# Endogeneity and Imperfect Instruments: Estimating Bounds for the Effect of Early Childbearing on High School Completion

Steffen Reinhold* Tiemen Woutersen†

September 20, 2010

**Abstract**

This paper derives informative bounds of the effect of early childbearing on high school completion. We allow the exclusion restriction of the instrument to be violated. In particular, we assume that the correlation between the instrument and the structural error is smaller than the correlation between the structural error and the endogenous regressor. We derive a confidence interval using the regular bootstrap and find that the least squares estimate is outside this confidence interval. That is, the bias of the least squares estimator is both substantial and statistically significant.

Keywords: Instrumental Variables, Validity, Identification, Bounds, Teenage Childbearing, Educational Attainment.
JEL Classification: C310, J130

*Mannheim Research Institute for the Economics of Aging, University of Mannheim; L13,17; 68131 Mannheim; Germany; email: reinhold@mea.uni-mannheim.de

†Johns Hopkins University, Department of Economics; 3400 N. Charles Street; MD 21218; email: woutersen@jhu.edu. We would like to thank Colleen Carey, Norman Swanson, Roger Klein, and Anne Morrison Piehl for valuable comments.

# 1 Introduction

There is a close association between early childbearing and adverse economic outcomes for mothers and their children such as poverty risk, lower educational achievement, and depressed earnings. While it is plausible to describe these outcomes as consequences of early fertility decisions, this correlation can also arise from a correlation of unobserved factors with early childbearing. The instrumental variable technique is one approach to deal with the problems of endogenous regressors. However, depending on the exact choice of the instruments, a wide range of estimates emerges.

An instrument needs to be correlated with the endogenous regressor teenage childbearing (relevance) but it must be uncorrelated with the structural error term in the outcome equation (validity). One problem with this approach is that often doubts linger as to whether the instruments satisfy this second assumption. In this paper, we employ a set identification approach and allow the instruments to be imperfect. In particular, the correlation between the instrument and the structural error is assumed to be smaller than the correlation between the structural error and the endogenous regressor childbearing. Nevo and Rosen (2008) were the first to use this inequality. This approach relaxes the validity assumption, allowing limited correlation with the error term and thus widens the set of potential instruments. Nevo and Rosen's work builds on a recent literature on partial identification. Horowitz and Manski (1998) develop interval estimates that asymptotically cover the entire identified region with fixed probability. Chernozhukov, Hong and Tamer (2007) extend this approach through formulating the problem of covering the entire identified region as a minimization problem. Imbens and Manski (2004) derive the confidence interval that covers each element in the identified region with fixed probability. This confidence interval is in general shorter than the earlier derived confidence intervals. Woutersen (2008) develops an easier and perhaps more intuitive way to calculate the confidence interval that covers each element in the identified region with fixed probability. An advantage of this method is that the regular bootstrap[1] can be used.

[1] rather than the subsampling bootstrap

We investigate whether we can identify informative bounds for the causal effect of teenage pregnancy on high school completion under the relaxed assumptions. A researcher may also be interested in the question of how much her results are driven by the identifying assumptions by employing a weaker assumptions on the instruments. In addition, this paper helps to assess whether set estimators, in general, provide informative bounds on the parameters of interest in an application that is typical for the literature in labor and demographic economics.

In the literature on the effects of teenage childbearing on high school completion, the problem of potentially invalid instruments is particularly urgent: Hotz et al. (1997, 1999, 2005) use the occurrence of a miscarriage among women who have become pregnant as teenagers as an instrument for teenage childbearing. They use the sample of teenagers who have become pregnant because they argue that miscarriages are a random phenomenon for pregnant women. But since pregnancy itself is not a random event, their inference does not extend to the general population. This is because the occurrence of a miscarriage does not satisfy the strong assumption of validity when applied to the full sample of women. But even in the restricted sample there is the possibility that the instrument does not satisfy the validity assumption if miscarriages are correlated with unobserved family background or the woman's health behavior. However with the identifying assumptions of Nevo and Rosen (2008), the occurrence of a miscarriage becomes a legitimate instrument even in the full sample if one believes that the correlation between it and the unobserved factors in the outcome equation is smaller than the correlation between teenage childbearing and the unobserved factors.

Another potential instrument is age at menarche (Ribar, 1994). Age at menarche is correlated with teenage pregnancy and via this pathway also with teenage childbearing. Ribar therefore uses this instrument without a sample restriction to women who have become pregnant as teenagers. Using only age at menarche, Ribar finds a detrimental effect of teenage childbearing on educational achievement. Although Ribar's empirical setup is different, the same fundamental problem emerges. Age at menarche may be correlated with other unobserved factors like nutrition during childhood (Freedman, Khan, Serdula, Dietz, Srinivasan, and Berenson 2002) which may itself influence education. This would render this instrument invalid.

