Roy-Model Bounds on Differential Treatment Effects

John Gardner*

April, 2016

Abstract

Concerns about positive selection bias in observational treatment effect studies are often motivated by a Roy model of enrollment and counterfactual outcomes. 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 this effect is nonnegative, differences in differences also identifies a lower bound on the ATT itself for the high-treatment-rate group. The conditions required for identification are either relatively weak or falsifiable. I apply the results to interpret black-white differences in North-South wage differences in terms of the causal effects of the Great Migration.

JEL Codes: C50, C34, J71, R23.

*Department of Economics, University of Mississippi, jrgardne@olemiss.edu. I thank John Conlon, Lowell Taylor, Robert Miller, and seminar participants at Carnegie Mellon University, the Western Economics Association conference, and the University of Mississippi for helpful comments.

1 Introduction

Consider the following scenario, all-too-familiar in conferences and seminars: A presenter contends that, because those who enrolled in some program, or engaged in some behavior, or otherwise received some treatment, experienced better outcomes than those who did not, the treatment has a positive causal effect on outcomes. An audience member raises a hand and asks the inevitable: "Could selection bias explain your findings? What if the individuals that volunteer for the treatment would have had better outcomes anyway, and there is no causal effect?" If only implicitly, what motivates the interlocutor's question—and many like it—is a hypothesis that outcomes and enrollment into the treatment follow a Roy model. That is, individuals enroll in the treatment if they expect to benefit from it, and this benefit depends on the same factors that determine the counterfactual outcomes individuals would experience with and without the treatment. In this paper, I show that when two different groups selectively enroll in a treatment according to such a Roy model, but do so at different rates, information about group differences in treatment rates can often be used to recover some of the information about the causal effect of the treatment contained in outcome comparisons between treated and untreated individuals. In particular, I show that a simple difference in differences—that is, the difference between the high- and low-treatment-rate groups in treated-untreated mean outcome differences—often identifies a lower bound on the difference in the average causal effect of the treatment on the treated between the high- and low-treatment-rate groups. If the treatment is, at worst, ineffective, so that the average effect of the treatment on the treated is nonnegative for both groups, this group difference is itself a lower bound on the average effect of the treatment for treated members of the high-rate-group. Thus, the identification arguments that I develop below are of particular use when interest centers on group heterogeneity in treatment effects or when theory or prior empirical evidence suggest that the treatment in question is, at worst, ineffective. Another interpretation of this paper's results is that they establish what differences in differences identifies when it is applied across groups rather than over time, and when it does so.

The basic idea behind the identification result can be motivated in multiple ways. In general, when enrollment is voluntary, the difference in mean outcomes between those who enroll and those who do not reflects a combination of the average effect of the treatment on the treated (i.e., the ATT) and selection bias that arises because enrollment is nonrandom. When enrollment follows a Roy model of positive selection like the one described above, this selection bias tends to lead outcome comparisons between the treated and the untreated to overstate the ATT, since enrollees would have experienced better outcomes even absent the treatment. Now suppose that some members of two different groups have

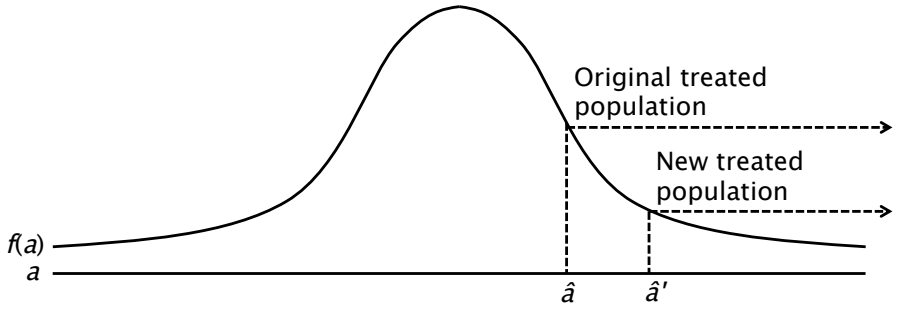
the opportunity to enroll in the treatment, but the enrollment is more selective for one group in the sense that members of this group are less likely to enroll. Intuition suggests that the selection bias component of the treated-untreated mean outcome difference will be larger for this more selective group. In this case, subtracting the treated-untreated difference for the low-treatment-rate-group from that difference for the high-rate group (i.e., *differences in differences)* will over-control for selection bias among the high-rate group, bounding the group difference in ATTs from below. My identification results amount to elucidating the conditions under which this intuition is correct.

A more analytical motivation is available when, as in many models of self selection, individuals enroll in a treatment if their realization of some random variable exceeds a threshold, and counterfactual outcomes depend on this same variable. For example, the net benefit of participating in a job training program may increase with latent skill, so that only those sufficiently skilled that their benefit is positive enroll. In such a model, the selection bias component of the treated-untreated mean outcome difference will depend on the difference in the mean of this random variable between the treated and the untreated (i.e., the difference between the left- and right-truncated expectations). An increase in the enrollment threshold will discourage the would-be enrollees with the lowest realizations of the random variable from seeking the treatment, making them the highest-realization non-enrollees, and increasing the mean of this variable among both the treated and untreated populations. If, as illustrated in Figure 1, the treated population is relatively small and the density of the random variable is decreasing at the threshold, we might expect a larger increase in the treated mean than in the untreated one, since the effect of reassigning these individuals out of enrollment will be spread over a smaller treated population. In this case, increasing the enrollment threshold will also increase the selection bias component of the treated-untreated mean outcome difference. Extending this intuition, if there are two groups who enroll at different rates because one group has a higher threshold for enrollment, the selection bias component will be larger for the low-treatment-rate (i.e., high-enrollment-threshold) group. Differences in differences will therefore bound the group difference in ATTs from below.

Threshold models similar to the one described above are common in theoretical and econometric analyses of self-selection. In this paper's formal identification results, I show that the identification argument motivated above can be applied in a large class of Roy models. I also develop similar results that can be applied in cases when idiosyncratic factors, besides those that determine counterfactual outcomes, also influence the decision to enroll in the treatment.

My results are perhaps best understood through their limitations. This paper proceeds

Figure 1: The effect of an increase in the enrollment threshold



by developing a series of decreasingly-restrictive Roy models and establishing the conditions under which identification results similar to the one sketched above hold. The results all place similar types of restrictions on the environments that agents face and how they behave within those environments. At the highest level, each requires that enrollment behavior follow some underlying Roy model—that is, individuals enroll if they expect to benefit from the treatment. This is a strong assumption, and one that imposes an explicit decision-making framework on the data. However, the purpose of my identification results is to interpret sample statistics in causal terms when selection bias is a concern. Since the hypothesis that enrollment follows a Roy model almost always underlies such concerns, if the behavioral assumptions of the Roy model are inappropriate, selection bias is unlikely to be a problem.[1]

Each of the results that I develop holds for a specific model differentiated by the assumptions that it makes about the functional forms that counterfactual outcomes and the enrollment decision rule take and the distributions that govern the unobserved arguments to these functions. I begin by developing the result in a setting similar to Roy's (1951) original model, in which outcomes are a linear function of a normally-distributed random variable. I go on to show that the lower-bound approach to identification means that analogous results apply under functional form and distributional assumptions that are much weaker than those used in standard econometric models of sample selection, truncation, discrete choice, duration, and reliability, including those implemented in popular statistical packages. Though my identification procedure is more robust to functional form and distributional misspecification than these econometric models, it shares their overt reliance on them: if the data-generating process cannot be well-approximated by a model that meets the conditions required for identification, there is no guarantee that my procedure will bound the population object of interest (either a difference in ATTs, or the ATT itself for the high-rate group) from below.

At the lowest level of abstraction, my results require that the parameters of a given Roy

[1]It is also worth noting that, if there is reason to suspect negative selection into the treatment, treated-untreated mean outcome differences themselves are likely to bound the treatment effect from below.

model satisfy certain restrictions. In some cases, these restrictions can be tested directly, or at least approximately, using the sample analogs of appropriate population moments (most of my results, e.g., require that both groups are treated with probability less than one half, which is easily tested). In others, direct tests are unavailable because the sample analogs of the implied moments are not observable. Although I develop heuristic falsification tests for appreciable deviations from these conditions, the possibility remains that these parametric restrictions are not met by data drawn from a process that is apparently consistent with them. Unsurprisingly, there is also a tradeoff between the severity of the functional form assumptions imposed by the model and the testability of the parametric restrictions required under it for identification.

Arguably the most binding drawback of my results is their object of identification. 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. This makes my results 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. For example, in the empirical portion of this paper I apply my identification results to interpret black-white differences in North-South wage differentials in terms of black-white 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 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 (e.g., whether a treatment improves outcomes, either relative to another group or absolutely). 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.

My results draw on a number of seminal methodological and theoretical papers on sample-selection problems, and can be applied in situations similar to those studied in these papers. Tobin (1958) demonstrates the effects of truncation on estimates of linear models and develops an estimation procedure to circumvent the resulting bias. Heckman (1979) shows that similar biases arising due to sample-selection can be viewed as specification error and accounted for using his celebrated sample-selection correction. Heckman (1976) and Amemiya (1984) show that many sample-selection estimators can be viewed as extensions of Tobin's (1958) truncated regression procedure. 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 the Roy model to analyze selective migration, while Heckman and Honoré (1990) study the identifiability of the model, broadly construed, from population outcome densities. Heckman and Vytlacil (1999), Heckman, Urzua, and Vytlacil (2006), and Eisenhauer, Heckman, and Vytlacil (2015) use a Roy framework to estimate and analyze treatment effects.

In Roy's (1951) original work, individuals have some innate earnings potential in each of two occupations, the between-occupation difference in earnings potential is a normally distributed random variable, and individuals self-select into the occupation that offers the highest potential earnings. I begin Section 2 by developing the basic identification argument in this parsimonious setting, which is readily adapted to the study of general treatment effects. I then show that similar results apply under less restrictive distributional assumptions and enrollment decision models, so that the classical Roy model can be viewed as an approximation to more complex treatment effect settings. In Section 3, I discuss identification when counterfactual outcomes and the utility from enrolling are nonlinear functions of some unobserved random variable. In Section 4, I apply the identification results to the analysis of black-white differences in the casual effect of Northward migration on migrants' wages

during the Great Migration. I summarize and conclude in Section 5.

2 Identification in classical Roy models

Suppose that individuals self-select into one of two treatment states, $d \in \{0, 1\}$, where d is equal to 1 if the individual elects to receive the treatment and 0 otherwise. In addition to treatment status, individuals differ according to observable and exogenous membership in one of two groups, $g \in \{l, h\}$, where members of group l are, by definition, less likely to receive the treatment than members of group h (e.g., if women are more likely to enroll in the treatment in question, they belong to group h while men belong to group l).

Counterfactual outcomes (Rubin, 1974)—the hypothetical realizations of the outcome of interest individuals would experience in either treatment state—are a linear function of some unobserved, normally-distributed random variable a. In particular, the outcome that a member of group g would experience given a realization of a and a treatment choice d, is

$$y_{dg}(a) = \gamma_{dg} a, \tag{1}$$

where $\gamma_{dg} \geq 0$ for $(d, g) \in \{0, 1\} \times \{l, h\}$. For example, y_{1g} and y_{0g} might represent an unemployed individual's earnings with and without enrolling in a job training program, both of which might depend on the individual's latent skill a.[2]

In the original Roy economy, individuals differ only in their earnings potential, so there is no analog to group membership g. To accommodate between-group differences, I assume initially that the a are iid normal, both within and across groups. Later, I relax both the assumption that a is normally distributed and the assumption that its distribution is independent of group membership g.

