tilde{\gamma}(\theta,a)\right)\frac{\partial\tilde{\gamma}(\theta,a)}{\partial\theta'}d\theta\right]}{E\left\{\left[1-F\left(\tilde{\gamma}(\theta,a)\right)\right]\right\}} - \frac{f\left(\tilde{\gamma}(\theta,a)\right)}{1-F\left(\tilde{\gamma}(\theta,a)\right)}\frac{\partial\tilde{\gamma}(\theta,a)}{\partial\theta'}d\theta. \tag{18}$$

The first term in (18) is a positive constant. Since f log concave implies that $f/(1-F)$ is monotone increasing, the second term in (18) is monotone increasing whenever $[\partial\tilde{\gamma}(\theta,a)/\partial\theta']d\theta > 0$. Therefore, there is an a^* such that (18) is positive on $a < a^*$ and negative on $a > a^*$.[23] □

Proof of Proposition 4. If $\tilde{\gamma}(\Delta,\sigma a) = \Delta + \gamma(\sigma a)$ then

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

The change in this expression with respect to a is $\left(\gamma''\sigma a + \gamma'\right)d\sigma \gtreqless 0$ 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

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

The change in this expression is $\gamma'\cdot(\sigma d\Delta+\Delta d\sigma)+\Delta\gamma''\sigma ad\sigma \gtreqless 0$ 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 2, 3, and 4.

[23] Note that (18) positive for all a implies $[\partial p(a|0)/\partial\theta']d\theta$ is always positive, in contradiction to $\int c_0 = 0$.

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 (17) 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 4, $[\partial p(a|0)/\partial \theta']d\theta$ has the same sign as

$$\frac{E\left[f\left(\gamma(\Delta + \sigma a)\right) \frac{\partial \gamma(\Delta + \sigma a)}{\partial \theta'} d\theta\right]}{E\left\{\left[1 - F\left(\gamma(\Delta + \sigma a)\right)\right]\right\}} - \frac{f\left(\gamma(\Delta + \sigma a)\right)}{1 - F\left(\gamma(\Delta + \sigma a)\right)} \gamma'(\Delta + \sigma a)(d\Delta + a d\sigma). \tag{19}$$

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 linear 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 (19) will be monotone decreasing as soon as $a > -d\Delta/d\sigma$, so there must be an a^* such that (19) is positive on $a < a^*$ and negative on $a > a^*$. The conclusion then follows from Lemmas 2 and 3. This case does not use the assumption that γ is CRRA. □

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.

Amemiya, T. (1984). Tobit Models: A Survey. *Journal of Econometrics 24*(81), 3–61.

An, M. Y. (1995). Log-concave Probability Distributions: Theory and Statistical Testing. Working Paper 919.

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

Borell, C. (1975). Convex Set Functions in d-Space. *Periodica Mathematica Hungarica 6*, 111–136.

Borjas, G. J. (1988). Self-Selection and the Earnings of Immigrants. *American Economic Review 77*(4), 531–553.

Caplin, A. and B. Nalebuff (1991). Aggregation and Social Choice: A Mean Voter Theorem. *Econometrica 59*(1), 1–23.

Chamberlain, G. (1986). Asymptotic Efficiency in Semi-parametric Models with Censoring. *Journal of Econometrics 32*(2), 189–218.

Collins, W. J. and M. H. Wanamaker (2014). Selection and Economic Gains in the Great Migration of African Americans: New Evidence from Linked Census Data. *American Economic Journal: Applied Economics 6*(1), 220–252.

Dahl, G. (2002). Mobility and the Return to Education: Testing a Roy Model with Multiple Markets. *Econometrica 70*(6), 2367–2420.

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

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

D'Haultfoeuille, X., A. Maurel, and Y. Zhang (2014). Extremal Quantile Regressions for Selection Models and the Black-White Wage Gap. NBER working paper no. 20257.

Donohue, J. and J. Heckman (1991). Continuous Versus Episodic Change: The Impact of Civil Rights Policy on the Economic Status of Blacks. *Journal of Economic Literature 29*(4), 1603–1643.

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.

Head, K. (2011). Skewed and Extreme: Useful Distributions for Economic Heterogeneity. Working paper.

Heckman, J. J. (1976). The Common Structure of Statistical Models of Truncation, Sample Selection and Limited Dependent Variables and a Simple Estimator for Such Models. *Annals of Economic and Social Measurement 5*(4), 475–492.

