BOUNDING CAUSAL EFFECTS WITH CONTAMINATED AND CENSORED DATA: REASSESSING THE IMPACT OF EARLY CHILDBEARING ON CHILDREN

by

Charles H. Mullin

Working Paper No. 00-W39

September 2000

DEPARTMENT OF ECONOMICS
VANDERBILT UNIVERSITY
NASHVILLE, TN 37235

www.vanderbilt.edu/econ

Bounding Causal Effects with Contaminated and Censored Data: Reassessing the Impact of Early Childbearing on Children

By

Charles H. Mullin*
Vanderbilt University

September 2000

Abstract

Empirical researchers commonly use instrumental variable (IV) assumptions to identify treatment effects. However, the credibility of these assumptions are often questionable. In this paper, we consider what can be learned when the assumptions necessary for point identification are violated in two specific ways. First, we allow the data to be contaminated, meaning that the exclusion restrictions of the IV estimator hold for only a fraction of the sample. Second, we allow for the data to be censored. After relaxing these assumptions point identification is no longer feasible, but we are able to construct sharp bounds of the treatment effect. In particular, we show that miscarriages can be seen as generating a contaminated and censored sample with which to analyze the impact of a mother's age at conception on the subsequent development of her child. Utilizing the aforementioned bounds, we are able to demonstrate that for non-black children, a delay in their mother's age at first birth is detrimental to their well being.

* Department of Economics, Vanderbilt University, Nashville, TN 37235. Email comments to: charles.mullin@vanderbilt.edu. This research was supported by a grant from the National Institute of Child Health and Human Development and by a predoctoral fellowship from the National Institutes of Health.

1. Introduction

This paper addresses the identification of treatment effects when data are generated by a less than ideal randomized experiment. This issue has been a long standing problem in economics and the social sciences in general. The difficulty in identifying these treatment effects from observational or experimental data is that the data are inherently incomplete. Although we observe the outcome for each individual resulting from the treatment they received, we do not observe the counterfactual outcomes that they would have attained under alternative treatments. In lieu of observing the counterfactual outcomes, researchers typically invoke sufficient assumptions to recover their expected value.

Over the years various sets of alternative identifying assumptions have been proposed, one of the most common of which is instrumental variables. Under this approach, the researcher finds an instrumental variable (IV) that is correlated with the choice of treatment, but uncorrelated with treatment outcomes. Although this technique has its advantages, often it replaces the suspect assumption of random treatment selection with alternative exclusion restrictions that are equally questionable.[1] This paper follows in the spirit of Manski (1989, 1990) by asking what can be learned when some of these exclusion restrictions are relaxed. In general, without invoking further assumptions, point identification will no longer be feasible, but sharp bounds of the treatment effect can be derived.

In particular, we build on the work of Hotz, Mullin and Sanders (1997). These authors relax the traditional IV exclusion restriction by allowing it to be violated for a fraction of the population. They go on to demonstrate that such a "contaminated" instrumental variable can be viewed as generating a "contaminated" sample in the terminology of Horowitz and Manski (1995). Then, applying Horowitz and Manski's results on forming bounds on moments of

[1] See Heckman (1990, 1996), Imbens and Angrist (1994), Angrist, Imbens and Rubin (1996) and Manski and Nagin (1998).

random variables based upon contaminated samples, they provide sharp bounds of the treatment effect. This contribution is significant because it covers a wide variety of evaluation contexts. For example, many controlled experiments fall into this framework due to the inability of experiments to eliminate various forms of non-compliance.[2] This paper extends their estimation technique to construct sharp bounds in the presence of endogenously censored outcomes within a contaminated sample. This extension is important since it covers a large subset of the evaluation contexts with contaminated data. For example, many experiments suffer from attrition which can be viewed as a form of censored outcomes.

We apply our bounds to examine the impact of a mother's age at conception on the subsequent development of her child. A controlled experiment in which age at first birth is randomly assigned is clearly not feasible, but Hotz, McElroy and Sanders (1997) observe that random miscarriages closely mimic this experiment and can be used to construct an IV estimator. Unfortunately, not all miscarriages are random. The presence of behaviorally induced or non-random miscarriages potentially biases their IV estimator. In the terminology of Horowitz and Manski, the miscarriage sample is a contaminated sample of the random miscarriage population, since it is mixed with non-random miscarriages. Hotz, Mullin and Sanders (1997) construct sharp bounds of the causal effect of teenage childbearing on maternal outcomes from this contaminated sample.

We extend this work to the analysis of child outcomes. In addition to the problems of selection-bias and the inability to differentiate between random and non-random miscarriages present in the analysis of maternal outcomes, the analysis of child outcomes presents the additional difficulty of censored data. In particular, some women never have children, while others who intend to have children have not had their first child within the sampling frame of the

[2] See Hotz and Sanders (1996).

data. This censoring problem does not occur for the children of teenage mothers, but it does complicate the estimation of their counterfactual outcomes. However, we still can construct sharp bounds of the causal effect of early childbearing. Furthermore, we are able to use these bounds to test the assumptions underlying traditional OLS and IV estimators.

We find that the deleterious effects of teenage childbearing for non-black women not only disappear when estimators account for self-selection into early motherhood, but that a delay in childbearing for this sub-population of women is actually harmful to their children. This result adds to a growing literature which indicates that despite the strong negative correlation between a woman's age at first birth and the subsequent attainment of her child, a minor having a child is purely a signal of a woman already in need of assistance, not the cause of her or her child's troubles. For example, Geronimus and Korenman (1992) and Hoffman, Foster, and Furstenberg (1993) exploit differences in sisters' childbearing to estimate family fixed-effect models. They find that the estimated deleterious effects of early motherhood on maternal outcomes are about one half the size inferred from cross-sectional estimates on the same data. Ribar (1994) estimates a bivariate probit model of the joint decision to complete high school and bear a child as a teenager with three alternative exclusion restrictions to attain identification. Under all specifications, the causal-effects of early childbearing on educational attainment are small and more often than not, positive. Angrist and Evans (1996), Bronars and Grogger (1994), and Hotz, McElroy, and Sanders (1997) employ natural experiments (legislation, twins and miscarriages, respectively) and attain IV estimates indicating few negative effects and some positive effects of teenage motherhood for those who choose it. Hotz, McElroy, and Sanders' results are confirmed by the more robust bounds constructed in Hotz, Mullin and Sanders (1997).

Turning to the effects on children, the literature is smaller, but has a similar flavor. In a review of the available statistical methods of identifying the effects of teenage childbearing, Rosenzweig and Wolpin (1995) conclude that teenage mothers have lower-gestation births, but at birthweights *greater* than older mothers. Geronimus, Korenman and Hillemeier (1994) use differences in sisters' childbearing to estimate a family fixed-effect model of the consequences of teenage childbearing on children and conclude that their estimates offer little support for a detrimental effect on children's achievement. Moreover, the only statistically significant estimates they attain, suggest a positive impact of teenage childbearing. Finally, our estimates provide a more robust verification of the conclusion that the children of teenage mothers are not harmed and may benefit from their mothers' choice to bear children. Furthermore, the results tell the same story over a diverse range of measures, including birth weight, behavioral problems and cognitive attainment.

2. Pregnancy Resolution and the Evaluation Problem

In this section, we characterize the problem of identifying the casual effect of teenage women bearing children. The exposition maintains the notation and closely parallels that of Hotz, Mullin and Sanders (1997). Those familiar with this previous work may wish to skip to the next section which addresses the additional complications involved in the analysis of child outcomes as opposed to maternal outcomes.

A pregnancy can be resolved in one of four ways: it can end in a birth (*B*); an induced abortion (*A*); or one of two types of miscarriages, non-random (*NR*) and random ones (*RM*). The non-random miscarriage category includes those which are induced by such behaviors as smoking and drinking. From a choice theoretic perspective, the first three ways of resolving a pregnancy can be viewed as choices women make, either directly or indirectly. As such, these choices, and their determinants, may be correlated with the outcome variables of interest. Let *D* indicate the way in which a women chooses to resolve her pregnancy, where $D = B, A$ or *NR*.

In contrast to the first three methods of resolving a pregnancy, random miscarriages represent events that are exogenously imposed on pregnant women. Let Z^* indicate the occurrence of a random miscarriage, where $Z^* = RM$ or ~*RM*. A key feature of random miscarriages is that they prevent women from choosing how their pregnancies are resolved.[3] So, when a random miscarriage occurs, a pregnant woman's preference for how her pregnancy would be resolved may not be revealed. Nonetheless, it is useful to characterize what a woman's choice would have been if she had not experienced a random miscarriage. Let D^L indicate a woman's *latent* pregnancy resolution choice, where D^L = *B*, *A* or *NR*. Her latent choice status is defined as how her pregnancy would be resolved in the absence of a random miscarriage.