To incorporate Roy's positive selection hypothesis into this more general treatment-effect setting, assume that the decision to enroll in the treatment takes the form

$$d(a, g) = \begin{cases} 1 & \text{if } a \geq \hat{a}_g \\ 0 & \text{otherwise} \end{cases}. \tag{2}$$

Since the groups $g \in \{l, h\}$ are defined by the probability that they enroll in the treatment and the distribution of a is the same for both groups, $\hat{a}_h < \hat{a}_l$. This decision rule encompasses

[2]Note that (1) can be interpreted as the expectation of y given a in order to accommodate the case where outcomes are random given a (e.g., if y is measured with error or includes an additively separable prediction error). The omission of a constant term is for expositional simplicity; the following argument applies without modification if counterfactual outcomes are given by $y_{dg}(a) = \beta_{dg} + \gamma_{dg} a$.

the standard Roy model assumption that the enrollment decision depends only on outcomes (net of costs), while allowing for the possibility that non-outcome factors that also depend on a influence enrollment (e.g., if high-latent-skill individuals who would earn more with or without a job-training program find training more enjoyable or less costly, they may be more likely to enroll in such a program). Like the standard model, however, it assumes that selection into treatment status depends purely on the same unobserved factors that determine counterfactual outcomes (another assumption that I relax below).

Let $y = dy_1 + (1-d)y_0$ denote observed outcomes (Quandt, 1958). For either group, the difference in mean outcomes between the treated and the untreated can be decomposed into the average effect of the treatment on the treated (ATT) and a bias term that reflects positive selection into the treatment:

$$\begin{aligned} E(y|d=1,g) - E(y|d=0,g) &= \gamma_{1g}E(a|a \geq \hat{a}_g) - \gamma_{0g}E(a|a < \hat{a}_g) \\ &= \underbrace{(\gamma_{1g} - \gamma_{0g})E(a|a \geq \hat{a}_g)}_{\text{ATT}} + \underbrace{\gamma_{0g}[E(a|a \geq \hat{a}_g) - E(a|a < \hat{a}_g)]}_{\text{Bias}}. \end{aligned} \tag{3}$$

The bias term arises because those with higher values of a are more likely enroll in the treatment, but would experience better outcomes even without it. While the ATT is the estimand of interest in many studies of treatment effects, the possibility of selection bias often threatens the causal interpretation of treated-untreated comparisons; concerns about such bias are almost always motivated, if only implicitly, by a Roy model of enrollment and outcomes.

The purpose of my argument is to show when treated-untreated comparisons for the low-treatment-rate group can be used to partially control for selection bias present in treated-untreated comparisons for the high-treatment-rate group, in order to recover some of the causal content of the latter comparison. Using (3), the difference in treated-untreated mean outcome differences between the high- and low-treatment-rate groups (the *difference in differences*) can be decomposed into differential treatment effect and selection bias terms as

$$\begin{aligned} &[E(y|d=1,h) - E(y|d=0,h)] - [E(y|d=1,l) - E(y|d=0,l)] \\ &\qquad = \underbrace{[(\gamma_{1h} - \gamma_{0h})E(a|a \geq \hat{a}_h) - (\gamma_{1l} - \gamma_{0l})E(a|a \geq \hat{a}_l)]}_{\text{ATT}_h - \text{ATT}_l} \\ &\qquad + \underbrace{\{\gamma_{0h}[E(a|a \geq \hat{a}_h) - E(a|a < \hat{a}_h)] - \gamma_{0l}[E(a|a \geq \hat{a}_l) - E(a|a < \hat{a}_l)]\}}_{\text{Bias}_h - \text{Bias}_l}. \end{aligned} \tag{4}$$

If the selection bias component for the low-rate group exceeds that for the high-rate group,

the second term in the above decomposition will be negative. In this case, subtracting the low-rate group's treated-untreated difference from the high-rate group's difference over-controls for selection bias among the high-rate group, bounding the high-low group difference in ATTs from below. In the special case where the ATT is believed to be nonnegative for the low-rate group (either on the basis of theoretical reasoning or prior empirical evidence), this group difference in ATTs can also be interpreted as a lower bound on the ATT itself for members of the high-treatment-rate group.

Rearranging, the differential selection bias term in (4) will be negative, so that differences in differences will identify a lower bound on the group difference in ATTs, if

$$\frac{\gamma_{0h}}{\gamma_{0l}} \leq \frac{E(a|a \geq \hat{a}_l) - E(a|\hat{a} < \hat{a}_l)}{E(a|a \geq \hat{a}_h) - E(a|\hat{a} < \hat{a}_h)}. \tag{5}$$

In applications, neither side of (5) is likely to be known, or even estimable, with much precision. However, under the *slope condition* that

$$\gamma_{0h} \leq \gamma_{0l}, \tag{6}$$

the left-hand side of (5) can be replaced with unity. In this case, the identification of a lower bound reduces to a comparison of truncated mean differences between the high- and low-treatment-rate groups.

Several aspects of the slope condition, (6), warrant additional discussion. It is sufficient, but necessary; if there is reason to believe that the right hand side of (5) is large, the lower bound argument may apply even if the slope condition is violated. Since it implies that members of the high-rate-group fare worse in the untreated state and may therefore have more to gain from the treatment, it is consistent with observed group differences in treatment rates, and may be acceptable on theoretical grounds alone (this reasoning will not always apply, however, since the present model allows for the possibility that non-outcome factors also influence the enrollment decision). It also places no restrictions on the "returns," γ_{1g}, to a in the treated state.

If theory provides insufficient justification, it is also possible to test whether the slope condition is consistent with data on outcomes and group membership. If outcomes are observed before selection into treatment states (e.g., with panel data or repeated cross sections), the full pre-treatment outcome distribution is observable. In this case, the model implies that slope condition can be tested against the ratio $Var(y|g = h)/Var(y|g = l) = \gamma_{0h}^2/\gamma_{0l}^2$ of pre-treatment outcome variances between the high- and low-treatment-rate groups. If the data are recorded after selection into treatment status, the outcome distributions are

only observable conditional on treatment decisions. In this case, since a normal implies that $Var(a|a < \hat{a})$ is non-decreasing in $\hat{a}$ (see Heckman and Honoré, 1990, Proposition 1), and $\hat{a}_h < \hat{a}_l$ by definition, the inequality

$$\frac{Var(y|d=0, g=h)}{Var(y|d=0, g=l)} = \frac{\gamma_{0h}^2 Var(a|a < \hat{a}_h)}{\gamma_{0l}^2 Var(a|a < \hat{a}_l)} \leq \frac{\gamma_{0h}^2}{\gamma_{0l}^2}$$

implies that the same variance ratio among untreated individuals can be used to test for large deviations from (6). Informally, this ratio of variances can be used to approximate the unconditional ratio (particularly when the treatment rate is small for both groups), and a finding that it does not exceed one suggests that the data are consistent with the slope condition.[3]

When slope condition (6) holds, and since $\hat{a}_h < \hat{a}_l$, the following *increasing-difference-in-truncated means* property is a sufficient condition for (5):

$$\frac{d}{d\hat{a}}[E(a|a \geq \hat{a}) - E(a|a < \hat{a})] \geq 0 \quad \text{for} \quad \hat{a} \geq \hat{a}_h. \tag{7}$$

To understand this condition, note that an increase in the enrollment threshold $\hat{a}$ will increase both the left- and right-truncated means (i.e., the means among those in the treated and untreated states), having an indeterminate effect on their difference. If this crucial property holds, the former effect dominates, so that the selection bias component of treated-untreated comparisons is increasing in the enrollment threshold (equivalently, decreasing in the probability of receiving the treatment) and differences in differences identifies a lower bound on the group difference in ATTs. As the following result shows, this property holds automatically when a is normally distributed and both groups are treated with probability less than $1/2$ (a condition which can be verified directly from data on group membership and treatment status).

Proposition 1. *Suppose that a is normally distributed. Then*

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

when the treatment probability is less than 1/2.

All proofs are presented in the appendix. The following theorem summarizes the basic identification result developed in this section.

[3]When the $\hat{a}_g$, and hence the group-specific treatment rates, are sufficiently different, (5) will hold even when (6) does not, so the argument that follows will still apply as long as γ_{0h}/γ_{0l} is not "too" large, although obviously this statement cannot be made rigorous.

Theorem 1. *Suppose that*

(i) counterfactual outcomes and enrollment follow the Roy model defined by (1) and (2),

(ii) the a are drawn from a normal distribution that does not depend on group membership g,

(iii) slope condition (6) holds, and

(iv) both groups are treated with probability less than 1/2.

Then differences in differences identifies a lower bound on the group difference in ATTs. If, in addition, the ATT is nonnegative for the low-treatment-rate group, differences in differences also identifies a lower bound on the ATT itself for the high-treatment-rate group.

2.1 Non-normality

Many theoretical analyses of self-selection (Borjas 1988; Heckman and Honoré 1990, e.g.) and popular econometric estimators (such as the Heckit, Tobit and switching regression models, see Heckman 1976, 1979; Amemiya 1984) assume that the unobservables that determine counterfactual outcomes and enrollment into treatment are normally distributed, making the result developed above a natural starting point. Such normality assumptions, however, have have been criticized for playing too-important a role in identifying the parameters of economic and econometric models and for introducing inconsistency into estimates of these parameters when unobservables are misspecified as being normal.[4] Because my approach only delivers set identification, it is not subject to the same critiques. The only role that normality plays in the identification argument outlined above is in establishing the existence of a region over which the increasing-difference-in-truncated means property, (7), holds, in which case the selection-bias component of treated-untreated mean differences is decreasing in the probability of receiving the treatment. If the a are drawn from any distribution which shares this property with the normal distribution, the lower bound result will still apply.

The question of robustness to non-normality then becomes one of whether it is reasonable to assume that the a are drawn from a distribution that satisfies the increasing-difference-in-truncated means property. Proposition 1, below, establishes two classes of distributions for which there is some a^* in the support of a beyond which this property holds. The conditions of the proposition are not as strong as they may appear at first glance. As I show below, the

[4]For example, Arabmazar and Schmidt (1982) show that Tobit-type estimators based on normality assumptions can suffer from considerable inconsistency when the data-generating process is non-normal, and, as Olsen (1980) discusses, without additional exclusion restrictions, Heckman's (1979) sample-selection model is not identified under certain distributional assumptions. In addition, Heckman and Honoré (1990) show that when the log-normality assumption is relaxed, the parameters of the classical Roy model are not identified from the cross-sectional distributions of outcomes and treatment states without additional sources of variation.

distributions used in popular econometric models of sample selection, truncation, discrete choice, duration, and reliability fall into one of these classes. This makes the distributional assumptions on which my result relies weaker than those commonly used in practice, and provides an empirical argument for the applicability of the identification procedure. I also show that the conditions of the proposition can be delivered by more primitive restrictions on densities, providing an alternative method of verifying whether those conditions are met, and helping to explain why many distributions meet them.

Proposition 2. *Suppose that a is distributed over $[L, \infty]$ with density f (and distribution function F) satisfying $\lim_{a\to\infty} f = 0$.*

(i) 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.