Heckman, J. J. (1979). Sample Selection Bias as a Specification Error. *Econometrica 47*(1), 153–161.

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

Heckman, J. J., S. Urzua, and E. Vytlacil (2006). Understanding Instrumental Variables in Models with Essential Heterogeneity. *Review of Economics and Statistics 88*(3), 389–432.

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.

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.

Mares, V. and J. M. Swinkels (2014). On the Analysis of Asymmetric First Price Auctions. *Journal of Economic Theory 152*, 1–40.

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.

Olsen, R. J. (1980). A Least Squares Correction for Selectivity Bias. *Econometrica 48*(7), 1815–1820.

Prékopa, A. (1971). Logarithmic Concave Measures with Application to Stochastic Programming. *Acta Sci. Math.(Szeged) 32*, 301–316.

Prékopa, A. (1973). On Logarithmic Concave Measures and Functions. *Acta Sci. Math.(Szeged) 34*, 335–343.

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*.

Smith, J. and F. Welch (1989). Black Economic Progress after Myrdal. *Journal of Economic Literature 27*(2), 519–564.

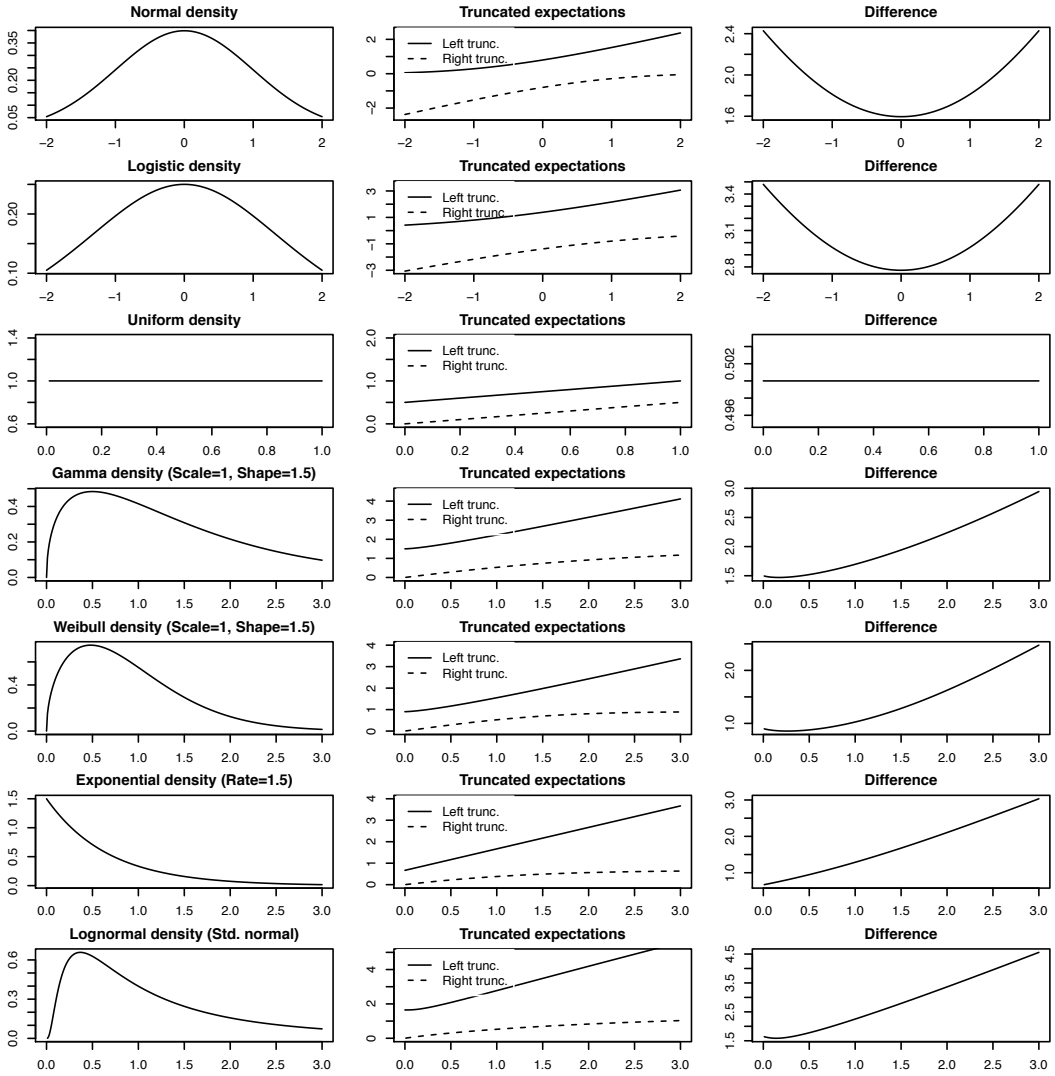
Tobin, J. (1958). Estimation of Relationships for Limited Dependent Variables. *Econometrica 26*(1), 24–36.