Finally, let *Y* denote an outcome of interest at age *t*, where we forego indexing by *t* to avoid excessive notation. Following the framework of the literature on the identification of treatment effects, we define Y_1 to be the outcome that would result if a woman's first birth occurs when she is a teen and Y_0 to be the outcome that would result if her childbearing is delayed. Our interest is in the effect of a woman having a birth as a teen verses delaying it on subsequent outcomes for the population of women who first gave birth as teens. More precisely, we are interested in identifying

$$\alpha(X) \equiv E\{Y_1 - Y_0 | D = B, X\} \tag{2.1}$$

where α may vary with *X*, a vector of exogenous characteristics. From this point onward, the conditioning on *X* is left implicit to avoid excessive notation.

The casual effect in (2.1) characterizes how different a teen mother's subsequent *Y*s would be if she postponed the birth of her first child. This casual effect is analogous to the mean

[3] Here, we assume that random miscarriages occur early in a woman's pregnancy, before she has the opportunity to abort the fetus or induce a non-random miscarriage. Below, we discuss the implications of some random miscarriages being preempted by abortions or non-random miscarriages for the identification of the casual effect of early childbearing.

effect of *treatment on the treated* in the evaluation literature. As discussed in Heckman and Robb (1985), α is a non-standard parameter from the vantage point of structural modeling in econometrics. Structural modeling is typically used to identify the average effect of a potentially endogenous event (a teen birth) on Y for everyone in the population, rather than just those who are observed to have chosen the event (teen mothers). We focus on the casual effect for teen mothers for two reasons. First, this effect is more readily identified from available data. Second, policies that seek to reduce the rate of teenage childbearing will likely target women who, under the status quo, would become teenage mothers. Knowledge of α is sufficient to assess the potential consequences of eliminating teenage childbearing for these women.

The fundamental problem in identifying α is that while we observe the outcomes of teenage mothers, Y_1, we can never directly observe the counterfactual outcome, Y_0, for these women, i.e. what the outcome would have been for these women in the absence of a teen birth. At issue is what comparison group to use to obtain data on Y_0 and its expectation. Often the outcomes for women who chose not to have teen births are used. In general, using the outcomes for the latter group to measure the counterfactual outcomes for the former will not identify α. Rather, it will identify

$$E\left\{Y_1 \middle| D=B\right\}-E\left\{Y_0 \middle| D\neq B\right\}=\alpha+\left[E\left\{Y_0 \middle| D=B\right\}-E\left\{Y_0 \middle| D\neq B\right\}\right] \tag{2.2}$$

where the expression in brackets is the *selection-bias* term -- i.e. the mean difference in outcomes that would have existed between women who had births and women who did not if both had delayed their childbearing.

In many evaluation problems the selection-bias term is eliminated by conducting a randomized experiment. In this case, the appropriate randomized experiment would involve randomly allowing some latent-birth type women to have children as teenagers and preventing others. Such an experiment is clearly unethical and, thus, not feasible to implement. This study considers three alternative estimation techniques to address the selection-bias term: Ordinary Least Squares (OLS), Instrumental Variables (IV), and Horowitz-Manski (HM) bounds.

2.1 Ordinary Least Squares (OLS)

The standard OLS estimator attempts to correct for the selection-bias term by conditioning on enough covariates such that it is zero, i.e. condition on a set of covariates *X* such that $E\{Y_0 | D = B, X\} = E\{Y_0 | D \neq B, X\}$. However, if the observable covariates do not control for all of the self-selection into early childbearing, then the estimator remains biased.

We consider two OLS estimators. The first, conditions on all exogenous variables available in the data. A complete list of these variables is provided in the discussion of the data. Although we do not believe the available exogenous variables are sufficient to eliminate the selection bias, this estimator is included to facilitate comparisons between our results and previous work.

Second, we restrict the sample to women who became pregnant as teens. If the covariates in the traditional regression controlled for the selection bias, then this restricted regression will be unbiased as well. However, the standard errors will be greater due to the reduction in sample size. On the other hand, if the covariates failed to correct for all of the self-selection into teenage childbearing and teenage mothers are more similar to teens who become pregnant and fail to have a birth than teens who do not conceive, then the restricted regression could suffer from less selection-bias. Since this restricted sample is more alike on observable characteristics, it is likely that they are more similar on unobservable characteristics. Furthermore, this second OLS estimator provides a Hausman like test of the assumptions underlying the first OLS estimator.

2.2 Miscarriages as an Instrumental Variable

As discussed earlier, a controlled experiment in which latent-birth type women are randomly assigned childbearing status is not feasible. However, a naturally-occurring experiment, in which teenagers experience random miscarriages, would seem to mimic this controlled experiment and provide ideal data for identifying α. Three conditions are necessary for random miscarriages to identify α.

Condition 1. The occurrence of a random miscarriage precludes the occurrence of a birth, while the absence of a random miscarriage ensures the occurrence of a birth for latent-birth type women.

Condition 2. Random miscarriages do not affect outcomes of latent-birth type women, i.e. $E\left\{Y_0 \middle| D^L = B, Z^* = RM\right\} = E\left\{Y_0 \middle| D^L = B\right\}$.

Condition 3. Latent-birth type women and their miscarriage status are observable.

If these conditions are met, α is identified and can be estimated as the difference between the mean outcomes of latent-birth type women who have births and those who have random miscarriages.

While Condition 1 is not controversial, the validity of the remaining two conditions are. A random miscarriage may cause a behavioral response, such as depression, if the child was desired, or elation, if it was unwanted. This effect of the randomizing event is labeled the "Hawthorne Effect" in randomized experiments and Heckman and Smith (1995) refer to this phenomena as "randomization bias." To satisfy Condition 3 the researcher must know the identities of latent-birth type women. For those that bear children, this identification is easy. However, among the miscarriage population, it is impossible to disentangle the latent-birth type women from latent-abortion and non-random miscarriage type women without further assumptions. In particular, when Condition 3 fails to hold, the following two additional conditions are sufficient to construct a consistent instrumental variables estimator:

Condition 4. Random miscarriages do not affect outcomes of latent-abortion type women, i.e. $E\left\{Y_0 \middle| D^L = A, Z^* = RM\right\} = E\left\{Y_0 \middle| D^L = A\right\}$.

Condition 5. (a) Random miscarriages do not affect outcomes of latent-non-random-miscarriage type women, i.e. $E\left\{Y_0 \middle| D^L = NR, Z^* = RM\right\} = E\left\{Y_0 \middle| D^L = NR\right\}$; and (b) the expected outcome for women experiencing non-random miscarriages can be expressed as a weighted average of the expected outcome for latent-birth and abortion type women, i.e. $E\left\{Y_0 \middle| D^L = NR\right\} = \left(P_B/[1-P_{NR}]\right)E\left\{Y_0 \middle| D^L = B\right\} + \left(P_A/[1-P_{NR}]\right)E\left\{Y_0 \middle| D^L = A\right\}$.

Condition 4 provides an avenue through which the population of women who abort their pregnancies can be used to purge the miscarriage population of latent-abortion type women. Condition 5 enables one to extract the non-random-miscarriage population from the miscarriage population. Although this final condition is strong, Hotz, McElroy and Sanders (1997) point out that if the epidemiological literature is correct, then conditioning on smoking and drinking during pregnancy is sufficient for Condition 5 to hold. See Hotz, Mullin and Sanders (1997) for the details on how to implement the appropriate IV estimator under Conditions 1, 2, 4 and 5.

2.3 Horowitz-Manski Bounds

In this section we develop bounds of the casual effect of teenage childbearing which employ weaker assumptions than those needed for point identification. The general intuition is that miscarriages can be thought of as a "contaminated" instrumental variable. As such, the estimate provided by it can be viewed of as a combination of an unbiased estimate from random miscarriages plus bias from non-random miscarriages. Horowitz and Manski (1995) provide the necessary tools to construct sharp bounds on this potential bias.