(ii) 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 $\hat{a}$.[5]

Under the Roy model developed above (i.e., if counterfactual outcomes and enrollment are determined by (1) and (2) and the slope condition, (6), holds), Proposition 2 implies that, if a is drawn from any distribution falling into one of these classes, there is some enrollment threshold above which (or treatment probability below which) differences in differences bounds the differential ATT from below. Note that, in this section, I maintain the assumption that the distribution of a does not depend on group membership g. If the distribution is symmetric and falls into the first class, the lower-bound result will apply as long as both groups are treated with probability less than one half. Without additional specification of the distribution, the treatment probability below which the increasing-differences-in-truncated means property holds is not known for asymmetric members of the first class of distributions. However, if the distribution is skewed right in the (informal) sense that its mean exceeds its median, this property will hold at treatment probabilities less than one half.[6] For mem-

[5] A function is log concave (log convex) if its logarithm is concave (convex).

[6] Proposition 2, and hence the lower-bound identification result, still hold when the median exceeds the mean. In this case, however, the treatment probability below which these results apply is not known without further specification of the distribution of a. For modest treatment probabilities, they will apply unless this distribution is extremely skewed.

bers of the second class of distributions, it holds at any treatment probability. Thus, when the mean is no less than the median—a property exhibited by many social and economic phenomena—a treatment probability of one half can be considered conservative. In fact, the identification argument is sufficiently robust to distributional assumptions in this case that it is unnecessary to specify the class to which f belongs.

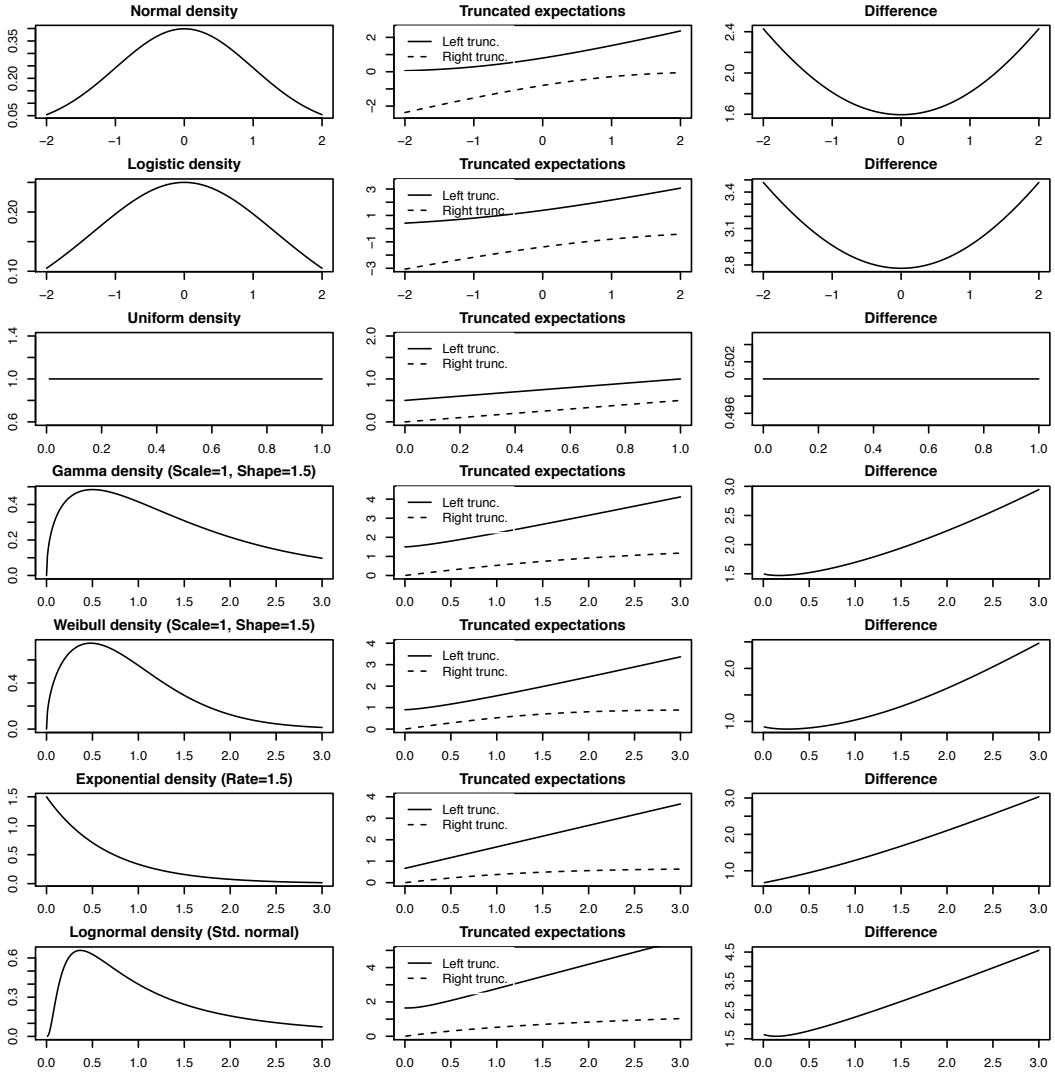
Figures 2 and 3 demonstrate that the distributions used in many common econometric modeling procedures (and implemented in popular statistical packages; see Toomet and Henningsen, 2008 and StataCorp, 2013) fall into one of the classes specified in Proposition 2.[7] Figure 2 plots the density functions, left- and right-truncated expectations, and the difference between these expectations, for the (standard) normal, logistic, uniform, gamma (with shape parameter 1.5), Weibull (with shape parameter 1.5), exponential (with rate parameter 1.5), and (standard) lognormal distributions. For each of these distributions, the left-truncated expectation is at least weakly convex and the right-truncated expectation is at least weakly concave. Accordingly, for each distribution, there exists a range over which the increasing-difference-in-truncated-means property holds at least weakly. For the normal and logistic distributions, this difference is increasing whenever $\hat{a}$ exceeds the mean (i.e., the treatment probability is less than $1/2$). Under, for example, the gamma and Weibull distributions, the threshold at which this difference reaches its minimum is lower (and the corresponding treatment probability is greater).

Figure 3 plots the same functions for the gamma distribution (with shape parameter .75), the Weibull distribution (with shape parameter 1.5), the Pareto distribution (with shape parameter 1.1) and the lognormal distribution (with $\log a \sim N(1, 2)$). For each of these distributions, both the left- and right-truncated expectations are concave and the slope of the left-truncated expectation exceeds that of the right-truncated expectation over the entire support. For these distributions, the difference in truncated means is increasing everywhere, implying that differences in differences identifies a lower bound on the group difference in average treatment effects at any treatment probability or enrollment threshold.

A comparison of Figures 2 and 3 reveals that log concave distributions tend to satisfy the first set of conditions given in Proposition 2 while log convex distributions tend to satisfy the second set. The density functions for the normal, logistic, and exponential densities are log concave, while the gamma and Weibull densities are log concave when their shape parameters exceed one (see Bagnoli and Bergstrom, 2005). As Figure 2 illustrates, these

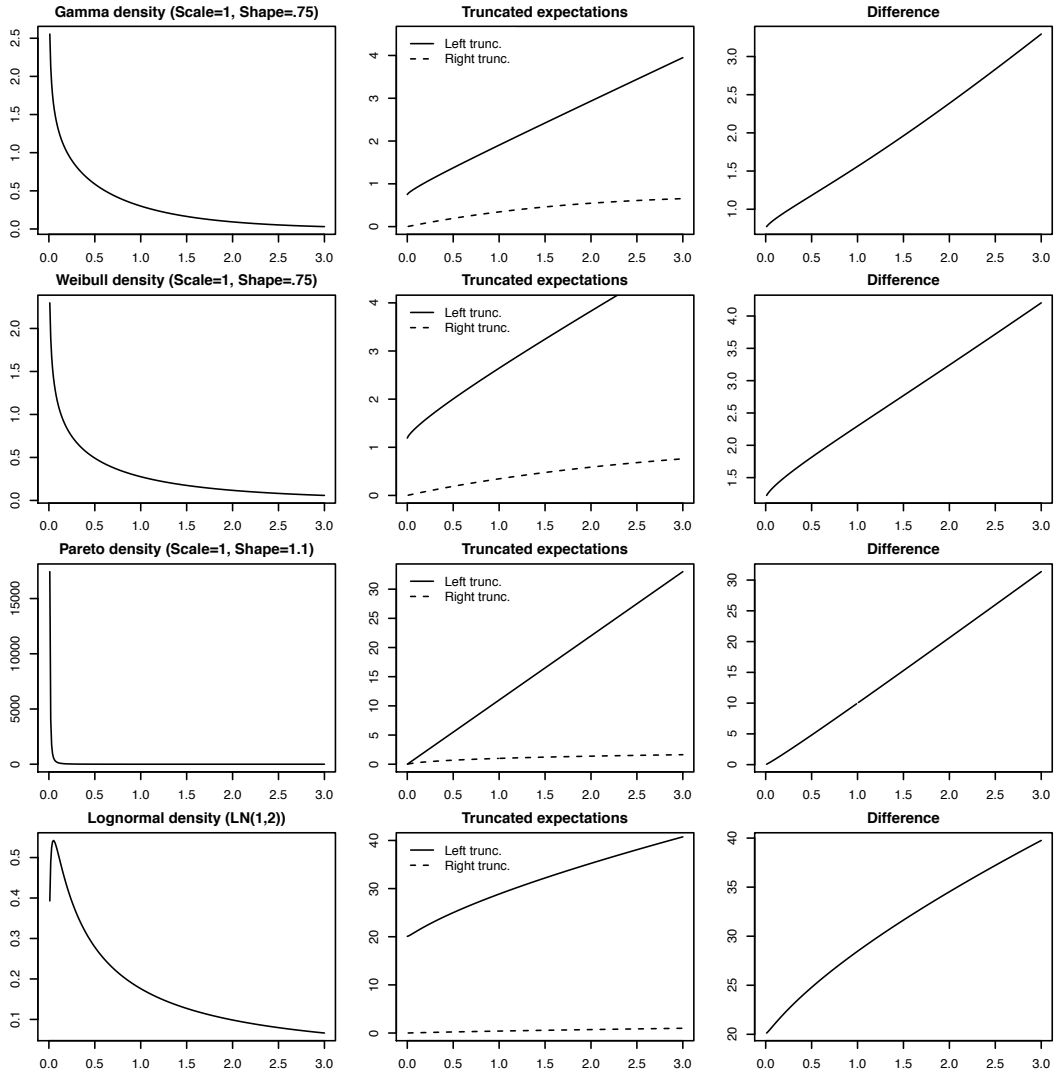
[7]The linear functional form for counterfactual outcomes also suggests a falsification test for whether the type variable meets the conditions of Proposition 2. If there exists a subpopulation among whom treatment is very unlikely, the distribution of y will be close to the distribution of a, making it possible to estimate the curvature of the truncated expectations or test whether the empirical distribution of a approximates some hypothesized distribution.

Figure 2: 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 3: 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).

densities have convex (concave) left-truncated (right-truncated) expectations. In contrast, the Pareto density is log convex and the gamma and Weibull distributions are log convex when their shape parameters are less than one. Figure 3 shows that these distributions have monotone decreasing density functions (and hence log concave distribution functions), concave left-truncated expectations with slopes that exceed one, and concave right-truncated expectations with slopes less than one. The uniform and lognormal densities straddle these cases: the uniform density is both log concave and log convex while the lognormal density is log concave when a is small and log convex when a is large.

Proposition 3 shows that this pattern is not coincidental. Under certain conditions, met by many distributions, log concave densities generate convex and concave left- and, respectively, right-truncated expectations while monotone decreasing log convex densities generate concave left-truncated expectations with slopes that are everywhere smaller than their concave right-truncated expectations.