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

Wooldridge, J. M. (2002). *Econometric Analysis of Cross Section and Panel Data* (1st ed.). MIT Press.

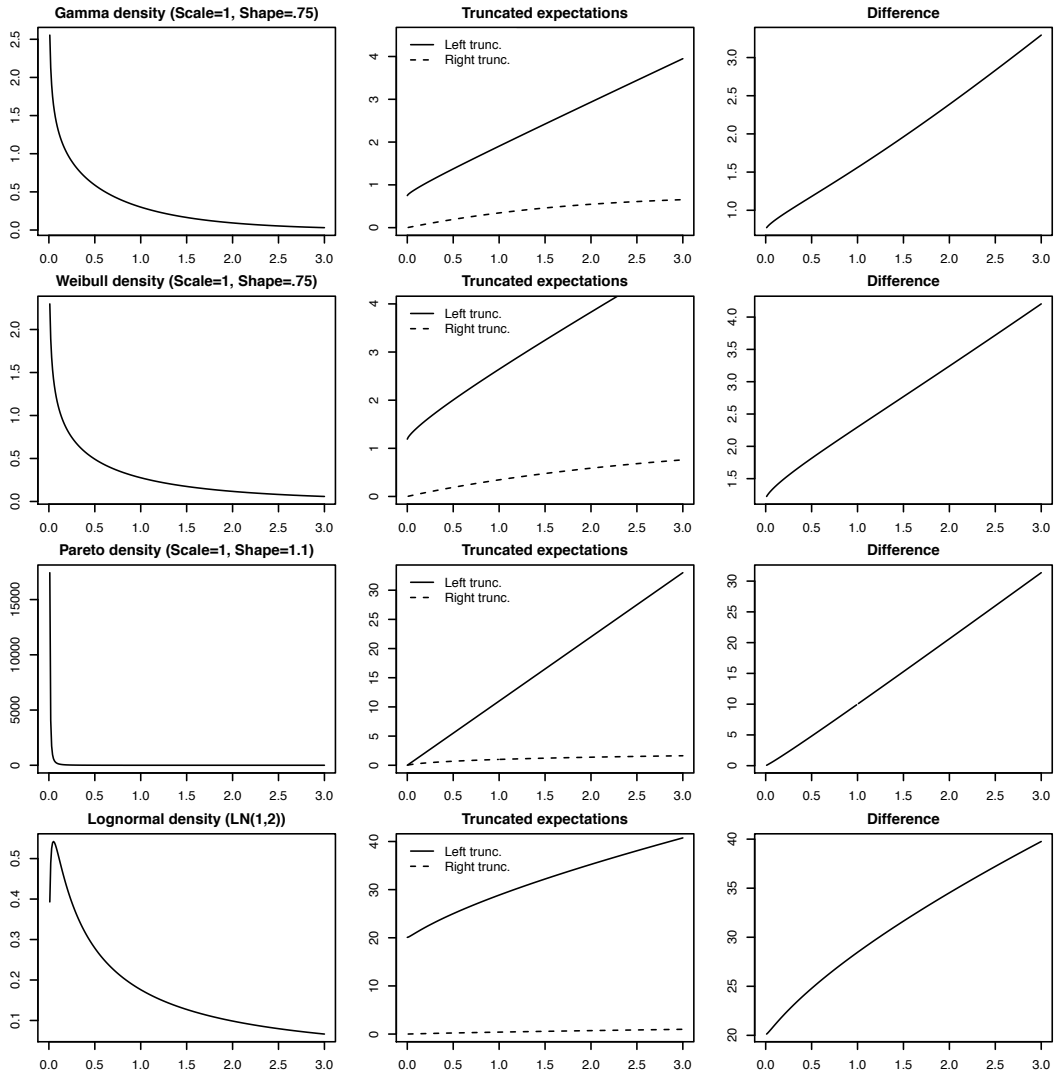
Figures and tables

Figure 1: Distributions with convex (concave) left- (right-) truncated expectations