Initially, assume that only Conditions 1 and 2 hold. So, we are unable to distinguish between random and non-random miscarriages, but we do observe when miscarriages occur. Let Z indicate the occurrence of a miscarriage, either random or otherwise, where $Z = M$ or $\sim M$. Therefore, the distribution of outcomes for women experiencing a miscarriage is identified and can be expressed as a mixture of the outcomes for the women of the three latent types.

$$E\left\{Y_0 \middle| Z=M\right\}=\left[\frac{P_{RM}}{P_M}\right]\left[\begin{array}{l}P_B \cdot E\left\{Y_0 \middle| D^L=B, Z^*=RM\right\}+P_A \cdot E\left\{Y_0 \middle| D^L=A, Z^*=RM\right\} \\ +P_{NR} \cdot E\left\{Y_0 \middle| D^L=NR, Z^*=RM\right\}\end{array}\right] \\ +\left[\frac{\left(1-P_{RM}\right) P_{NR}}{P_M}\right] E\left\{Y_0 \middle| D^L=NR, Z^*=\sim RM\right\} \tag{2.3}$$

where $P_{RM} \equiv \Pr\left(Z^*=RM\right)$ is the probability of a random miscarriage; $P_j \equiv \Pr\left(D^L=j\right)$ is the probability of the j^{th} pregnancy resolution being a woman's latent choice, j = B, A and NR for which $P_B+P_A+P_{NR}=1$; and P_M is the probability of a miscarriage, where $P_M=P_{RM}+\left(1-P_{RM}\right)P_{NR}$. While in the absence of Condition 3 we are unable to point identify α, we are able to place a bound on it. Recall from (2.1) that $\alpha \equiv E\left\{Y_1 \middle| D=B\right\}-E\left\{Y_0 \middle| D^L=B\right\}$. Since $E\left\{Y_1 \middle| D=B\right\}$ is identified by women who had a birth as a teen, placing a bound on α rests entirely on forming a bound for $E\left\{Y_0 \middle| D^L=B\right\}$. Rearranging and consolidating terms in (2.3) yields

$$E\left\{Y_0 \middle| D^L=B\right\}=\left(\frac{1}{\lambda^*}\right)\left[E\left\{Y_0 \middle| Z=M\right\}-\left(1-\lambda^*\right)E\left\{Y_0 \middle| D^L \neq B, Z=M\right\}\right] \tag{2.4}$$

where λ^* $\left(\equiv P_B P_{RM}/P_M\right)$ is the proportion of miscarriages that occur randomly to latent-birth type women. While $E\left\{Y_0 \middle| Z=M\right\}$ is identified, the second term in brackets on the right-hand side of equation (2.4), $E\left\{Y_0 \middle| D^L \neq B, Z=M\right\}$, is not identified. So, neither $E\left\{Y_0 \middle| D^L=B\right\}$ nor α are identified.

But the fact that equation (2.4) holds implies that there is a tight set of bounds on $E\left\{Y_0 \middle| D^L=B\right\}$. This follows from results in Horowitz and Manski (1995) on the identification of bounds for moments of random variables using data from contaminated samples so long as one knows (or can estimate) either λ^* or a lower bound on this proportion. Let λ denote this lower bound. Below, we discuss how one estimates λ from epidemiological studies of random miscarriages and vital statistics data. For now, we assume that λ is known.

The intuition for the bounds is as follows: Suppose that half of the miscarriage population are latent-birth type women. Then, the population of non-latent-birth type women cannot have a distribution below that of the bottom half of the miscarriage population. So, the mean outcome for the bottom half of the miscarriage population is a lower bound on the average outcome for non-latent-birth type women. In general, the fraction $(1-\lambda)$ of miscarriages are non-latent-birth type women and the expected value of the $(1-\lambda)$-quantile of the distribution of outcomes among the miscarriage population is a lower bound on the average outcome for non-latent-birth type women. Notationally, define $Y_{M,1-\lambda}$ to be the $(1-\lambda)$-quantile of the distribution of Y, i.e. $\Pr\left(Y \leq Y_{M,1-\lambda} \middle| Z = M\right) = 1-\lambda$. Then, the greatest lower bound on $E\left\{Y_0 \middle| D^L \neq B, Z = M\right\}$ is given by the truncated mean

$$E\left\{Y_0 \middle| Y \leq Y_{M,1-\lambda}, Z = M\right\}. \tag{2.5}$$

By similar reasoning, the greatest upper bound is

$$E\left\{Y_0 \middle| Y \geq Y_{M,\lambda}, Z = M\right\}. \tag{2.6}$$

It follows from Corollary 4.1 in Horowitz and Manski (1995) that these bounds are sharp.

Using (2.5) and (2.6) as bounds for $E\left\{Y_0 \middle| D^L \neq B, Z = M\right\}$, we define the Horowitz-Manski bounds on α as

$$\alpha \in \left[A_{1L}(\lambda), A_{1U}(\lambda)\right]. \tag{2.7}$$

where

$$A_{1L}(\lambda) \equiv E\left\{Y_1 \middle| D = B\right\} - \frac{1}{\lambda}\left[E\left\{Y_0 \middle| Z = M\right\} - (1-\lambda)E\left\{Y_0 \middle| Y \leq Y_{M,1-\lambda}, Z = M\right\}\right], \tag{2.8}$$

$$A_{1U}(\lambda) \equiv E\left\{Y_1 \middle| D = B\right\} - \frac{1}{\lambda}\left[E\left\{Y_0 \middle| Z = M\right\} - (1-\lambda)E\left\{Y_0 \middle| Y \geq Y_{M,\lambda}, Z = M\right\}\right]. \tag{2.9}$$

Notice that these bounds are defined even if *Y* does not have bounded support.

If in addition to Conditions 1 and 2, we assume that Condition 4 holds, i.e. we assume there is no Hawthorne Effect for latent-abortion type women, then the bounds can be tightened. Rearranging and consolidating terms in equation (2.3) yields

$$E\left\{Y_0 \middle| D^L = B\right\} = \left(\frac{1}{\lambda^*}\right)\begin{bmatrix} E\left\{Y_0 \middle| Z = M\right\} - \theta^* E\left\{Y_0 \middle| D^L = A\right\} \\ -\left(1-\lambda^*-\theta^*\right)E\left\{Y_0 \middle| D^L = NR, Z = M\right\} \end{bmatrix} \tag{2.10}$$

where θ^* $\left(P_A P_{RM} / P_M\right)$ is the proportion of miscarriages that occurred randomly to latent-abortion types. Using analogous bounds to (2.5) and (2.6) for $E\left\{Y_0 \middle| D^L = NR, Z = M\right\}$ and letting θ be a lower bound for θ^*, the modified Horowitz-Manski bounds are

$$\alpha \in \left[A_{2L}(\lambda,\theta), A_{2U}(\lambda,\theta)\right]. \tag{2.11}$$

where

$$A_{2L}(\lambda,\theta) \equiv E\left\{Y_1 \middle| D = B\right\} - \frac{1}{\lambda}\begin{bmatrix} E\left\{Y_0 \middle| Z = M\right\} - \theta E\left\{Y_0 \middle| D = A\right\} \\ -\left(1-\lambda-\theta\right)E\left\{Y_0 \middle| Y \le Y_{M,1-\lambda-\theta}, Z = M\right\} \end{bmatrix}, \tag{2.12}$$

$$A_{2U}(\lambda,\theta) \equiv E\left\{Y_1 \middle| D = B\right\} - \frac{1}{\lambda}\begin{bmatrix} E\left\{Y_0 \middle| Z = M\right\} - \theta E\left\{Y_0 \middle| D = A\right\} \\ -\left(1-\lambda-\theta\right)E\left\{Y_0 \middle| Y \ge Y_{M,\lambda+\theta}, Z = M\right\} \end{bmatrix}. \tag{2.13}$$

Whenever $E\{Y_0|D=A\} > E\{Y_0|Y_{M,1-\lambda-\theta} \le Y \le Y_{M,1-\lambda}, Z=M\}$, $A_{2L}(\lambda,\theta) > A_{1L}(\lambda)$. Similarly, whenever $E\{Y_0|D=A\} < E\{Y_0|Y_{M,\lambda} \le Y \le Y_{M,\lambda+\theta}, Z=M\}$, $A_{2U}(\lambda,\theta) < A_{1U}(\lambda)$. These bounds are sharp given the maintained conditions.[4]

3. Censored Outcomes

In addition to the problems of selection-bias and the unobservable nature of non-random miscarriages, the analysis of child outcomes presents the following additional difficulty: Some women never have children, while others who intend to have children have not had their first child within the sampling frame of the data. This censoring problem does not occur for the sample of teenage mothers' children, but it does complicate the estimation of counterfactual outcomes.

We start by dealing with the easier case of estimating the expected outcome of latent-abortion type women. An implication of Condition 4 (no Hawthorne Effect for latent-abortion type women) is that these women's desire to eventually have a child is independent of whether they experience a random miscarriage or have an abortion. Therefore, an equal proportion of the latent-abortion type women in both the miscarriage and abortion populations will remain childless. A possible exception to this conclusion is that miscarriages and abortions can lead to future fertility problems. However, given that we know of no evidence indicating otherwise, we assume that infertility caused by the abortion or miscarriage that terminated the first pregnancy occurs randomly. Thus, such infertility randomly censors child outcomes, at potentially different

[4] A proof can be provided from the authors upon request. Also, the bounds could be tightened further if one were to assume that the distribution of outcomes for women intending to have an abortion is unaffected by a random miscarriage. Under this assumption, we could extract the latent-abortion outcome distribution from the miscarriage distribution leaving the distribution of outcomes for those that had miscarriages, but did not intend to have an abortion. Then, the first type of Horowitz-Manski bounds can be computed for this resultant distribution (the resultant distribution is still contaminated with non-random miscarriages). However, in this paper, these bounds are not informative conditional on the other estimators because of the imprecision of the estimates of the density functions.

rates, in both the miscarriage and abortion populations. But random censoring does not affect the identification of any of the estimators.[5]

Turning to latent-birth type women in the miscarriage population, we consider three approaches to implementing the Horowitz-Manski bounds. First, we assume that children are randomly distributed to latent-birth type women in the miscarriage population.