Proposition 3. *Suppose that a is distributed over $[L, H]$ with density f and*

$$\frac{d}{da}\left|\frac{[\log f(a)]''}{\left\{[\log f(a)]'\right\}^2}\right| \lesseqgtr 0 \quad \textit{when} \quad f'(a) \lesseqgtr 0. \tag{8}$$

(i) 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.

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

In addition to explaining the pattern exhibited by the distributions shown in Figures 2 and 3, Proposition 3 provides easily verifiable conditions under which a hypothesized distribution will satisfy the first criterion given in Proposition 2 (the second criterion can already be verified given an expression for the density function). For example, $|(\log f)''/(\log f)'^2|$ is $1/a^2$ and $2\exp(a)/[1-\exp(a)]^2$ when a is standard normal and, respectively, logistic. These functions are decreasing when a exceeds zero, and the left-truncated expectations for these distributions are convex. When a is Pareto with shape parameter β, this ratio is $1/(\beta+1)$ and the left-truncated expectation is linear. The proposition shows that log concavity is not sufficient for the convexity of truncated expectations. Noting the similarity between condition (8) and the measure of absolute risk aversion, the requirement is, informally, that the log density become less concave as the density decreases.[8]

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

In the present model, increasing-differences-in-truncated-means is only sufficient for differences in differences to identify a lower bound when the slope condition, (6), holds. Tests analogous to those in the normally distributed case are available for whether the data are consistent with this condition. As before, the ratio $Var(y|h)/Var(y|l) = \gamma_{0h}^2/\gamma_{0l}^2$ of pre-enrollment-period outcome variances between the high- and low-treatment-rate groups provides a direct test of this condition. When only post-enrollment data are available, Proposition 4, below, shows that the same ratio among untreated individuals suggests a test for violations of the slope condition when the density, f, of a is log concave, and a test for the slope condition itself when f is log convex. Note that, regardless of whether outcomes are observed before or after selection into treatment, it is possible to test whether f is log concave (see, e.g., An, 1995).

Proposition 4. *Suppose that counterfactual outcomes and enrollment are determined by (1) and (2). Then*

(i) if f is log concave, $Var(y|0,h)/Var(y|0,l) > 1$ implies that $\gamma_{0h} > \gamma_{0l}$;

(ii) if f is log convex, $Var(y|0,h)/Var(y|0,l) < 1$ implies that $\gamma_{0h} < \gamma_{0l}$.

The following theorem amends the previous result to reflect the conclusions of this section.

Theorem 2. *Suppose that*

(i) counterfactual outcomes and enrollment follow the Roy model defined by (1) and (2),

(ii) f satisfies the conditions of Proposition 2, does not depend on group membership g, and the mean of a is no less than its median,

(iii) slope condition (6) holds, and

(iv) both groups are treated with probability less than 1/2.

Then differences in differences identifies a lower bound on the group difference in ATTs. If, in addition, the ATT is nonnegative for the low-treatment-rate group, differences in differences also identifies a lower bound on the ATT itself for the high-treatment-rate group.

2.2 Distributional differences

The results presented so far have been developed under the simplifying assumption that the distribution of a is the same for both the high- and low-treatment-rate groups. In some applications, theory or prior empirical evidence may support this assumption. However, there may equally well be reason to believe that there are between-group differences in the distributions of these unobservables. Indeed, such differences may help explain observed group differences in treatment rates. The conventional, and most direct, way of thinking about distributional

distribution function, and many log convex densities have log concave distribution functions.

differences is to suppose that the group-specific distributions are different members of the same parametric family. In this case, as I show below, the identical distributions assumption can be viewed as a normalization under which the identification results developed above hold without modification. Identification may fail if the group-specific distributions are arbitrarily different. However, it would be unusual—and require strong justification—to suspect that two groups belonging to the same parent population face unobservable distributions that do not belong to, or cannot be represented approximately as belonging to, the same family.

To incorporate distributional differences into the model, assume that the densities f_g, $g \in \{l, h\}$, governing a belong to the same location-scale family with support over $\mathbb{R}$ or the same scale family with support over $\mathbb{R}^+$.[9] Normalize this family of distributions so that f_l is the standard distribution (i.e., set $\sigma_l = 1$ and, in the location-scale case, $\mu_l = 0$). Note that, in this case, $F_h(\hat{a}_h) = F_l[(\hat{a}_h - \mu_h)/\sigma_h]$, so that $(\hat{a}_h - \mu_h)/\sigma_h < \hat{a}_l$ (by the definitions of l and h) and $E(a|h, a \gtreqless \hat{a}_h) = \mu_h + \sigma_h E[a|l, a \gtreqless (\hat{a}_h - \mu_h)/\sigma_h]$.[10] Then the decomposition of group g's treated-untreated mean difference into ATT and bias components can be expressed as[11]

$$\begin{aligned} E(y|1,g) - E(y|0,g) &= \gamma_{1g} E(a|g, a \geq \hat{a}_g) - \gamma_{0g} E(a|a, g < \hat{a}_g) \\ &= (\gamma_{1g} - \gamma_{0g}) E(a|g, a \geq \hat{a}_g) + \gamma_{0g}[E(a|g, a \geq \hat{a}_g) - E(a|g < \hat{a}_g)] \\ &= \underbrace{(\gamma_{1g} - \gamma_{0g}) \left[\mu_g + \sigma_g E\left(a|l, a \geq \frac{\hat{a}_g - \mu_g}{\sigma_g} \right) \right]}_{\text{ATT}} \\ &\quad + \underbrace{\gamma_{0g} \sigma_g \left[E\left(a|l, a \geq \frac{\hat{a}_g - \mu_g}{\sigma_g} \right) - E\left(a|l, a < \frac{\hat{a}_g - \mu_g}{\sigma_g} \right) \right]}_{\text{Bias}}. \end{aligned} \tag{9}$$

Differencing (9) between the high- and low-treatment-rate groups, the selection bias component will be larger for the low-rate group, and differences in differences will bound the group difference in ATTs from below, if (recalling that group l is normalized to have the standard distribution)

$$\frac{\gamma_{0h}\sigma_h}{\gamma_{0l}} \leq \frac{E(a|l, a \geq \hat{a}_l) - E(a|l, a < \hat{a}_l)}{E[a|l, a \geq (\hat{a}_h - \mu_h)/\sigma_h] - E[a|l, a < (\hat{a}_h - \mu_h)/\sigma_h)]}. \tag{10}$$

Here, the corresponding condition under which the left-hand side of (10) can be replaced with unity is that

$$\gamma_{0h}\sigma_h \leq \gamma_{0l}. \tag{11}$$

[9] For most distributions with support on $\mathbb{R}^+$, the mean depends on the scale parameter, so it does not make sense to think of location-scale families over this support.

[10] This follows from the change of variable $a = \mu_h + \sigma_h t$ in $\int_{\hat{a}_h}^{\infty} a f_h(a) da = \int_{\hat{a}_h}^{\infty} a \frac{1}{\sigma_h} f_l\left(\frac{a - \mu_h}{\sigma_h}\right) da$.

[11] When the f_g belong to a scale family, this expression applies with the μ terms omitted.

If this slope condition holds, then since $(\hat{a}_h - \mu_h)/\sigma_h < \hat{a}_l$ by definition, the increasing-difference-in-truncated-means property (7) is a sufficient condition for (10). Thus if the family f_g of distributions satisfies the conditions of Proposition 2, differences in differences will identify a lower bound on the group difference in ATTs.

As in the identical distribution case, the group ratio of pre-treatment-period outcome variances gives a direct test of the slope condition, since $Var(y|h)/Var(y|l) = (\gamma_{0h}\sigma_h)^2/\gamma_{0l}^2$ under the present distributional assumptions. When only post-enrollment data are available, the following result motivates tests analogous to those in the identical-distribution case.

Proposition 5. *Suppose that counterfactual outcomes and enrollment are determined by (1) and (2) and the f_g belong to the same scale or location-scale family. Then*

(i) if the f_g are log concave, $Var(y|h,0)/Var(y|l,0) > 1$ implies that $\gamma_{0h}\sigma_h > \gamma_{0l}$;

(ii) if the f_g are log convex, $Var(y|h,0)/Var(y|l,0) < 1$ implies that $\gamma_{0h}\sigma_h < \gamma_{0l}$.

As before, the untreated outcome variance ratio can provide informal guidance about the plausibility of the slope condition even when only post-enrollment data are available, particularly when the treatment rate is low for both groups.[12]

In summary:

Theorem 3. *Suppose that*

(i) counterfactual outcomes and enrollment follow the Roy model defined by (1) and (2),

(ii) the f_g belong to a scale or location-scale family satisfying the conditions of Proposition 2 and the group-specific means of a are no less than their medians,

(iii) slope condition (11) holds, and

(iv) both groups are treated with probability less than 1/2.

Then differences in differences identifies a lower bound on the group difference in ATTs. If, in addition, the ATT is nonnegative for the low-treatment-rate group, differences in differences also identifies a lower bound on the ATT itself for the high-treatment-rate group.

2.3 Noisy selection

The results presented so far have been based on a particularly stark model model of enrollment that depends on exactly the same unobserved factors that influence counterfactual outcomes. Though this model might approximate the enrollment decision reasonably well in many applications, it might also be inadequate in cases where enrollment is likely to depend on a confluence of factors. Here I show that a simple adaptation of the foregoing Roy model

[12]Note that since log concavity is preserved under linear transformation, every member of a location-scale family has the same log concavity (Bagnoli and Bergstrom, 2005).

allows the identification results developed above to be applied in settings where idiosyncratic factors add noise to the enrollment decision (in Section 3, I allow for this possibility in a less restrictive, but more complex, way).

Suppose, as before, that counterfactual outcomes are given by (1). Let the net expected-utility benefit from receiving the treatment be $\Delta_g - \epsilon$, where Δ_g is a group-specific constant and ϵ is a random variable. For example, ϵ may reflect the combined influence of treatment-induced improvements in outcomes, non-outcome factors that affect the utility of receiving the treatment, and other idiosyncratic factors such as private information and optimization errors, while the Δ_g can be viewed as secular group-specific cost or benefit shifters. In this case, the enrollment decision rule takes the form

$$d(g, \epsilon) = \begin{cases} 1 & \text{if } \Delta_g - \epsilon \geq 0 \\ 0 & \text{otherwise} \end{cases}. \tag{12}$$

To relate this decision rule to those used in the models developed above, note that when only a influences enrollment, (12) reduces to (2).[13]

Because this model is designed to allow for noisy selection into treatment, and in the model the ϵ represent the contribution of a number of factors to the utility of enrolling, it is natural to assume that the density, f_ϵ, of ϵ is a symmetric and log concave member of the first class of distributions given in Proposition 2. These distributional assumptions are similar to (though still weaker than) those made by standard discrete choice models.[14] Note that, because the enrollment criterion is linear, no generality is lost in assuming that the ϵ are distributed identically across groups.[15] Observed group differences in treatment rates imply that $\Delta_h > \Delta_l$.

Under this enrollment process, a simple way to capture the notion of selective enrollment into treatment is to follow Olsen (1980) and Wooldridge (2002, Assumption 17.1) in assuming that

$$E(a|\epsilon) = \rho\epsilon, \tag{13}$$

with $\rho < 0$ to reflect the standard Roy-model hypothesis of positive selection. Olsen (1980) shows that (13) does not impose unreasonable requirements on the distribution of a. Furthermore, when a and ϵ are multivariate normal, this linearity is automatic and the noisy Roy model reduces to the canonical switching regression or Type-5 Tobit model (Amemiya, 1984), particularly if the model is applied within observable covariate strata. More gener-

[13] With $-\epsilon = a$ and $-\Delta_g = \hat{a}_g$; the sign conventions are for consistency with the next section.