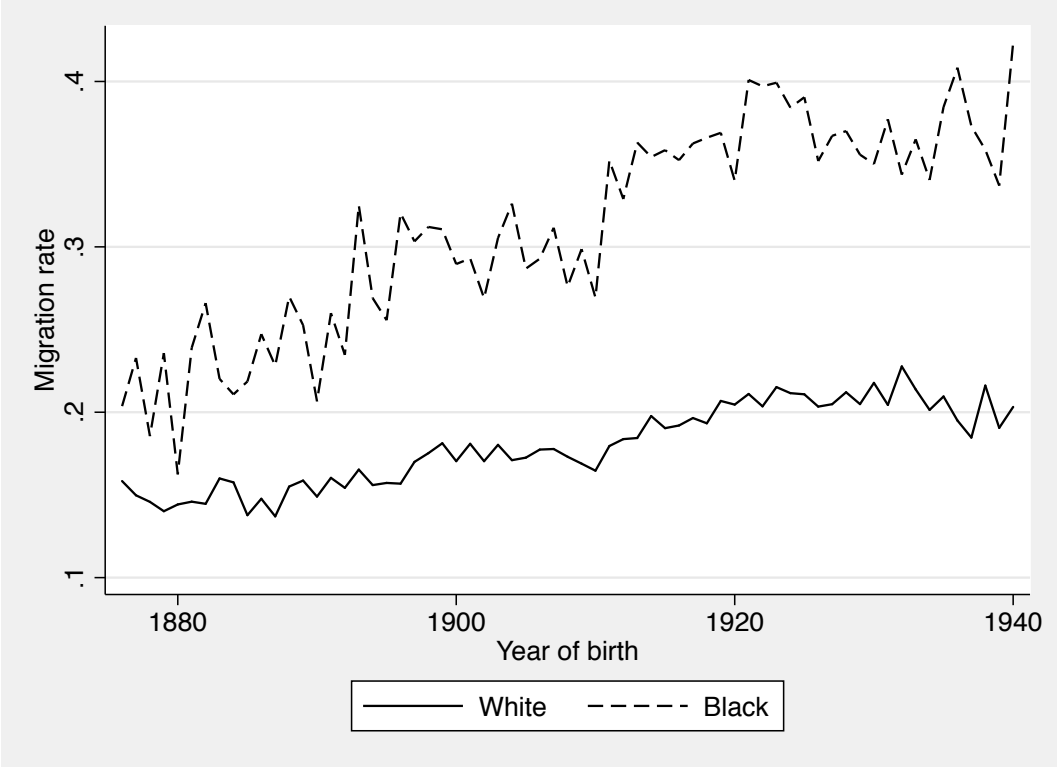
Notes—Difference denotes $E(a|a \geq \hat{a}) - E(a|a < \hat{a})$. Density formulae taken from Bagnoli and Bergstrom (2005). Expressions for the left- and right-truncated moments of the normal, logistic, gamma, Weibull and lognormal densities can be found in Arabmazar and Schmidt (1982), Heckman and Honore (1990) and Jawitz (2004). By direct calculation, if $a \sim U[0,1]$ then $E(a|a \geq \hat{a}) = 1/2 + \hat{a}/2$ and $E(a|a < \hat{a}) = \hat{a}/2$. If a is exponential with rate parameter λ, then it can be shown (integrate by parts and apply L'Hôpital's rule) that $E(a|a \geq \hat{a}) = 1/\lambda + \hat{a}$ and $E(a|a < \hat{a}) = 1/\lambda - \hat{a}/(e^{\lambda \hat{a}} - 1)$ (see also Head, 2011).

Figure 2: Distributions with concave truncated expectations



Notes—Difference denotes $E(a|a \geq \hat{a}) - E(a|a < \hat{a})$. Density formulae taken from Bagnoli and Bergstrom (2005). Expressions for the left- and right-truncated moments of the gamma, Weibull and lognormal densities can be found in Jawitz (2004). If a is Pareto distributed with shape parameter β then it can be shown that $E(a|a \geq \hat{a}) = \beta\hat{a}/(\beta - 1)$ and $E(a|a < \hat{a}) = [\beta/(\beta - 1)](1 - \hat{a}^{1-\beta})/(1 - \hat{a}^{-\beta})$ (see also Head, 2011).