Condition 6. Latent-birth type mothers in the miscarriage population are a random sample of latent-birth type women in the miscarriage population.

When Condition 6 holds, child outcomes for latent-birth type women have been randomly censored from the miscarriage population. As with latent-abortion type women, random censoring does not affect the identification of any of the estimators. However, it does change the proportion of children in the miscarriage population belonging to latent-birth type women. In particular, let δ^* be the proportion of latent-birth type women in the miscarriage population with children and δ be a lower bound for δ^*. Also, let η^* be the proportion of women in the miscarriage population with children and η be the proportion of such mothers in the sample. Then, in the Horowitz-Manski bounds given in (2.7) and (2.11), replace λ, the proportion of latent-birth type women, with $\lambda' \equiv \lambda\delta/\eta$, the proportion of children with latent-birth type mothers. Similarly, replace θ, the proportion of latent-abortion type women, with $\theta' \equiv \theta\gamma/\eta$, the proportion of children with latent-abortion type mothers, where γ is the proportion of latent-abortion type women in the miscarriage population with children.[6]

Although Condition 6 is a strong assumption, there is nothing in the data to cast doubt on it. In analysis not shown here, the length of delay from miscarriage to birth has no statistically significant impact on children's outcomes. In a linear specification, the point estimate indicates a

---

[5] If the rates of random censoring differ between the miscarriage and abortion populations, then the weights of the IV estimator need to be adjusted. In general, the weights are constructed such that the weighted ratio of women having abortions to those having children needs is constant across the miscarriage and non-miscarriage samples.

[6] Similar re-weighting of the observation needs to done for the IV estimator (see footnote 5).

slight decrease in children's attainment as the length of delay increases, but the size of the estimate is statistically insignificant and economically inconsequential. Non-parametric and alternative parametric specifications were unable to determine any pattern in child outcomes as a function of the length of delay between the first miscarriage and the birth of the child. However, the power of all of these tests are very low, since there are only 18 births in the miscarriage sample that occur with a delay in excess of three years.

The second approach takes a more agnostic viewpoint as to the source of censoring of child outcomes for latent-birth type women in the miscarriage population. When the outcome in question is bounded, the censored observations can be replaced by the minimum or maximum attainable value as the construction of the lower or upper bound dictates. In other words, the miscarriage population provides bounds on a fraction δ of the latent-birth types (those with children in the counterfactual state), while the remaining latent-birth types are bounded by the lower and upper bounds on Y, Y_L and Y_U, respectively. Therefore, after correcting for censoring, the A_1 bounds become

$$\alpha \in \left[A_{3L}(\lambda'), A_{3U}(\lambda') \right]. \tag{3.1}$$

where

$$\begin{aligned} A_{3L}(\lambda') \equiv {} & E\{Y_1 | D = B\} \\ & -\delta \left[\frac{1}{\lambda'} \left[E\{Y_0 | Z = M\} - (1-\lambda') E\{Y_0 | Y \leq Y_{M,1-\lambda'}, Z = M\} \right] \right] - (1-\delta) Y_U \end{aligned}, \tag{3.2}$$

$$\begin{aligned} A_{3U}(\lambda') \equiv {} & E\{Y_1 | D = B\} \\ & -\delta \left[\frac{1}{\lambda'} \left[E\{Y_0 | Z = M\} - (1-\lambda') E\{Y_0 | Y \geq Y_{M,\lambda'}, Z = M\} \right] \right] - (1-\delta) Y_L \end{aligned}. \tag{3.3}$$

Note that assuming the average censored outcome is bounded by the ε and κ percentiles of the population simultaneously tightens the bounds and makes them informative for unbounded outcomes. Computationally, replace the lower and upper bounds of Y with the ε and κ quantiles of the distribution of Y in equations (3.2) and (3.3). Given that ε and κ are choice parameters, the A_3 bounds are feasible for unbounded outcomes under arbitrarily weak additional assumptions since ε and κ can be set arbitrarily close to zero and one hundred. Finally, the A_2 bounds can be adjusted for censored outcomes in a similar manner to yield what we shall refer to as the A_4 bounds. In practice, we experiment with values for ε and κ to determine how robust the IV point estimates and the previous bounds are to nonrandom censoring of outcomes.

The third approach attempts to address the following two shortcomings of the A_3 and A_4 bounds: (1) When the outcome of interest is unbounded, these bounds are uninformative; and (2) in some situations, it is unclear whether the minimum and maximum values are appropriate bounds for the counterfactual outcome. To illustrate the latter case, consider the problem analyzed here of teenage childbearing. How does one measure the impact on a woman's first born child of a delay in childbearing when such a delay causes the woman not to have children?

As an alternative to attempting to answer this question, it is possible to reduce the treatment group (teenage mothers) to a set of women that matches those women with children in the miscarriage population. It is important to note that such a change alters the treatment effect being estimated. Instead of estimating α, the effect of treatment on the treated, we are estimating $\tilde{\alpha}$, the effect of treatment on a sub-sample of the treated. In particular, the effect of treatment on those women who would go on to have children at a later date if they were prohibited from having a child as a teenager. Note that the sub-sample of women being omitted are those women who would only have children as teenagers. The absence of a child in the counterfactual state makes it unclear how to incorporate these women into an analysis of child well being. So, dropping them may be preferable.

The bounds for the counterfactual outcome for the children of teenage mothers are identical to those constructed under Condition 6. The change is that we no longer have point identification of the expected outcome for the children of the appropriate sub-population of teenage mothers. However, using the results in section 2, the expected outcome for the children of this sub-population of teenage mothers is bounded below by

$$E\left\{Y_0 \middle| Y \leq Y_{B,1-\delta}, D = B\right\} \tag{3.4}$$

and above by

$$E\left\{Y_0 \middle| Y \geq Y_{B,\delta}, D = B\right\}. \tag{3.5}$$

Substituting these expression into the A_1 bounds given in equation (2.7) yields

$$\tilde{\alpha} \in \left[A_{5L}(\lambda'), A_{5U}(\lambda')\right]. \tag{3.6}$$

where

$$A_{5L}(\lambda') \equiv E\left\{Y_0 \middle| Y \leq Y_{B,1-\delta}, D = B\right\} - \frac{1}{\lambda'}\left[E\left\{Y_0 \middle| Z = M\right\} - (1-\lambda')E\left\{Y_0 \middle| Y \leq Y_{M,1-\lambda'}, Z = M\right\}\right], \tag{3.7}$$

$$A_{5U}(\lambda') \equiv E\left\{Y_0 \middle| Y \geq Y_{B,\delta}, D = B\right\} - \frac{1}{\lambda'}\left[E\left\{Y_0 \middle| Z = M\right\} - (1-\lambda')E\left\{Y_0 \middle| Y \geq Y_{M,\lambda'}, Z = M\right\}\right]. \tag{3.8}$$

Similarly, the A_2 bounds given in equation (2.11) become

$$\tilde{\alpha} \in \left[A_{6L}(\lambda', \theta'), A_{6U}(\lambda', \theta')\right]. \tag{3.9}$$

where

$$A_{6L}\left(\lambda',\theta'\right) \equiv E\left\{Y_0 \middle| Y \le Y_{B,\delta}, D = B\right\} - \frac{1}{\lambda'}\begin{bmatrix} E\left\{Y_0 \middle| Z = M\right\} - \theta' E\left\{Y_0 \middle| D = A\right\} \\ -\left(1-\lambda'-\theta'\right) E\left\{Y_0 \middle| Y \le Y_{M,1-\lambda'-\theta'}, Z = M\right\} \end{bmatrix}, \qquad (3.10)$$

$$A_{6U}\left(\lambda',\theta'\right) \equiv E\left\{Y_0 \middle| Y \ge Y_{B,1-\delta}, D = B\right\} - \frac{1}{\lambda'}\begin{bmatrix} E\left\{Y_0 \middle| Z = M\right\} - \theta' E\left\{Y_0 \middle| D = A\right\} \\ -\left(1-\lambda'-\theta'\right) E\left\{Y_0 \middle| Y \ge Y_{M,\lambda'+\theta'}, Z = M\right\} \end{bmatrix}. \qquad (3.11)$$

4. Estimating λ', θ' and δ

In constructing the bounds in Section 3 it was assumed that λ' and θ', lower bounds on the proportion of children in the miscarriage population with latent-birth and abortion type mothers, as well as δ, an upper bound on the proportion of latent-birth type women in the miscarriage population who have not yet had a child, were all known. Clearly, they are not known and must be estimated. The first step is to estimate λ and θ, lower bounds on the proportion of latent-birth and abortion type mothers in the miscarriage population, respectively.

We divide this estimation problem into two pieces. First, we estimate a lower bound on the fraction of miscarriages that occur randomly. Second, conditional on being a random miscarriage, we estimate the proportion of miscarriages that are of each type. Throughout the analysis we maintain the assumption that non-random miscarriage are reported less frequently than random miscarriages.