[14] The argument can be applied with minor modification when f_ϵ is asymmetric or log convex.

[15] That is, if the $f_{\epsilon g}$ belong to the same location-scale family, put $\Delta_h = (\tilde{\Delta}_h - \mu_{\epsilon h})/\sigma_{\epsilon h}$ where $\tilde{\Delta}_h$ is the unnormalized threshold.

ally, (13) can be viewed as a first-order approximation to an underlying nonlinear conditional expectation.

If the densities f_g, $g \in \{l, h\}$, for a belong to the same scale or location-scale family (where f_l is the standard distribution), the treated-untreated mean difference can be decomposed into ATT and selection bias components as

$$\begin{aligned} E(y|1,g) - E(y|0,g) &= (\gamma_{1g} - \gamma_{0g})E(a|g, \epsilon \leq \Delta_g) + \gamma_{0g}[E(a|g, \epsilon \leq \Delta_g) - E(a|g, \epsilon > \Delta_g)] \\ &= (\gamma_{1g} - \gamma_{0g})[\mu_g + \sigma_g E(a|l, \epsilon \leq \Delta_g)] + \gamma_{0g}[E(a|l, \epsilon \leq \Delta_g) - E(a|l, \epsilon > \Delta_g)] \\ &= \underbrace{(\gamma_{1g} - \gamma_{0g})[\mu_g + \rho\sigma_g E(\epsilon|l, \epsilon \leq \Delta_g)]}_{\text{ATT}} + \underbrace{\gamma_{0g}\rho\sigma_g[E(\epsilon|l, \epsilon \leq \Delta_g) - E(\epsilon|l, \epsilon > \Delta_g)]}_{\text{Bias}}. \end{aligned}$$

Then differences in differences will bound the group difference in ATTs from below if

$$\frac{\gamma_{0h}\sigma_h}{\gamma_{0l}} \leq \frac{[E(\epsilon|\epsilon > \Delta_l) - E(\epsilon|\epsilon \leq \Delta_l)]}{[E(\epsilon|\epsilon > \Delta_h) - E(\epsilon|\epsilon \leq \Delta_h)]}, \tag{14}$$

where the left-hand side of (14) can be replaced with unity as long as slope condition (11) (which asserts that $\gamma_{0h}\sigma_h \leq \gamma_{0l}$) holds. As in the pure Roy model, when pre-enrollment data are available, the ratio $Var(y|h)/Var(y|l)$ represents a direct test of this condition, and the ratio $Var(y|0,h)/Var(y|0,l)$ can be used to provide heuristic guidance about this condition when outcomes are only observed after selection into treatment states, as the following result shows.[16]

Proposition 6. *If counterfactual outcomes and enrollment follow (1), (12), and (13), and f_ϵ is log concave, $Var(y|0,h)/Var(y|0,l) > 1$ implies that $\gamma_{0h}\sigma_h > \gamma_{0l}$.*

When the slope condition holds, and since $\Delta_h > \Delta_l$, a sufficient condition for (14) is that

$$\frac{d}{d\Delta}[E(\epsilon|\epsilon > \Delta) - E(\epsilon|\epsilon \leq \Delta)] \leq 0 \quad \text{when} \quad \Delta < \Delta_l. \tag{15}$$

Note that, because of the enrollment rule used in this model, the sign of inequality in (15) is reversed relative to the analogous inequality in (10), requiring a decreasing-differences-in-truncated-means property to hold. Since f_ϵ is symmetric and belongs to the first class of distributions specified in Proposition 2, (15) will hold as long as both groups are treated with probability less than one half (apply Proposition 2 to $-\epsilon$). The following theorem summarizes this conclusion.

Theorem 4. *Suppose that*

[16]Note, however, that it is not possible in this case to test whether f_ϵ is log concave although, as discussed above, the assumption is natural.

(i) counterfactual outcomes and enrollment follow the Roy model defined by (1) and (12),

(ii) the f_g belong to the same location-scale family, f_ϵ belongs to a symmetric and log concave member of the first class of distributions established in Proposition 2, and the expectation of a given ϵ satisfies (13),

(iii) slope condition (11) holds, and

(iv) both groups are treated with probability less than 1/2.

Then differences in differences identifies a lower bound on the group difference in ATTs. If, in addition, the ATT is nonnegative for the low-treatment-rate group, differences in differences also identifies a lower bound on the ATT itself for the high-treatment-rate group.

2.4 Further modeling considerations

2.4.1 Observable covariates

The Roy models developed above abstract away from observable covariates that may also affect outcomes and enrollment. The most flexible way to allow for observables within the framework of these models is to apply the identification argument (including all restrictions on distributions, outcome variances, and treatment rates) within covariate strata. In this case, the unconditional (on covariates) difference in differences identifies an average lower bound on strata-specific group differences in ATTs.[17] A more conventional approach is to model covariates as affecting outcomes and enrollment through a scalar index $a = h(x, u)$, where h is some function and u is an unobserved random variable, in which case the procedure outlined above can be applied across covariate strata without modification (a noisy selection model can be used to accommodate the possibility that covariates have an effect on enrollment beyond their contribution to a). This approach may be preferable when there are many covariate strata, and can also be implemented within specific strata in order to examine heterogeneity across observables.

2.4.2 Proportional treatment effects

The identification argument can also be used to bound group differences in the average proportional effect, $E[(y_{1g} - y_{0g})/y_{0g}|d = 1, g]$, of the treatment on the treated, rather than the absolute effect. The most straightforward way to apply the identification procedure introduced above to inference about proportional treatment effects is to assume that (1)

[17] With the average being taken with respect to the population covariate distribution. Strata-specific differences in differences can also be aggregated according to another weighting scheme (e.g., a group-specific covariate distribution). Because the strata-specific bounds will differ, a difference-in-differences regression that conditions on covariates will identify a more difficult-to-interpret weighted average of the strata-specific bounds.

determines the logs, rather than levels, of counterfactual outcomes, in which case the group difference in mean log outcome differences can be interpreted as an approximate lower bound on the group difference in proportional treatment effects.

One drawback to modeling proportional treatment effects this way is that it implicitly restricts the form that absolute counterfactual outcomes take. Another is that the empiricist may be interested in both absolute and proportional treatment effects. Maintaining the linear model for absolute counterfactual outcomes, it may be possible to use differences in differences to identify lower bounds on group differences in both the proportional and absolute effects of the treatment on the treated. When absolute outcomes follow (1), the proportional treatment effect, $(\gamma_{1g}-\gamma_{0g})/\gamma_{0g}$, is constant and the group difference in treated-untreated mean log outcome differences decomposes into differential proportional treatment effect and bias terms as

$$
\begin{aligned}
&[E(\log y_{1h}|a \geq \hat{a}_h) - E(\log y_{0h}|a < \hat{a}_h)] - [E(\log y_{1l}|a \geq \hat{a}_l) - E(\log y_{0l}|a < \hat{a}_l)] \\
&= \underbrace{\left[\log\left(\frac{\gamma_{1h}}{\gamma_{0h}}\right) - \log\left(\frac{\gamma_{1l}}{\gamma_{0l}}\right)\right]}_{\approx \mathrm{ATT}_h - \mathrm{ATT}_l} + \underbrace{[E(\log a|a \geq \hat{a}_h) - E(\log a|a < \hat{a}_h)] - [E(\log a|a \geq \hat{a}_l) - E(\log a|a < \hat{a}_l)}_{\mathrm{Bias}_h - \mathrm{Bias}_l}.
\end{aligned}
$$

Provided that the distributions of both a and $\log a$ satisfy the requirements of Proposition 2, there will be a treatment probability below which the differential bias term will be negative and differences in log differences will identify a lower bound on the difference in proportional treatment effects between the high- and low-rate groups. For example, Figure 3 shows that both the normal and lognormal distributions satisfy the requirements of Proposition 2, so differences in differences identifies lower bounds on group differences in both absolute and proportional ATTs when outcomes are linear in a, a is lognormal, and both groups are treated with probability less than one half (this treatment probability is conservative for the absolute effect since the mean of lognormal exceeds its median).

Which model of proportional treatment effects is appropriate depends on the assumptions that the empiricist is willing to make about the underlying Roy model; the log linear model restricts the form that absolute outcomes take while the linear model imposes constant proportional treatment effects. In the next section, I show that the identification argument can be applied under a broad class of counterfactual outcome functions, obviating the need to specify a model for outcomes. However, since identification is more fragile in this case, it may be preferable to use the results derived in this section unless theory or evidence suggest that counterfactual outcomes cannot be adequately described, or approximated, by a linear model.

3 Identification in nonlinear Roy models

The results so far have been presented for Roy models where counterfactual outcomes are a linear function of some unobserved random variable. In this section, I show that similar results can be applied with this linearity relaxed. Before developing the identification argument for the nonlinear case, some comments are in order about the generality of the simpler linear models. The purpose of the identification argument is to provide a causal interpretation of treated-untreated outcome comparisons when the treated population is positively selected on an unobserved factor that also affects counterfactual outcomes. In some applications, this unobserved factor may have a specific interpretation that makes a linear functional form for counterfactual outcomes unrealistic (e.g., if there are diminishing returns to unobserved skill in the labor market). In such applications, the above identification results can still be applied as long as counterfactual outcomes can be approximated by a linear function of a nonlinear transformation of the underlying factor and the distribution of this transformed random variable meets the appropriate conditions.[18] In other applications, there may be no articulable source of concern about selection bias. When it has no clear interpretation, the unobserved factor can be defined in terms of the counterfactual untreated outcome itself. The identification results can then be applied without modification as long as treated outcomes can be approximated as a linear function of untreated outcomes and the distribution of untreated outcomes satisfies the appropriate conditions.

From this viewpoint, the linear models developed in Section 2 are widely applicable. In some specific applications, however, these models' assumptions and restrictions may seem unacceptably strong. For example, theory or existing evidence may suggest that a linear function will not provide a good approximation to the effect of the treatment. Alternatively, researchers may be concerned about selection on a specific unobserved factor, but be hesitant to take a stance on the functional forms through which this factor determines counterfactual outcomes, or make distributional assumptions about transformations of this factor. The results in this section extend the identification argument to a large class of semi-parametric Roy models of enrollment and counterfactual outcomes. As I discuss below, a drawback to the additional generality that these results provide is that the conditions they require for differences in differences to identify a lower bound are more intricate, and more difficult to verify, than in the linear case. The results may nevertheless be useful if a linear model is judged inadequate.

[18] The natural example is where $y_{dg}(a) = \gamma_{dg} \log a$, and the distribution of $\log a$ satisfies the conditions of Proposition 2 (note that this is different from the models of proportional treatment effects discussed previously). Since log-concavity (-convexity) is preserved by weakly concave (convex) transformations, these distributional assumptions are not particularly strong (An, 1995; Bagnoli and Bergstrom, 2005).