Figure 3: Migration rates by year of birth



Notes—Probability of living in the North, by birth year, for black and white men, aged 30 or later and born after 1850.

Table 1: Migration rates

	All decades	1940	1950	1960	1970
Black	0.108***	0.0569***	0.120***	0.130***	0.126***
	(0.00685)	(0.00749)	(0.00993)	(0.00872)	(0.00810)
Constant	0.117***	0.130***	0.166***	0.193***	0.187***
	(0.00733)	(0.00736)	(0.00944)	(0.0105)	(0.0100)
Observations	517,361	133,059	45,130	162,833	176,339

Notes—Pooled models include decade effects. Sample consists of Southern-born men greater aged 16-64 with nonzero wages. Standard errors clustered on state-year of birth. *** $p<0.01$, ** $p<0.05$, * $p<0.1$.

Table 2: Difference-in-difference regressions

	Wage				
	All decades	1940	1950	1960	1970
Black	-11,037***	-6,080***	-7,152***	-11,034***	-12,678***
	(168.9)	(137.0)	(199.5)	(246.5)	(376.6)
North	5,825***	3,572***	3,554***	4,998***	6,433***
	(336.2)	(279.2)	(372.9)	(534.2)	(794.0)
Black*North	3,338***	466.6*	2,615***	3,139***	4,340***
	(278.7)	(250.2)	(366.2)	(403.8)	(561.5)
Observations	384,818	83,198	32,479	125,706	143,435
	Log wage				
	All decades	1940	1950	1960	1970
Black	-0.687***	-0.684***	-0.587***	-0.667***	-0.532***
	(0.0116)	(0.0153)	(0.0219)	(0.0183)	(0.0177)
North	0.321***	0.368***	0.306***	0.282***	0.236***
	(0.0190)	(0.0274)	(0.0327)	(0.0368)	(0.0421)
Black*North	0.422***	0.325***	0.400***	0.388***	0.338***
	(0.0162)	(0.0243)	(0.0335)	(0.0257)	(0.0249)
Observations	384,818	83,198	32,479	125,706	143,435

Notes—Pooled models include decade effects. Sample consists of Southern-born men greater aged 16-64 with nonzero wages. 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 3: Variance tests

	Wages				
	All decades	1940	1950	1960	1970
White SD	19964.55	10037.33	11463.72	17116.83	23719.99
Black SD	10801.45	4299.75	6418.75	8920.82	13230.61
p-value	0.00	0.00	0.00	0.00	0.00
	Log wages				
	All decades	1940	1950	1960	1970
White SD	1.13	1.07	1.06	1.05	1.04
Black SD	1.13	0.91	1.04	1.07	1.05
p-value	0.99	0.00	0.07	0.00	0.04

Notes—P-values are for the null that $\sigma_W/\sigma_B = 1$ under the alternative that $\sigma_W/\sigma_B \neq 1$.

Table 4: Covariance tests

	Black wages		White wages	
Decade	Covariance	Conf. interval	Covariance	Conf. interval
1940	2.844	(-16.939, 44.158)	-81.554	(-95.751, 2.633)
1950	12.496	(-64.012, 86.344)	-12.317	(-134.710, 81.069)
1960	-109.875	(-138.227, -18.251)	-106.838	(-176.898, 6.311)
1970	-145.987	(-252.687, -76.388)	-122.319	(-265.139, -22.607)

	Black log wages		White log wages	
Decade	Covariance	Conf. interval	Covariance	Conf. interval
1940	0.007	(0.001, 0.012)	0.002	(-0.002, 0.009)
1950	0.004	(-0.008, 0.013)	0.005	(-0.007, 0.014)
1960	0.000	(-0.004, 0.009)	0.006	(0.001, 0.011)
1970	0.001	(-0.008, 0.003)	0.005	(-0.001, 0.008)

Notes—Table entries represent the covariance 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 95% percentile-based confidence intervals from a nonparametric bootstrap in which cell-specific wages and group differences in migration probabilities, and the covariance between these quantities, are estimated in each of 999 bootstrap samples.