Assumption 1. The rate of underreporting by non-random miscarriage types is at least as great as underreporting in the entire population.

Assumption 1 implies that the proportion of miscarriages that are random is greater in our sample than the population as a whole. To date, there is substantial evidence that smoking or drinking during pregnancy, or having an intrauterine device in place at the time of conception, raises the rate of random miscarriages. However, according to Kline, Stein and Susser (1989) the medical evidence on an association between other behavioral factors and miscarriages remains less conclusive.[7] So, we presume that a miscarriage is random if it occurs to a woman who never smoked cigarettes nor drank alcohol during her pregnancy.[8] For women who smoked cigarettes, we assume they smoked at least 15 a day.[9] Kline, Stein and Susser (1989) finds that these women are 60 percent more likely to experience a miscarriage. Similarly, for women who drank alcohol, we assume they drank one to two drinks per day during their pregnancy. Harlap, Shiono and Ramcharan (1980) finds that consumption at such levels leads to no more than a 100 percent increase in miscarriages. These factors lead to the following percentages of miscarriages being random: 62.5 percent for women who smoked, 50 percent for those who drank and 39.5 percent for those who did both.[10] Given these rates and utilizing the data in the NLSY on respondent's smoking and drinking behavior, we estimate that 87 percent of the miscarriages in our sample were random.[11]

Now, we turn to dividing the random miscarriages into the proportion belonging to each of the three latent types. Initially, we determine an upper bound on the fraction of latent-non-random miscarriages in the random miscarriage group. Given Assumption 1, non-random miscarriage types compose no more than three percent of the random miscarriage group.[12] To be conservative, we set the percentage of latent-non-random miscarriage types in the random

[7] Kline, Stein and Susser (1989) note that other factors (nutrition, cocaine use, etc.) affect gestational age, birth weight, and infant mortality, but there is no evidence that they affect the incidence of miscarriage.

[8] We ignore the effect of an intrauterine device for two reasons: (1) It is uncommon among teen women; and (2) our data has no information on contraceptives used while women were teens.

[9] We have no reliable data on the quantity of cigarettes or alcohol consumed. So, we assume that all women who smoked or drank did so in large quantities, guaranteeing a lower bound on the fraction of random miscarriages.

[10] We obtained the 39.5 percent figure by assuming that the effects of smoking and drinking are independent.

[11] We have data on smoking and drinking during the year of a women's pregnancy. To be conservative, we assume that if a woman smoked or drank in that year, then she smoked or drank during her pregnancy.

[12] 13 percent of miscarriages are non-random. Since, 45 percent of pregnancies end in births and 41 percent end in abortions, 14 percent end in a miscarriage. Thus, 1.8 percent of pregnancies end in a non-random miscarriage. Adjusting for the risk exposure of both abortions (1/3) and non-random miscarriages (1/2), the percentage of non-random miscarriages is 1.5.

miscarriage population at its upper bound, three percent, guaranteeing that random miscarriage types are not over-represented.

Before dividing the remaining 97 percent into λ and θ, the percent of latent-birth and abortion types in the miscarriage population, respectively, note that the bounds are continuous in these parameters. This fact suggests the following strategy: Generate two decompositions, one over-representing latent-birth types and the other under-representing latent-birth types. By continuity and monotonicity of the truncated sample means, we know that any decomposition of the random miscarriage group falling between these two decompositions, which includes the true decomposition, will produce bounds that fall between those obtained using these two decompositions. Hence, if the bounds under the two approaches are similar, then we have confidence in their robustness.

To justify the first decomposition, we assume that underreporting is equal across the three latent types.

Assumption 2. While we allow underreporting of pregnancies, underreporting is at a constant rate across latent-types.

When Assumption 2 holds, λ and θ can be estimated from the NLSY sample. The sample proportions of births and abortions among non-miscarried pregnancies, adjusted for the lower risk exposure to a miscarriage of latent-abortion types, can serve as estimates of λ and θ. Using data on the number of births and abortions in our sample, we estimate that latent-birth types compose 75 percent of the miscarriage sample, while latent-abortion types account for 9 percent of the miscarriages. Since Assumption 2 implies that latent-birth types report all of their pregnancies, this decomposition probably over-represents the proportion of latent-birth types. We refer to this decomposition as Decomposition A2.

For the second decomposition, we assume that the proportion of latent-birth and abortion types in our sample of random miscarriages is the same as that of the population as a whole.

Assumption 3. Under-reporting of miscarriages is constant across latent types.

Based on the number of births and abortion in the population as reported by U.S. Vital Statistics and the Alan Guttmacher Institute, we estimate that latent-birth types account for 65 percent of miscarriages, while latent-abortion types compose 20 percent of the miscarriage population. If latent-birth types report their miscarriages more often than the population as a whole, then this decomposition under-represents the fraction of latent-birth types. Furthermore, since this decomposition places fewer latent-birth type women in the miscarriage population, the resultant bounds are wider whenever Condition 4 (no Hawthorne Effect for latent-abortion types) is not imposed. However, when Condition 4 is imposed, the bounds corresponding to the two decomposition are of identical width, although centered at potentially different locations. We refer to this second decomposition as Decomposition A3.

Finally, we need estimates of γ, the fraction of latent-abortion types who remain childless, δ, the proportion of latent-birth types in the miscarriage population who remain childless, and η, the fraction of the miscarriage population without children. The first and third of these can be estimated by the analogous proportions in our sample. These estimates are $\hat{\gamma}$ equal to 21 percent and $\hat{\eta}$ equal to 84 percent. To estimate δ, we assume that latent-non-random miscarriage types are the least likely to have had a child.[13] When this assumption holds,

$$\left[\lambda\delta^{*}+\theta\gamma\right]/(\lambda+\theta)\geq\eta. \qquad (4.1)$$

Rearranging terms, we have a lower bound for δ^{*}, i.e.

$$\delta^{*}\geq\left[\eta(\lambda+\theta)-(\lambda+\theta)\right]/\lambda. \qquad (4.2)$$

[13] Since latent-non-random miscarriage types compose less than 16 percent of the miscarriage sample, the results are robust to this assumption so long as the birth rate for latent-non-random miscarriage types does not greatly exceed that of the population as a whole.

Substituting our estimates of the corresponding population parameters in equation (4.2) yields an estimated lower bound of 85 percent of latent-birth types in the miscarriage population being childless.

Putting these estimates together implies that the fraction of children in the miscarriage population belonging to latent-birth type mothers is 76 and 66 percent under decompositions A2 and A3, respectively. Similarly, the percent of children belonging to latent-abortion type mothers is 9 and 18, respectively. Using these estimates of the fraction of mothers in the miscarriage sample who were latent-birth and abortion types, we construct non-parametric kernel estimates of the bounds. All of the parameter estimates are given in Table I.

5. Data

We use the National Longitudinal Survey of Youths (NLSY). The NLSY is an annual survey of a nationally representative sample of youths who were 14 to 21 years old when the series began in 1979. Despite attrition, this dataset remains representative of its intended population after adjusting by the original 1979 sampling weights (MaCurdy, Mroz and Gritz (1998)). In 1983, a retrospective pregnancy history was administered, and thereafter a pregnancy history was administered approximately every two years. We restrict our attention to women, including those in the oversamples of Blacks and Hispanics, and their first-born children. We exclude later births to avoid complications that arise from differences in child outcomes due to birth order.

We use for our analysis the 978 women in the NLSY who reported a pregnancy before their 18th birthday. Of those pregnancies, 723 resulted in births, 185 terminated in abortions and 70 ended in miscarriages.[14] After adjusting for population weights, these numbers imply that 73 percent of non-miscarried pregnancies are brought to term in our sample. However, the

[14] Jones and Forrest (1992) compare the responses on pregnancy in the self-administered questionnaire to the responses in the open survey and find that it is very rare that a pregnancy is reported with different resolutions in the two mediums.

corresponding number for the population is only 52 percent.[15] So, abortions are almost certainly underreported. The fact that abortions are underreported, leads one to believe that miscarriages will also be underreported. But, it is difficult to determine the degree of this problem since there are no data sources, comparable to the Alan Guttmacher Institute data on abortions, with which to verify their accuracy. If underreporting is correlated with outcomes of interest, then it can bias all of the estimators.

In addition to interviewing the original members of the NLSY, the children of women in the NLSY were interviewed in a child supplement starting in 1986. This supplement contains information about: the child's cognitive attainment, the behavioral problems of the child, and the household in which the child lives. We use the following assessments of children: birth weight, the Peabody Individual Achievement Tests (PIATs), the Peabody Picture Vocabulary Test (PPVT), the behavioral problem indices (BPI), and measures of the child's home environment.[16] The first four of these categories are child outcomes, while the last, home environment, provides indices of inputs into the child. All outcomes except birth weight are measured in percentile scores, normalized such that a higher score is better and, where appropriate, scores have been adjusted for cohort and age at the time of measurement. Finally, all of the child assessments used are listed in Table II with a brief description of the objective of the assessment, the manner in which it is compiled and the ages at which it is taken.