To extend the noisy Roy model developed in Section 2.3 to allow the unobserved factor a that determines counterfactual outcomes to enter nonlinearly into the enrollment decision, let the net benefit from enrollment be given by $\tilde{\gamma}(\Delta_g, a) - \epsilon$. In this expression, $\tilde{\gamma}(\Delta_g, a)$ represents the component of this benefit explained by a, and ϵ represents the contribution of idiosyncratic factors that do not also influence counterfactual outcomes to this benefit. The Δ_g are group-specific cost or benefit shifters included to reflect group differences in treatment rates. To model these group differences flexibly, let $\tilde{\gamma}$ take the semi-parametric form

$$\tilde{\gamma}(\Delta_g, a) \in \{\Delta_g + \gamma(a), \Delta_g \gamma(a), \gamma(\Delta_g + a)\}. \tag{16}$$

That is, the group-specific functions $\tilde{\gamma}(\Delta_g, a)$ that summarize the component of the benefit of enrolling explained by a are either vertical, multiplicative, or horizontal translates. In addition, assume that γ satisfies:

$$\gamma'(a) \geq 0, \ \gamma''(a) \leq 0. \tag{17}$$

The assumption that γ is increasing reflects the hypothesis of positive selection into treatment. The assumption that it is concave reflects the standard hypotheses of diminishing marginal utility or returns. These functional forms capture the most natural sources, and most common models, of group differences in utility (i.e., increasing, decreasing, or constant returns to Δ).[19]

The enrollment decision rule is then

$$d(g, a, \epsilon) = \begin{cases} 1 & \tilde{\gamma}(\Delta_g, a) - \epsilon \geq 0 \\ 0 & \text{otherwise} \end{cases}, \tag{18}$$

and group differences in treatment rates imply that $\Delta_h > \Delta_l$. Note that, as in the previous models, this rule allows for the possibility that factors other than counterfactual outcomes influence the decision to enroll. Here, unlike the model of Section 2.3, the dependence of the enrollment decision on the factor a that determines counterfactual outcomes is modeled directly through the $\tilde{\gamma}$ term, and the ϵ only represent idiosyncratic factors that make this decision noisy (in the previous model, the ϵ represented both idiosyncratic factors and those

[19]It should be noted that the restriction that $\tilde{\gamma}$ takes one of the forms in (16) is stronger than necessary. In fact, as the proof of this section's main result makes clear, the argument can be applied for any function for which $\partial\tilde{\gamma}/\partial\Delta$ is monotone in a. Accordingly, the identification procedure may remain applicable when there is reason to believe that enrollment follows a rule that is not well-approximated by (16). However, assuming that (16) holds simplifies the analysis in the next section and allows me to avoid presenting a complicated set of functional-form-contingent results. It is also trivial to extend the argument to the case where $\tilde{\gamma}$ is a linear function of $\gamma(a)$ (and hence depends on a vector of parameters rather than a scalar Δ).

related to outcomes). Accordingly, I assume that ϵ is distributed independently of a and g with support $\mathbb{R}$, density f, and distribution function F.[20] To accommodate nonlinearity in counterfactual outcomes and the decision rule, I assume in this section that a has support over $\mathbb{R}^+$ with density $\pi(a)$. Unlike the linear case, the requirements for identification are somewhat stronger when this density varies by group. Since in certain applications there may be reason to believe that the group-independence holds, I consider the case of different distributions separately below.

Suppose that counterfactual outcomes $y_{dg}(a)$ satisfy

$$y'_{dg}(a) \geq 0 \tag{19}$$

and

$$y'_{0h}(a) \leq y'_{0l}(a) \tag{20}$$

for $d \in \{0, 1\}$ and $g \in \{l, h\}$. Conditions (19) and (20) are the nonlinear counterparts of the corresponding conditions (1) and (6) for the linear Roy models of Section 2. The assumption that outcomes are increasing in a is consistent with the hypothesis of positive selection. Prior knowledge of, or hypotheses about, the forms that the counterfactual outcome functions take may motivate the use of a nonlinear Roy model and, consequently, justify the nonlinear slope condition, (20). Otherwise, though this condition is not directly verifiable, the following result shows that untreated outcome variance ratios between the high- and low-treatment-rate groups can once again be used to test whether the slope condition is consistent with data on outcomes and group membership, regardless of whether outcomes are observed before or after enrollment decisions are made.

Proposition 7. *Suppose that enrollment and counterfactual outcomes follow (16), (17), (18), and (19), and that π is independent of g. Then, by a first-order approximation, the following hold:*

(i) In the pre-enrollment period, $Var(y|h)/Var(y|l) > 1$ implies that $E(y'_{0h}) > E(y'_{0l})$.

(ii) In the post-enrollment period, if π is log concave then $Var(y|0,h)/Var(y|0,l) > 1$ implies that $E(y'_{0h}) > E(y'_{0l})$. If π is log convex then $Var(y|0,h)/Var(y|0,l) < 1$ implies that $E(y'_{0h}) < E(y'_{0l})$.

These variance ratio tests are weaker than their linear-model counterparts in that they convey only approximate information about whether the slope condition holds on average,

[20] When $\tilde{\gamma}$ is linear in γ, it is straightforward to extend the argument to allow the distributions of ϵ for the high- and low-treatment-rate groups to belong to the same location-scale family (the group-specific means and variances can be absorbed into the function $\tilde{\gamma}$).

while identification formally requires that it hold pointwise. Note that the log concavity conditions in Proposition 7 refer to the distribution governing a and not ϵ.

In this model, the difference in mean outcomes between treated and untreated individuals can be decomposed into ATT and selection bias terms according to

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

where $p(a|g,d)$ denotes the density of a conditional on group membership g and treatment status d. The difference in treated-untreated mean differences between the high- and low-treatment-rate groups will therefore bound the group difference in ATTs from below if the selection bias component is larger for the low-treatment-rate group than for the high-rate group, or if

$$\int \left[p(a|1,h) - p(a|0,h)\right] y_{0h}(a)da \leq \int \left[p(a|1,l) - p(a|0,l)\right] y_{0l}(a)da. \tag{21}$$

If, in addition, the ATT is nonnegative for the low-treatment group, differences in differences will also represent a lower bound on the ATT itself for the high-rate group.

Proposition 8, below, provides a sufficient condition for (21). The intuition behind the lower-bound identification result is similar to in the linear case. Imagine that initially both groups enroll in the treatment with equal probability, and consider the effect of an increase in one group's cost or benefit shifter Δ_g. This change will make the treatment more attractive for members of group g, increasing the probability of enrollment at every level of a within this group by $f(\tilde{\gamma})(\partial\tilde{\gamma}/\partial\Delta)d\Delta$. If the associated increases in the probability of enrolling are higher for those with relatively low untreated outcomes, selection bias will be decreasing in the treatment rate because the marginal enrollee will have a lower expected realization of a than among the initial population of enrollees. This would hold, for example, if high-outcome (i.e. high-a) individuals enrolled with high probability even before the increase in Δ_g.[21] The following result makes this intuition rigorous.

Proposition 8. *Suppose that enrollment and counterfactual outcomes follow (16), (17), (18), (19), and (20). Suppose, in addition, that π is independent of g, f is log concave, and both groups are treated with probability less than 1/2. Then (21) holds if there is a $g \in \{l,h\}$*

[21] Indeed, in a simplified case with $a \in \{0,1\}$ and $\tilde{\gamma}(\Delta,a) = \Delta + \gamma 1_{a=1}$, the key condition for identification is that $f(\Delta+\gamma) \leq f(\Delta)$. The details of this argument are available upon request.

such that

$$Cov\left[f\left(\tilde{\gamma}(\Delta,a)\right)\frac{\partial\tilde{\gamma}(\Delta,a)}{\partial\Delta},y_{0g}(a)\right]\leq 0 \tag{22}$$

between Δ_l *and* Δ_h.

Although, in the interest of brevity, I present the analytical details behind Proposition 8 in its proof in the Appendix, the underlying structure of the argument is straightforward. First, I show that, under the slope condition, if (21) holds when both groups are assigned the same untreated outcome functions, then (21) itself holds. I then show that the derivative of the selection bias term with respect to Δ, holding untreated outcomes constant, is bounded from above by the covariance defined in (22). This *covariance condition* formalizes the above intuition that selection bias will be decreasing in the treatment rate if an increase in the cost or benefit shifter causes a disproportionate number of low-untreated-outcome individuals to enroll. Because it is critical for identification, I discuss this condition at greater length, and develop an informal test of its consistency with data, in Section 3.1, below.

The other conditions required by Proposition 8 are largely similar to those required for identification in the preceding Roy models. As in the noisy Roy model of Section 2.3, the requirement that the distribution of the idiosyncratic component, ϵ, of the enrollment equation be log concave is natural and met by distributions used in discrete choice models. Since log concavity of the density is not sufficient for the convexity of the truncated expectation, the distributional requirements of Proposition 8 are weaker than the corresponding requirements for the linear Roy models developed above. Note that Proposition 8 itself places no restrictions on the distribution, π, governing a.

The following theorem presents a formal summary of the identification results developed in this section.

Theorem 5. *Suppose that*

(i) enrollment and counterfactual outcomes follow the Roy model defined by (16), (17), (18), and (19),

(ii) the group-specific densities, π_g, $g\in\{l,h\}$, *of a belong to the same scale family and the density,* f, *of* ϵ *is log concave,*

(iii) outcomes satisfy slope condition (20),

(iv) covariance condition (22) holds, and

(v) both groups are treated with probability less than 1/2.

Then differences in differences identifies a lower bound on the group difference in ATTs. If, in addition, the ATT is nonnegative for the low-treatment-rate group, differences in differences also identifies a lower bound on the ATT itself for the high-treatment-rate group.

3.1 Assessing the covariance condition

The functional form and distributional assumptions required by Theorem 5 for differences in differences to identify a lower bound on the group difference in treatment effects when enrollment and outcomes are nonlinear are relatively mild. The key requirement is the covariance condition, (22), that untreated outcomes covary negatively with cost- or benefit-shifter-induced changes in the probability of enrolling. Because this covariance depends simultaneously on four potentially nonlinear functions (the distributions of ϵ and a as well as the selection and untreated outcome equations), it is natural to ask when these conditions are likely to hold. Since y_{0g} is increasing by assumption, relatively large values of y_{0g} will be associated with relatively small values of the Δ-induced change $f(\tilde{\gamma})(\partial\tilde{\gamma}/\partial\Delta)d\Delta$ in the probability of receiving the treatment when this change tends to be a decreasing function of a. Informally, the covariance between untreated outcomes and these changes will be negative if Δ is sufficiently large that this change tends to be decreasing (i.e., if the treatment probability is high even before the change in Δ, so that f tends to be decreasing), if the density of a is larger where this change is decreasing, or if untreated outcomes increase more rapidly with a when this change is decreasing.[22]

Thus, the covariance condition does not appear to impose unreasonable requirements on the data, although it is unlikely to be acceptable on theoretical grounds alone. Putting $d\Delta = \Delta_h - \Delta_l$, (22) implies that the difference, $P(1|h,a) - P(1|l,a)$, in treatment probabilities between the high- and low-treatment-rate groups covaries negatively with untreated outcomes, $y_{0g}(a)$. Since a is not observed and $y_{0g}(a)$ is nonlinear, the signs of these covariances cannot be verified directly. However, if there is a proxy, q, related to enrollment through its correlation with a, it possible to estimate the components

$$Cov\left[P(1|h,q) - P(1|l,q), E(y_{0g}|q)\right] \tag{23}$$

of these covariances that are explained by q. Since, by the law of total covariance, the unconditional covariance consists of the component explained by q and the average covariance within q strata, a finding that the sample analogs of (23) are negative would add considerable

[22] A simple numerical example illustrates this logic. Suppose that the outcome equations take the linear forms $y_1 = 3a$ and $y_0 = a$ and that the selection equation is a probit in the difference in counterfactual outcomes, so that $d = 1(\Delta + 2a - \epsilon)$ with ϵ standard normal. When $\log(a)$ is standard normal, simulation reveals that $Cov[f(\Delta+2a), a]$ is negative when Δ is large enough to generate an average treatment probability $E[F(\Delta + 2a)]$ greater than about .3. Instead, if $\log a \sim N(0,2)$, in which case the right tail of the type distribution is fatter (and hence places more weight on points of the support of a where f is decreasing), the simulated covariance is negative when Δ is large enough to generate an average treatment probability of about .15 or greater.

credence to the hypothesis that covariance condition (22) itself holds in a given application.[23] Observable covariates, which are often included in empirical treatment-effect models to control for unobservable determinants of enrollment and outcomes rather than because they are of intrinsic interest, are natural candidates for proxies with which this heuristic test can be implemented (this approach is similar in spirit to that of Altonji, Elder, and Taber, 2005).[24]

3.2 Distributional differences

As in the pure and noisy linear models of Section 2, the identification results for the nonlinear Roy model can still be applied when the distribution of a differs between the low- and high-treatment-rate groups, although the requirements for robustness to distributional differences are slightly stronger in the nonlinear case. As before, a simple and flexible way to accommodate distributional differences is to let the group-specific densities π_g, $g \in \{l, h\}$, for a belong to the same scale family (since these distributions have support on $\mathbb{R}^+$, their means will also depend on the group-specific scale parameters), normalized so that $\sigma_l = 1$.