6. Results

In this section we discuss the empirical estimates from the alternative estimators of the casual effect of teenage childbearing for the variables discussed in Section 5. All of the OLS and IV estimates computed control for the following background characteristics of a woman: her AFQT

[15] Population numbers are based on data from U.S. Vital Statistics and the Alan Guttmacher Institute.

[16] When the same measure is available for multiple years, the average of the normalized scores is used in order to reduce the noise associated with these measures.

score (adjusted for age and cohort), whether she lived in an intact family at age 14, whether she lived in a female-headed household at age 14, her family's income in 1978, whether her family was on welfare in 1978, and the education of her mother and father. Finally, all estimates of bounds only control for the quartile of the woman's AFQT score.

Before discussing any particular estimates, it is useful to discuss a couple findings that facilitate the exposition of the results. First, the empirical estimates are qualitatively similar under the two proposed decompositions, Decomposition A2 and A3.[17] So, for the sake of brevity, we focus the discussion on the bounds constructed under Decomposition A2.[18] Second, although at times the bounds can reject the assumptions underlying some of the point estimator, the bounds themselves are always internally consistent.

6.1 Consistency Across Estimators

Tables III and IV present the point estimates for black and non-black women, respectively. If the assumptions underlying all of the estimators are true, then these estimates should be mutually consistent. In particular, when the assumptions underlying the unrestricted OLS estimator hold, the restricted OLS estimator can be viewed as this unrestricted estimator plus noise. Thus, we can perform a Hausman test of the assumptions underlying the unrestricted OLS estimator. Also, if the assumptions underlying a set of bounds are valid, then any unbiased point estimator should fall within those bounds.[19]

[17] Since Decomposition A3 places fewer latent-birth type women in the miscarriage population, the bounds not imposing Condition 4 (no Hawthorne effect for latent-abortion type women) are wider under this decomposition. However, when Condition 4 is imposed, the bounds are of identical width and centered slightly higher than under Decomposition A2.

[18] The estimates of the alternative bounds under Decomposition 3A can be obtained from the authors upon request.

[19] For discussions of testing joint inequality constraints such as we have here see Perlman (1969) and Wolak (1989). In our application, since at most one of the two constraints can be violated, the squared number of standard deviations outside of the bounds is distributed $(1/2)\chi_1^2$ under the null hypothesis that our assumptions hold.

For black women, from a strict statistical perspective the results are very consistent. The Hausman test never comes close to rejecting the assumptions underpinning the unrestricted OLS estimator (the lowest *p*-value across the seven measures is 0.17). Furthermore, only three of the 21 point estimates generated by the OLS and IV estimators fall outside the most restrictive A_2 bounds (see Table V) and none of these violations exceed half a standard deviation. However, despite the inability of the formal tests to detect any violations in the underlying assumptions, for all seven measures analyzed, the restricted OLS estimator produces a greater point estimate than the unrestricted OLS estimator; an outcome consistent with a negative selection bias into early childbearing. If the estimated effects for the seven measures were independent of one another, the Sign test would indicate that all seven estimates increasing with the additional restriction should occur less than one percent of the time. Since all seven estimates are based on the same sample, they are not independent, but this outcome does shed doubt on the assumptions underlying the unrestricted OLS estimator.

Turning to non-black women, the assumptions underlying the OLS estimators are strongly rejected. The Hausman test rejects the unrestricted OLS estimator for four of the seven measures (behavioral problems, PIAT mathematics and both the emotional and cognitive home environment with *p*-values of 0.01, 0.04, 0.08 and 0.01, respectively). Additionally, six of the seven unrestricted point estimates fall below their corresponding A_2 lower bound (see Table V) and four of these violation are statistically significant. As we continue to relax assumptions, the unrestricted OLS estimator continues to lie outside the bounds, but the deviations start to loose statistical significance. Similarly, the restricted OLS estimator is routinely below its corresponding A_2 lower bound with three of the deviations being statistically significant. In contrast to the OLS estimates, the point estimates from the IV estimator always lie within all of the bounds.

In summary, the results indicate that the OLS estimator suffers from negative selectivity bias. There is strong statistical support of this conclusion for both the restricted and unrestricted OLS estimators for non-black women, so neither of these estimators will be discussed further. For black women the data are unable to reject the absence of selectivity bias, but the point estimates are strongly suggestive of its presence.

6.2 Point Estimates

Due to the selectivity bias detected in the OLS estimators, we focus the discussion on the IV point estimates. The estimates tell a very different story for black and non-black women.[20] For the former, although the point estimates are never statistically significant, they give the impression that teenage childbearing has a negative effect on children. With the exception of the PPVT, the five measures of child development all decrease for the children of teens. On the other hand, all five measure increase for the children of non-black teens, including a statistically significant *ten* percentile improvement in behavioral problems and mathematics.[21] Furthermore, the emotional support the child receives improves substantially for non-black children.

The following sub-sections assess the robustness of these findings to the assumptions underlying the IV estimator.

6.3 Allowing for Non-Random Miscarriages

One concern with the IV estimator is that it places strong assumptions on the non-random miscarriage population. To assess the robustness of the previous results to these assumptions, we treat the non-random miscarriage population as a source of contamination in our sample and

[20] Differences across racial lines should not be surprising, since the population of teenage mothers differs greatly by race. In particular, 35 percent of black women in the NLSY conceive their first child before their eighteenth birthday and 97 percent of those children are born out-of-wedlock. The corresponding numbers for non-black women are 13.5 percent bearing children of which 76 percent are out-of-wedlock.

[21] Although the point estimate for birth weight is not statistically significant, the estimate of a 3.99 ounce increase is in line with the preferred fixed-effects instrumental-variables sibling estimator of Rosenzweig and Wolpin (1995) which produces an estimate of 3.36 ounces.

construct the A_2 Horowitz-Manski bounds given in equation (2.11). Table V displays the corresponding estimates.

As one may have suspected given the large standard errors in the point estimates for the children of black women, the bounds are uninformative for this population. The point estimates are relatively inaccurate and even if the sampling variability is ignored, the bounds all contain zero.

In contrast to the results for black children, the bounds for non-black children are informative. In particular, they give an almost identical story as the IV point estimates. For all measures except birth weight, the A_2 bounds are non-negative and the lower bounds are significantly greater than zero for the same three measures that have statistically significant IV point estimates. Furthermore, the lower bounds for those three outcomes are only one to two percentile points below their corresponding IV point estimates. Thus, the conclusion that non-black children of teenage mothers benefit from their mothers' early childbearing is robust to the inclusion of non-random miscarriages.

6.4 Allowing for Non-Random Censoring

A second alternative explanation for our finding an increase in well being for the children of non-black teenage mothers is that within the miscarriage population there is negative selection into future childbearing. In other words, members of the miscarriage population who have not yet born a child are a non-random sample of this population, so child outcomes suffer from non-random censoring. To assess the robustness of our results to the presence of non-random censoring, we turn to the A_4 Horowitz-Manski bounds, which allow for both non-random miscarriages and non-random censoring. Table VI contains the point estimates of these bounds.

The first two columns present the lower and upper bounds for each measure constraining the censored observations to fall between their minimum and maximum attainable values.[22] Under this specification, the bounds contain zero for all for all of the measures except the emotional support provided to the child, which has a lower bound very close to zero. Hence, if we are unwilling to place any constraints on the types of women yet to have children in the miscarriage population, we are unable to sign the causal effects of early childbearing. However, with very mild assumptions on this population of childless women, we can recover our earlier results.

The middle two columns of Table VI contain the bounds when we assume that the expected outcome for the potential children of childless women falls within two standard deviations of the population mean. With this relatively weak assumption, the bounds for behavioral problems, mathematics and emotional support are non-negative and the *p*-values corresponding to the test of whether the lower bounds are greater than zero are 0.17, 0.22 and 0.07, respectively. Strengthening the assumption such that the expected outcome falls within one standard deviation drops these *p*-values to 0.07, 0.10 and 0.03, respectively (see the final two columns of Table VI). In summary, the conclusion that non-black children of teenage mothers benefit from their mothers' early childbearing is robust to severe selectivity in the censoring of child outcomes, but not robust to absolutely all forms of censoring.

6.5 Further Robustness Checks

We compute two additional checks of the robustness of our conclusions. First, we relax Condition 4, the assumption that the eventual children of latent-abortion type women are

[22] Since birth weight is unbounded from above, in the presence of censored data the Horowitz-Manski bounds on the counterfactual birth weight is unbounded from above. So, there is no lower bound on the causal effect of teenage childbearing on children's birth weight.

unaffected by a miscarriage preempting the abortion of their mother's first pregnancy.[23] The A_1 bounds given in equation (2.7) allow both for non-random miscarriages and the relaxation of Condition 4. As seen in Table VII, under these weaker assumptions the bounds for behavioral problems, mathematics and emotional support remain non-negative, but the *p*-values corresponding to the test of whether the lower bounds are greater than zero rise to 0.08, 0.22 and 0.18, respectively. If in addition to relaxing Condition 4 we allow for non-random censoring of child outcomes, then all of the bounds become uninformative.

Second, for those uncomfortable bounding the well being of a yet unborn child, we compute the A_6 bounds given in equation (3.9) which reduce the treatment group (teenage mothers) to a set of women that matches those women with children in the miscarriage population. As discussed during the derivation of these bounds, this change alters the treatment effect being estimated. Instead of estimating the effect of treatment on the treated, we are estimating the effect of treatment on those women who would have already had a children if they were prohibited from having one as a teenager. These bounds are not presented in any table since the results are almost identical to the A_4 bounds which allowed for censored child outcomes, but constrained the degree of censoring to keep the expected censored outcome within two standard deviations of the population mean.[24]

7. Conclusion

We conclude by discussing the substantive implications of what we have learned about the impact of teenage childbearing on several measures of child well being through exploiting the natural experiment which miscarriages provide. First, as repeatedly found in the literature, the

---

[23] When Condition 4 is relaxed, we are unable to use the observed outcomes for the children of women who aborted their first pregnancy to extract these women from the miscarriage population. Instead, latent-abortion type women in the miscarriage population are treated as an additional source of contamination in our sample.

[24] The A_6 bounds are available from the authors upon request.

traditional OLS estimator is rejected by the data. This finding is supported both by the bounds and the Hausman test of the internal consistency between the two nested OLS estimators.

Second, for non-black women the traditional IV estimator indicates that a delay in childbearing for those women currently choosing to be teenage mothers would harm their children. This finding is consistent across all seven measures analyzed and statistically significant in the following three areas: behavioral problems, mathematical achievement and emotional support provided to the child. Furthermore, relaxing the assumptions of the IV estimator to allow for non-random miscarriages, non-random censoring of child outcomes or the behavior of women intending to have an abortion to be affected by a miscarriage does not alter the qualitative nature of these findings. Note that these results are a far cry from endorsing early or illegitimate childbearing, but they do indicate that those women having children as minors may be acting in their own and their children's best interests given the environment they face.

Third, the data are not as informative for black children, but they are suggestive that the effects may be substantially different than we observed for non-black children. In particular, the IV point estimates are typically negative, meaning that black children suffer due to their mothers' early childbearing. Furthermore, the positive effects estimated for non-black children more often than not exceed the upper bound (A_2 bounds) on the same measure for black children. Thus, despite the fact that the data do not allow us to reject that there are no differences across racial lines, the point estimates and bounds, coupled with the substantial racial difference in both the incidence of teenage childbearing and the fraction of those births which are illegitimate, lead these authors to conclude that the effect of teen fertility probably differ qualitatively by race.

Finally, this study demonstrates that the information generated by less than ideal natural experiments is capable of resolving conflicts about causal effects. Furthermore, the ability to peel back one identifying assumption at a time and see the affect this relaxation produces in the bounds allows researchers to ascertain the robustness of their conclusions to their assumptions. For example, in this study the three measures for which the IV estimator produces statistically significant impacts for non-black children are robust to the to the inclusion of non-random miscarriages. Additionally, the results are qualitatively similar when either non-random censoring of child outcomes is permitted or the no Hawthorne effect assumption is relaxed. However, if all three of these assumptions are relaxed simultaneously, the conclusion is seriously weakened.

References

Angrist, Joshua D. and William N. Evans, “Schooling and Labor Market Consequences of the 1970 State Abortion Reforms,” NBER Working Paper #5406 (January, 1996).

Angrist, Joshua D., Guido W. Imbens and Donald B. Rubin, “Identification of Causal Effects Using Instrumental Variables,” *Journal of the American Statistical Association* Volume 91, Issue 434 (June, 1996): 444 - 455.

Bronars, Stephen G. and Jeff Groger, "The Economic Consequences of Unwed Motherhood: Using Twins as a Natural Experiment." *The American Economic Review*, Volume 84, Number 5 (December, 1994): 1141 - 1156.

Geronimus, Arline and Sanders Korenman, “The Socioeconomic Consequences of Teen Childbearing Reconsidered.” *The Quarterly Journal of Economics* Volume 107, Issue 4 (November 1992): 1187 - 1214.

Geronimus, Arline, Sanders Korenman, and Marianne M. Hillemeier, “Does Young Maternal Age Adversely Affect Child Development? Evidence from Cousin Comparisons in the United States.” *Population and Development Review* Volume 20, Issue 3 (September, 1994): 585 - 609.

Harlap, S., P. Shiono and S. Ramcharan, “A Life Table of Spontaneous Abortions and the Effects of Age, Parity and Other Variables.” *Human Embryonic and Fetal Death*, Edited by I. Porter and E. Hook, New York: Academic Press, 1980.

Hausman, J. A. "Specification Tests in Econometrics," *Econometrica* Volume 46, Number 6, (November, 1978): 1252 - 1271.

Heckman, James J. “Alternative Approaches to the Evaluation of Social Programs: Econometric and Experimental Methods,” (1990, Barcelona Lecture, World Congress of the Econometric Society).

Heckman, James J. “Randomization as an Instrumental Variable,” *Review of Economics and Statistics* Volume 78, Issue 2 (May, 1996): 336 - 342.

Heckman, James J. and R. Robb, “Alternative Methods for Evaluating the Impact of Interventions,” in Heckman and Singer (eds.) *Longitudinal Analysis of Labor Market Data*, New York, Cambridge University Press (1985).

Heckman, James J. and Jeffrey A. Smith, “Assessing the Case for Social Experiments,” *Journal of Economic Perspectives* Volume 9, Number 2 (Spring, 1995): 85 - 110.

Hoffman, Saul, E. Michael Foster, and Frank Furstenberg, “Reevaluating the Costs of Teenage Childbearing.” *Demography* Volume 30, Issue 1 (February, 1993): 1 - 13.

Horowitz, J. and C. Manski “Identification and Robustness with Contaminated and Corrupted Data,” *Econometrica* Volume 63, Number 2 (March, 1995): 281 - 302.

Hotz, V. Joseph, Susan McElroy, and Seth Sanders, “The Costs and Consequences of Teenage Childbearing for Mothers,” in *Kids Having Kids: Economics Costs and Social Consequences of Teen Pregnancy*, Urban Institute Press, 1997: 55 - 94.

Hotz, V. Joseph, Charles H. Mullin, and Seth Sanders, "Bounding Causal Effects Using Data from a Contaminated Natural Experiment: Analyzing the Effects of Teenage Childbearing," *Review of Economic Studies* Volume 64, Issue 4 (October 1997): 575 - 603.

Hotz, V. Joseph and Seth Sanders, "Bounding Treatment Effects in Experimental Evaluations Subject to Post-Randomization Treatment Choice," (1996 Unpublished, University of Chicago).

Imbens, G. and J. Angrist, "Identification and Estimation of Local Average Treatment Effects," *Econometrica* Volume 62, Number 2 (March, 1994): 467 - 475.

Jones, E. and J. Forrest, "Underreporting of Abortions in Surveys of U.S. Women: 1976 – 1988," *Demography* Volume 29, Issue 1 (February, 1992), 113 - 126

Kline, J., Z. Stein, and M. Susser, *Conception to Birth: Epidemiology of Prenatal Development*, New York: Oxford University Press, 1989.

MaCurdy, Thomas, Thomas Mroz, and R. Mark Gritz, "An Evaluation of the NLSY." *Journal of Human Resources* Volume 33, Issue 2 (Spring, 1998), 345 - 436.

Manski, Charles F. "Anatomy of the Selection Problem," *Journal of Human Resources* Volume 24, Issue 2 (Spring, 1989): 343 - 360.

Manski, Charles F. "Non-Parametric Bounds on Treatment Effects," *American Economic Review* Volume 80, Issue 2 (May, 1990): 319 - 323.

Manski, Charles F. and Daniel S. Nagin "Bounding Disagreements About Treatment Effects: A Case Study of Sentencing and Recidivism," *Sociological Methodology* Volume 28 (1998): 99 - 137.

National Center on Addiction and Substance Abuse at Columbia University, "Who Uses During Pregnancy?" (1996).

Perlman, M. D., "One-Sided Testing Problems in Multivariate Analysis." Annals of Mathematical Statistics Volume 40, Issue 2 (April, 1969): 549 - 567.

Ribar, David C., "Teenage Fertility and High School Completion." *The Review of Economics and Statistics* Volume 76, Issue 3 (August, 1994): 413 - 424.

Rosenzweig, Mark R. and Kenneth I. Wolpin, "Sisters, Siblings, and Mothers: The Effect of Teen-Age Childbearing on Birth Outcomes in a Dynamic Family Context." *Econometrica* Volume 63, Number 2 (March, 1995): 303 - 326.