Assume, as in the identical distribution case, that the utility from enrolling and the enrollment decision rule follow (16), (17), and (18). Here, observed group differences in treatment rates may be explained by group differences in the enrollment decision rules, the distributions of a, or both. Writing the group-specific treatment probabilities in terms of draws from the low-rate-group's distribution of a as $P(d = 1|g) = E\left[F\left(\tilde{\gamma}(\Delta_g, \sigma_g a)\right)|g = l\right]$, a higher treatment rate for group h implies that $P(1|h) - P(1|l) \approx [\partial P(1|g)/\partial\theta']d\theta \equiv dP(1) > 0$, where $\theta = (\Delta, \sigma)$ and $d\theta = \theta_h - \theta_l$.

Similarly, assume that counterfactual outcomes satisfy (19) and the modified slope condition that

$$y'_{0h}(\sigma_h a)\sigma_h \leq y'_{0l}(a). \tag{24}$$

As before, the ratio of outcome variances between untreated members of the high- and low-rate groups provides an approximate test for whether the data are consistent with this slope

[23] In addition, if outcomes are only observed after enrollment decisions have been made, (23) 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 log concave densities are unimodal, the change $f(\tilde{\gamma})(\partial\tilde{\gamma}/\partial\Delta)$ in the treatment probability is likely to be small for those with large realizations of a (and therefore high enrollment likelihoods). As a consequence, the covariance among the untreated (which will exclude those with the lowest treatment probability changes but highest untreated outcomes) is likely to overstate the unconditional covariance.

[24] Naturally, this precludes the application of the identification argument within q strata. If covariates influence a through a single-index model of the form $a = h(x, u)$, any covariate can be considered such a proxy, though the test is more informative if x and u are themselves correlated. Informally, $\tilde{\gamma}$ can be interpreted as reflecting the component of any direct influence of the covariates on enrollment that depends on a (this interpretation can be formalized by letting the density of ϵ depend on, and be log concave for all, a).

condition holding on average, as the following result establishes.

Proposition 9. *Suppose that enrollment and counterfactual outcomes follow (16), (17), (18), and (19), and that the π_g belong to the same scale family. Then by a first-order approximation, the following hold:*

(i) In the pre-enrollment period, $Var(y|h)/Var(y|l) > 1$ implies that $E[y'_{0h}(\sigma_h a)\sigma_h|l] > E[y'_{0l}(a)|l]$.

(ii) In the post-enrollment period, if π is log concave, $Var(y|0,h)/Var(y|0,l) > 1$ implies that $E[y'_{0h}(\sigma_h a)\sigma_h|l] > E[y'_{0l}(a)|l]$. If π is log convex, $Var(y|0,h)/Var(y|0,l) < 1$ implies that $E[y'_{0h}(\sigma_h a)\sigma_h|l] < E[y'_{0l}(a)|l]$.

The decomposition of treated-untreated differences in mean outcomes into ATT and selection bias terms, and hence condition (21) under which the difference in this difference between the high- and low-treatment-rate groups bounds the group difference in ATTs from below, are the same as in the identical distribution case. The intuition behind identification is also similar: if the typical individual induced to enroll by simultaneous changes in the cost or benefit shifter Δ and the distribution of a would experience relatively low untreated outcomes, selection bias will be decreasing in the treatment probability. However, when the density of a is allowed to dilate at the same time that Δ changes, the relationship between outcomes and changes in the probability of enrolling may be non-monotonic. For example, suppose that an increase in Δ puts upward pressure on the enrollment probability while an accompanying decrease in σ puts downward pressure on that probability, but that these competing effects combine to increase the population treatment probability. It is possible that this increase is driven by those with relatively low and relatively high realizations of a, while in the middle of the distribution the second effect dominates, decreasing the number of enrollees. Such a reduction in enrollment among those with middling outcomes may increase selection bias, even if outcomes and changes in the treatment probability covary negatively across the population. As the following result shows, a simple and concise sufficient condition to rule out such non-monotonicity is that the function γ exhibit constant relative risk aversion.[25] This condition restricts the relative magnitudes of the competing effects described above, which operate through the derivatives of γ, ensuring that when untreated outcomes are negatively correlated with changes in the treatment probability, selection bias is decreasing in that probability.

Proposition 10. *Suppose that enrollment and counterfactual outcomes follow (16), (17), (18), (19), and (24). Suppose in addition that the π_g belong to the same scale family, γ is*

[25]Note the symmetry between this condition and the requirement of Proposition 3 that $\log f$ exhibit declining absolute risk aversion, which is implied by constant relative risk aversion.

CRRA, f is log concave, and both groups are treated with probability less than 1/2. Then (21) holds if there is a $g \in \{l, h\}$ such that

$$Cov\left[f\left(\tilde{\gamma}(\Delta, \sigma a)\right)\frac{\partial\tilde{\gamma}(\Delta, \sigma a)}{\partial\theta'}d\theta, y_{0g}(\sigma_g a)\right] \leq 0 \tag{25}$$

between θ_l and θ_h.

Note well that the CRRA condition is sufficient but by no means necessary, and that, while it places additional structure on the model, it still allows the identification procedure to be applied under a broad and flexible range of enrollment equations that includes the functional forms most frequently used in theory and applications.[26] The informal test developed in Section 3.1 can be applied to the modified covariance condition, (25), without change.

By way of peroration:

Theorem 6. *Suppose that*

(i) enrollment and counterfactual outcomes follow the Roy model defined by (16), (17), (18), and (19), with γ CRRA,

(ii) the group-specific densities, π_g, $g \in \{l, h\}$, of a belong to the same scale family and the density, f, of ϵ is log concave,

(iii) outcomes satisfy slope condition (24),

(iv) covariance condition (25) holds, and

(v) both groups are treated with probability less than 1/2.

Then differences in differences identifies a lower bound on the group difference in ATTs. If, in addition, the ATT is nonnegative for the low-treatment-rate group, differences in differences also identifies a lower bound on the ATT itself for the high-treatment-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 tremendous number of Southern-born blacks left the South in favor of cities in the

[26]The statement of Proposition 10 also sacrifices some generality in the interest of parsimony. A similar result could be obtained by placing weaker restrictions on the function $\tilde{\gamma}$, but these restrictions would appear arbitrary and difficult-to-verify while the CRRA restriction is easily recognizable and interpretable. Identification also holds under weaker conditions for certain specifications of the enrollment equation (the proof shows, e.g., that when $\tilde{\gamma}(\Delta, \sigma a) = \gamma(\Delta + \sigma a)$, the result holds for any concave function γ). The CRRA assumption allows me to present one condition that works for any enrollment decision rule.

North. This episode is widely 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 of the Great Migration and its effects). 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 log 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 declines in the black-white wage gap. This estimand can easily be recovered from black-white differences in North-South wage differentials. However, since characteristics not observed in the data 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 descriptive and not necessarily 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 presented above provide a framework for understanding racial differences in North-South wage differentials in terms of the causal impact of migration on

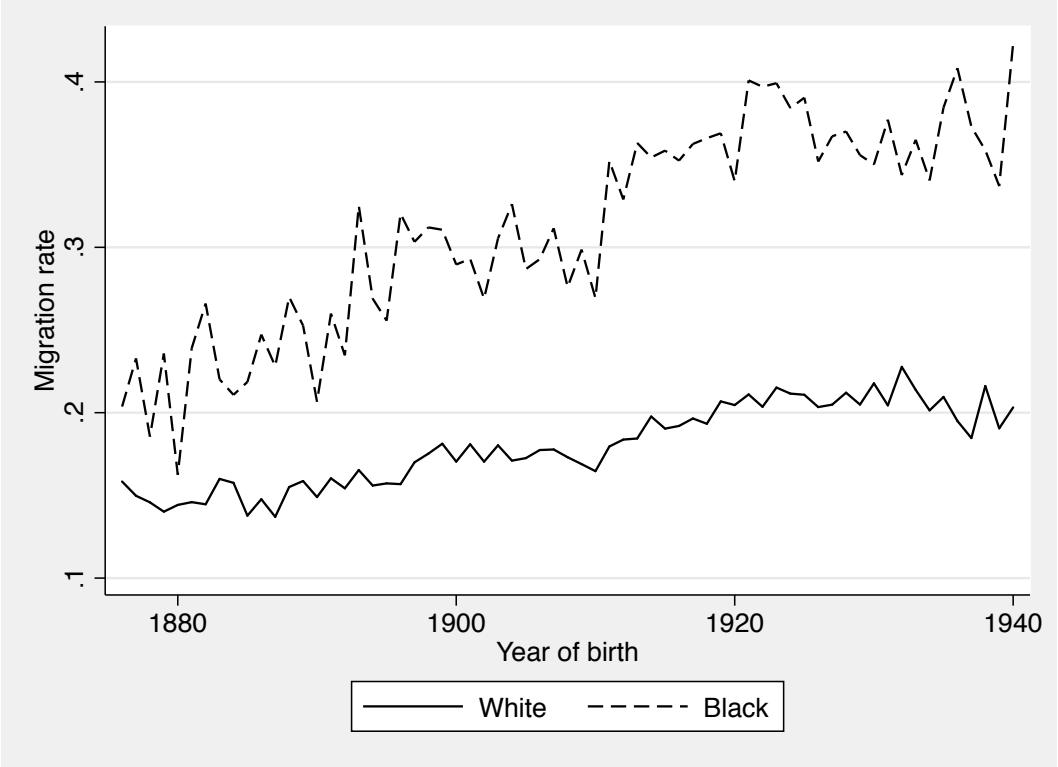
wages. In fact, because they recover bounds on group-differences in average treatment effects, the identification procedure developed in this paper is particularly well-suited to the analysis of the impacts of the Great Migration. As I note above, a limitation to my identification argument is that, unless 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, however, 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 subtracting the nominal wage effect of migration among whites from that for blacks removes the component of that effect due to these factors (at least insofar as this component is similar for blacks and whites). In the context of the Great Migration, then, between group heterogeneity conveys information about the extent to which black migrants earned more in the North because its denizens were less discriminatory than their Southern counterparts.

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, as well as the arduous undertaking that their data collection effort represents, underscore the usefulness of the identification results presented in this paper. To my knowledge, this is the first paper to provide rigorous evidence about the racial difference in the causal effect of the Great Migration on migrants' wages.

4.2 Descriptive evidence

Figure 4 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

Figure 4: 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.

black and white men.[27] 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 of migrating. 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.

[27]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 wage data are available.

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 presents regression estimates of the black-white difference in North-South (annual) wage and log wage differentials over the same periods.[28] 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, 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. The results are 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.

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. When counterfactual wages in the South and North are approximately

[28]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.

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 $\sigma_W/\sigma_B \neq 1$.

linear in unobserved skill (or some nonlinear transformation of skill), wages and migration can be viewed through the lens of Roy models developed in Section 2. The requirements for differences in differences to bound the black-white difference in ATTs (and the black ATT itself) from below are broadly similar for all of these models: (i) the distributions of the unobservables (skill, a transformation of skill, or a composite of skill and idiosyncratic factors) must meet the criteria specified in Proposition 2, (ii) the treatment probability must be less than 1/2 for both groups, and (iii) counterfactual outcomes must satisfy an appropriate slope condition (either or (11) or (14), depending on whether there are group differences in the skill distribution).

As discussed in Section 2, the distributional requirements for identification are met by those employed in standard econometric models of sample selection, truncation, discrete choice, duration, and reliability. Figure 4 and Table 1 show that blacks migrated with higher probability than whites and that the migration probability was less than 1/2 for both groups. Although wage data are available only after the beginning of the Great Migration, Propositions 4, 5, and 6 imply that the data are consistent with the appropriate slope condition 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 the assumption that the wage functions are linear in skill (or that the distributional criteria hold after applying a suitable transformation to skill) is strong, the nonlinear Roy model framework of Section 3 may still apply. In addition to the (previously verified) treatment probability conditions, and the (relatively weak) log concavity assumption on the distribution of the idiosyncratic component of the enrollment equation, identification under this model requires a nonlinear slope condition (either (20) or, if the black and white skill distributions differ, (24), in which case we also need that γ is CRRA) and the covariance condition (25).

In light of Propositions 7 and 9, the results of the variance tests presented in Table 3 provide no evidence against the appropriate slope conditions. To implement the heuristic test proposed in Section 3.1 for whether the data support the covariance condition, I assign observations to education$\times$age$\times$birthplace$\times$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, 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), implying that the black-white difference in selection bias components is bounded above by zero. The results of these tests suggest that the lower-bound argument can be applied even if wages and migration depend nonlinearly on unobserved skill.

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 all 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,

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.

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 perennial 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. In this context, group differences in treatment effects are particularly interesting because they convey information about the return to reduced discrimination in the North. I find 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 size of the white migrant flows and wage differentials suggests that migration did not decrease their wages, 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.

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.

Arabmazar, A., and P. Schmidt (1982): "An Investigation of the Robustness of the Tobit Esitmator to Non-Normality," *Econometrica*, 50(4), 1055–1063.

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.

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.

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.

Greene, W. H. (2011): *Econometric Analysis*. Prentice Hall, 7th edn.

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.

——— (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.

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.

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

Quandt, R. E. (1958): "The Estimation of the Parameters of a Linear Regression System Obeying Two Separate Regimes," *Journal of the American Statistical Association*, 53(284), 873–880.

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

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

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

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

StataCorp (2013): *Stata 13 Base Reference Manual*. Stata Press, College Station, Texas.

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.

Toomet, O., and A. Henningsen (2008): "Sample Selection Models in R: Package sampleSelection," *Journal of Statistical Software*, 27(7).

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

A Proofs

All of the proofs below assume that functions are continuously differentiable and that means are finite.

A.1 Section 2

Proof of Proposition 1. A standard result on truncated normal variables is that (supposing without loss of generality that a is standard normal) $E(a|a \geq \hat{a}) = \lambda(a) = \phi(a)/\Phi(-a)$ and $E(a|a < \hat{a}) = -\lambda(-a) = -\phi(a)/\Phi(a)$, where λ, ϕ and Φ are the inverse Mill's ratio, and the standard normal density and distribution functions, respectively (see, e.g., Heckman, 1979; Wooldridge, 2002; Greene, 2011). Thus, $E(a|a \geq \hat{a}) - E(a|a < \hat{a}) = \lambda(a) + \lambda(-a)$. By inspection, this function has a critical point at zero. Furthermore, Heckman and Honoré (1990) showed that $\lambda(a)$ is strictly convex, so $\lambda(a)+\lambda(-a)$ is strictly convex as well, implying that the function reaches its unique minimum at zero and is increasing for all $\hat{a} > 0$. □

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

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

At the median, $\tilde{a}$, of a, this expression becomes $4f(\tilde{a})\,[E(a) - \tilde{a}]$. 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 $\hat{a}$. □

The proof of Proposition 3 relies on an extension of the Prékopa-Borell theorem (Prékopa, 1971, 1973; Borell, 1975) due to Mares and Swinkels (2014).[29] 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 3. 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.

[29] 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,

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

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. □

Proof of Propositions 4 and 5. I prove proposition 5; Proposition 4 follows by setting $\mu_h = 0$ and $\sigma_h = 1$. 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 6. Because $E(a|g,\epsilon) = \mu_g + \rho\sigma_g\epsilon$, we can write $a = \mu_g + \rho\sigma_g\epsilon + u_g$, where u_g is uncorrelated with ϵ and $Var(u_g) = \sigma_g^2 - \rho^2\sigma_g^2\sigma_\epsilon^2$. Then

$$\begin{aligned} Var(a|g, \epsilon > \Delta_g) &= \rho^2\sigma_g^2 Var(\epsilon|\epsilon > \Delta_g) + \sigma_g^2 - \rho^2\sigma_g^2\sigma_\epsilon^2. \\ &= \sigma_g^2\{1 + \rho^2[Var(\epsilon|\epsilon > \Delta_g) - \sigma_\epsilon^2]\}. \end{aligned}$$

Then since $\Delta_h > \Delta_l$ and f_ϵ log concave implies that $dVar(\epsilon|\epsilon > \Delta)/d\Delta \leq 0$ (Heckman and Honoré, 1990),

$$\frac{Var(y|h,0)}{Var(y|l,0)} = \frac{(\gamma_h\sigma_h)^2}{\gamma_l^2}\frac{Var(a|\epsilon > \Delta_h)}{Var(a|\epsilon > \Delta_l)} < \frac{(\gamma_h\sigma_h)^2}{\gamma_l^2},$$

$Var(y|h,0)/Var(y|l,0) > 1$ implies $\gamma_h\sigma_h > \gamma_l$. □

A.2 Section 3

Proof of Propositions 7 and 9. I prove Proposition 9; Proposition 7 follows with $\sigma_l = \sigma_h = 1$ and $\mu_l = \mu_h$.

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,

$$Var(a|l, \tilde{\gamma}(\Delta, \sigma a) < \epsilon) = E\left[Var\left(a|l, \epsilon, a < \frac{\tilde{\gamma}^{-1}(\Delta, \epsilon)}{\sigma}\right)\right] + Var\left[E\left(a|l, \epsilon, a < \frac{\tilde{\gamma}^{-1}(\Delta, \epsilon)}{\sigma}\right)\right], \tag{26}$$

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 (26) is approximately

$$\frac{\partial}{\partial(\tilde{\gamma}^{-1}/\sigma)} Var\left(a|l, a < \frac{\tilde{\gamma}^{-1}(\Delta, \mu_\epsilon)}{\sigma}\right)\left(\frac{\partial \tilde{\gamma}^{-1}(\Delta, \mu_\epsilon)/\sigma}{\partial \Delta} d\Delta - \frac{\tilde{\gamma}^{-1}(\Delta, \mu_\epsilon)}{\sigma^2} d\sigma\right). \tag{27}$$

By Proposition 1 of Heckman and Honoré (1990), π log concave implies that the leading term in (27) 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 (26) is approximately negative if π is log concave and positive if π is log convex.

The second term in (26) 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 as limit operations can be interchanged, the derivative of this term with respect to a vector of parameters θ can be expressed

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

Since, by a first-order approximation about μ_ϵ,

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

the change in the second term in (26) 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 proofs of Propositions 8 and 10 make 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.[30] 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 limit operations 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

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

[30] 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 well 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

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

□

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

$$\frac{f\left(\tilde{\gamma}(\theta,a)\right)}{F\left(\tilde{\gamma}(\theta,a)\right)} \frac{\partial \tilde{\gamma}(\theta,a)}{\partial \theta'} d\theta - \frac{E\left[f\left(\tilde{\gamma}(\theta,a)\right) \frac{\partial \tilde{\gamma}(\theta,a)}{\partial \theta'} d\theta\right]}{E\left[F\left(\tilde{\gamma}(\theta,a)\right)\right]}. \tag{28}$$

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 (28) 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 (28) 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

$$\frac{E\left[f\left(\tilde{\gamma}(\theta,a)\right) \frac{\partial \tilde{\gamma}(\theta,a)}{\partial \theta'} d\theta\right]}{E\left\{\left[1 - F\left(\tilde{\gamma}(\theta,a)\right)\right]\right\}} - \frac{f\left(\tilde{\gamma}(\theta,a)\right)}{1 - F\left(\tilde{\gamma}(\theta,a)\right)} \frac{\partial \tilde{\gamma}(\theta,a)}{\partial \theta'} d\theta. \tag{29}$$

The first term in (29) is a positive constant. Since f log concave implies that $f/(1-F)$ is monotone increasing, the second term in (29) is monotone increasing whenever $[\partial \tilde{\gamma}(\theta,a)/\partial \theta']d\theta > 0$. Therefore, there is an a^* such that (29) is positive on $a < a^*$

and negative on $a > a^*$.[31] □

Proof of Proposition 8. Set $\theta = \Delta$. If $\tilde{\gamma}(\Delta, a) = \Delta + \gamma(a)$, then $(\partial\tilde{\gamma}/\partial\Delta)d\Delta = d\Delta$ is constant. If $\tilde{\gamma}(\Delta, a) = \Delta\gamma(a)$ then $(\partial\tilde{\gamma}/\partial\Delta)d\Delta = \gamma(a)d\Delta$ is increasing. If $\tilde{\gamma}(\Delta, a) = \gamma(\Delta+a)$ then, since γ is concave, $(\partial\tilde{\gamma}/\partial\Delta)d\Delta = \gamma'(\Delta + a)d\Delta$ is decreasing. In either case, the proof follows directly from Lemmas 2, 3 and 4. □

Proof of Proposition 10. 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 $(\gamma''\sigma a + \gamma')\, d\sigma \gtreqless 0$ as $-(\gamma''\sigma a/\gamma')d\sigma \lesseqgtr d\sigma$. Since γ is CRRA, $(\partial\tilde{\gamma}/\partial\theta')d\theta$ is (weakly) monotone increasing or decreasing.

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

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

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

In either case, the proposition follows from Lemmas 2, 3, and 4.

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

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

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 (28) 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

$$\frac{E\left[f\left(\gamma(\Delta + \sigma a)\right)\frac{\partial\gamma(\Delta+\sigma a)}{\partial\theta'}d\theta\right]}{E\left\{\left[1 - F\left(\gamma(\Delta + \sigma a)\right)\right]\right\}} - \frac{f\left(\gamma(\Delta + \sigma a)\right)}{1 - F\left(\gamma(\Delta + \sigma a)\right)}\gamma'(\Delta + \sigma a)(d\Delta + ad\sigma). \tag{30}$$

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,

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

[31] Note that (29) positive for all a implies $[\partial p(a|0)/\partial\theta']d\theta$ is always positive, in contradiction to $\int c_0 = 0$.

is monotone increasing. Hence (30) will be monotone decreasing as soon as $a > -d\Delta/d\sigma$, so there must be an a^* such that (30) 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. $\square$