Wolak, Frank A., "An Exact Test for Multiple Inequality and Equality Constraints in the Linear Regression Model." *Journal of the American Statistical Association* Volume 82, Number 399 (September, 1989): 782 - 793.

Table I
Parameter Estimates

	Miscarriage Population		
	Everyone	**Black**	**Non-Black**
Percent of Random Miscarriages	0.87	0.83	0.88
Gamma	0.79	0.88	0.77
Eta	0.84	0.87	0.83
Lambda			
Decomposition A2	0.75	0.76	0.74
Decomposition A3	0.65	0.62	0.70
Theta			
Decomposition A2	0.09	0.04	0.11
Decomposition A3	0.20	0.19	0.16
Delta			
Decomposition A2	0.85	0.87	0.84
Decomposition A3	0.86	0.87	0.84
Lambda Prime			
Decomposition A2	0.76	0.76	0.75
Decomposition A3	0.66	0.61	0.71
Theta Prime			
Decomposition A2	0.09	0.04	0.11
Decomposition A3	0.18	0.19	0.14

Table II
Child Assessments Administered in the NLSY: 1986, 1988, 1990, 1992, 1994 and 1996

Instrument	**Objective**	**Measurement**	**Age of Children**
Birth Weight	Health of child	Weight in ounces	At birth
Behavioral Problems	To measure the frequency, range and type of childhood behavior problems	Mothers respond to 28 questions on specific behavior problems over the past three months	4+ years
Cognitive Attainment			
PIAT Mathematics	To measure academic achievement in mathematics	84 multiple choice questions	5+ years
PIAT Reading Recognition	To measure word recognition and pronunciation	Child matches letter, names letters and reads single words aloud	5+ years
Peabody Picture Vocabulary Test	To measure receptive vocabulary	Child chooses one of four pictures that best matches a stated word	3+ years
Home Environment			
Cognitive	To assess the amount of cognitive stimulation the child receives	Interviewer observation of the interaction between mother and child Maternal report of child's interactions with adults, discipline methods, toys/educational materials and available activities	5+ years
Emotional	To assess the amount of emotional support the child receives	Same as cognitive	5+ years

Table III
Point Estimates of the Percentile Change in Children's Outcomes from Bearing Children as Minors for Black Women
(Standard Errors are Given in Parentheses)

	OLS		IV
Assessment	**All[A]**	**Pregnant[B]**	**Pregnant[B]**
Birth Weight	-1.41	0.18	-9.73
	(1.42)	(2.45)	(7.70)
Behavioral Problems	1.73	2.83	-3.08
	(1.36)	(2.41)	(5.95)
Cognitive Attainment			
PIAT Mathematics	0.27	1.70	-5.97
	(1.24)	(2.32)	(5.95)
PIAT Reading Recognition	-2.56**	-0.66	-2.79
	(1.28)	(2.58)	(5.26)
PPVT	2.50**	2.79	7.46
	(1.26)	(2.39)	(4.91)
Home Environment			
Cognitive	-1.44	1.25	-2.52
	(1.16)	(2.28)	(5.47)
Emotional	-0.49	1.74	1.53
	(1.13)	(2.17)	(4.23)

[A] All women with children are included in the analysis.
[B] Each women with a child who experienced her first pregnancy before 18 is include
* Significant at 0.10 level.
** Significant at 0.05 level.

Table IV
Point Estimates of the Percentile Change in Children's Outcomes from Bearing Children as Minors for Non-Black Women
(Standard Errors are Given in Parentheses)

Assessment	OLS All[A]	OLS Pregnant[B]	IV Pregnant[B]
Birth Weight	-1.29	-1.41	3.99
	(1.32)	(2.08)	(8.27)
Behavioral Problems	-2.60*	2.51	10.02**
	(1.50)	(2.23)	(4.38)
Cognitive Attainment			
PIAT Mathematics	-0.19	3.26	10.53**
	(1.22)	(2.08)	(5.10)
PIAT Reading Recognition	0.32	2.17	3.58
	(1.13)	(2.01)	(6.15)
PPVT	2.07	1.54	4.38
	(1.50)	(2.35)	(6.59)
Home Environment			
Cognitive	-1.31	1.45	4.02
	(1.13)	(1.96)	(4.28)
Emotional	-1.38	2.90	13.13**
	(1.16)	(2.02)	(4.85)

[A] All women with children are included in the analysis.
[B] Each women with a child who experienced her first pregnancy before 18 is include
* Significant at 0.10 level.
** Significant at 0.05 level.

Table V
Horowotz-Manski Bounds of the Percentile Change in Children's Outcomes from Having a Teen Mother: Allowing for Non-Random Miscarriages
(Standard Errors are Given in Parentheses)

	Black		Non-Black	
Assessment	**Lower**	**Upper**	**Lower**	**Upper**
Birth Weight	-19.31	-0.15	-0.82	6.79
	(8.07)	(8.45)	(8.34)	(8.51)
Behavioral Problems	-6.83	5.05	9.29**	14.30
	(6.61)	(6.78)	(5.16)	(4.38)
Cognitive Attainment				
PIAT Mathematics	-12.61	2.01	8.42*	13.35
	(6.70)	(6.58)	(5.73)	(4.46)
PIAT Reading Recognition	-10.09	3.91	0.76	9.16
	(5.97)	(5.50)	(7.30)	(6.15)
PPVT	4.46	15.73	0.16	8.19
	(5.90)	(3.42)	(7.78)	(6.14)
Home Environment				
Cognitive	-7.22	5.19	1.87	5.78
	(5.95)	(6.15)	(4.77)	(4.49)
Emotional	-3.31	8.37	11.15**	15.67
	(4.28)	(4.69)	(5.39)	(4.51)

* Significant at 0.10 level.
** Significant at 0.05 level.

Table VI
Horowotz-Manski Bounds of the Percentile Change in Non-Black Children's Outcomes from Having a Teen Mother:
Allowing for Non-Random Miscarriages and Non-Randoming Censoring of Child Outcomes
(Standard Errors are Given in Parentheses)

	The Expected Value of Censored Outcomes is Within					
	The Minimum and Maximum Values		Two Standard Deviations of the Sample Mean		One Standard Deviation of the Sample Mean	
Assessment	**Lower**	**Upper**	**Lower**	**Upper**	**Lower**	**Upper**
Birth Weight	-	24.31	-5.43	10.85	-2.96	8.37
	-	(7.26)	(7.12)	(7.26)	(7.12)	(7.26)
Behavioral Problems	-1.95	18.26	4.21	17.43	6.46*	15.18
	(4.40)	(3.76)	(4.40)	(3.76)	(4.40)	(3.76)
Cognitive Attainment						
PIAT Mathematics	-0.94	19.21	3.75	18.31	6.35*	15.70
	(4.87)	(3.82)	(4.87)	(3.82)	(4.87)	(3.82)
PIAT Reading Recognition	-6.27	16.79	-3.64	14.29	-0.92	11.57
	(6.19)	(5.23)	(6.19)	(5.23)	(6.19)	(5.23)
PPVT	-9.06	13.69	-4.22	13.69	-1.06	11.99
	(6.62)	(5.26)	(6.62)	(5.26)	(6.62)	(5.26)
Home Environment						
Cognitive	-6.76	12.53	-2.22	9.91	-0.01	7.7
	(4.11)	(3.88)	(4.11)	(3.88)	(4.11)	(3.88)
Emotional	0.89	20.68	6.76*	19.5	8.99**	17.26
	(4.60)	(3.88)	(4.60)	(3.88)	(4.60)	(3.88)

* Significant at 0.10 level.
** Significant at 0.05 level.

Table VII
Horowotz-Manski Bounds of the Percentile Change in Children's Outcomes from Having a Teen Mother: Allowing for Non-Random Miscarriages and Hawthorne Effects for Aborters
(Standard Errors are Given in Parentheses)

	Black		Non-Black	
Assessment	**Lower**	**Upper**	**Lower**	**Upper**
Birth Weight	-44.08	14.72	-1.24	10.42
	(5.99)	(6.37)	(7.64)	(7.80)
Behavioral Problems	-20.91	8.42	6.17*	14.93
	(4.82)	(5.20)	(4.32)	(4.04)
Cognitive Attainment				
PIAT Mathematics	-23.99	8.62	4.03	14.00
	(4.79)	(4.57)	(5.12)	(4.14)
PIAT Reading Recognition	-19.19	12.42	-1.52	12.12
	(4.35)	(3.87)	(6.64)	(5.57)
PPVT	-13.80	15.73	-3.90	10.10
	(4.30)	(2.67)	(7.04)	(5.54)
Home Environment				
Cognitive	-19.97	10.16	-1.72	6.61
	(4.15)	(4.56)	(3.93)	(4.39)
Emotional	-18.36	10.85	4.38	15.67
	(3.24)	(3.75)	(4.85)	(4.40)

* Significant at 0.10 level.
** Significant at 0.05 level.