Using different sets of instrument often leads to differing qualitative conclusions about the causal effect of teenage childbearing on high school completion as in this particular case. Whereas the results using the occurrence of a miscarriage indicate no adverse effect of teenage childbearing one finds a rather strong adverse effect on high school completion using age at menarche as an instrument. Geronimus and Korenman (1992 and 1993) and Hoffman, Foster and Furstenberg (1993) argue vigorously about the right set of assumptions to point identify the effect of childbearing on high school completion. No consensus emerged and, therefore, this seems a natural application for bounds. A paper that is closely related to ours is Hotz, Mullin, and Sanders (1997). They estimate the number of miscarriages that are "random" and use this to derive bounds. They find that teenage childbearing causes the labor market earnings and hours worked to increase. In this paper, we neither need to estimate the fraction of miscarriages that are "random," nor do we need to take a position on what constitutes a "random miscarriage." We view this as an advantage. Also, we report the confidence interval of the causal effect, rather than the confidence interval of the partially identified set.

The paper is organized as follows. After discussing the data, we present our model and discuss the identifying assumptions about the unobserved correlation between the instrument and the error term. We then derive confidence intervals for the causal effect of teenage childbearing using both Imbens and Manski (2004) and Woutersen (2008). We then check the robustness of these results by also estimating the partial linear model with endogeneity and the models derived by Lewbel (2000) and Magnac and Maurin (2008).

# 2 Description of Data

We use the 2002 cycle of the National Survey of Family Growth (NSFG) a representative sample of 7643 women from the ages 15-44. We restrict our sample to women older than 20 years. At this age, high school should normally be completed. With this sample restriction we have 6443 observations in our complete sample. In addition, we also construct subsamples of women having become pregnant before age 18 (N=1284), ever pregnant women (N=4810), and women having children in their household (N=3910).

Thus, we use the instruments in less restricted samples than Hotz et al. (1997, 1999, 2005) who only estimated the effect of teenage childbearing in a sample of woman who experienced teenage pregnancy. Under the stronger validity assumptions this restriction is clearly reasonable. However, under our relaxed assumptions we can use the occurrence of a miscarriage even in the full sample of women as long as the correlation between teenage childbearing and the structural error term is smaller as the correlation between the occurrence of a miscarriage and the structural error term.

As additional controls we also use dummies for religious affiliation (Catholic, Protestant, no religion, and other), race, age and age squared, a dummy for migrant status, intact family background, and dummies for parental education.

# 3 Empirical Model

## 3.1 Estimation of Bounds

Consider the following model:

$$Y = X\beta + W\delta + U \tag{1}$$

where $Y$ is a dummy which takes the value of 1 for not completing high school and zero otherwise, $X$ is a dummy which takes the value 1 if the individual gives birth as a teenager and zero otherwise.[2] $W$ is a vector of other covariates including age at interview and its square, race, religion, parents' educational background, intact family background, and migration background. There is also a set of imperfect instruments $Z$ including age at menarche and the occurrence of a miscarriage.

Controlling for other covariates in IV regressions is often important because the assumption of zero correlation between the error term and the instrument only needs to hold after conditioning on all exogenous variables. Since we do not require zero correlation between the instrument and the error term we can proceed with estimating the model without covariates. Also, for estimating we will partial out all other covariates, so that we essentially are estimating a simple bivariate model with one endogenous regressor. For this reason, we also discuss the following bivariate model:

$$Y = X\beta + U \tag{2}$$

We assume that the correlations between the endogenous regressor and the error term and between the instrument and the error term have the same sign:

$$\rho_{XU}\rho_{Z_jU} \geq 0 \tag{3}$$

where $j$ indexes the instrument in the case of multiple instruments. This is not restrictive because one can just multiply the instrument by negative 1 for convenience without loss of generality. But one needs to make an assumption about these correlations because one never can observe the true error term. Nevo and Rosen (2008) assume that the correlation between the instruments and the error term is weaker in absolute terms than the correlation between the endogenous regressor and the error term, that is:

$$|\rho_{XU}| \geq |\rho_{ZU}| \tag{4}$$

[2]For those who miscarry this dummy can still take the value of 1 if a second teenage pregnancy results in a live-birth, but these cases are rare.

This assumption considerably weakens the usual assumptions for instrumental variables which would require that $\rho_{ZU} = 0$. We believe that in our case these assumptions are satisfied.

Hotz et al. (1997, 1999, 2005) argue that pregnancies are not a random event in the population, and hence the occurrence of a miscarriage cannot be a valid instrument in the full sample because it is correlated with pregnancies. Becoming pregnant as a teenager may well be correlated with the error term, for example because of unobserved preferences for education or future earnings potential. Conditional on being pregnant, however, miscarriages are largely a random event with possible some behavioral risk factors such as illicit drug use. If miscarriages are largely random then the correlation between miscarriages and the error term must be smaller than the correlation between teenage childbearing and the error term. Similarly, it seems plausible that there is enough random variation in age at menarche so that the correlation between the structural error term and age at menarche is relatively small.

In addition, we impose the usual rank conditions ensuring that the probability limits of OLS and 2SLS estimators are well defined. We will rely for estimation on their second proposition, and here it is important to also know the correlation between the instruments and the endogenous regressor. Our assumptions on these correlations are detailed in the next section. Under these assumptions, the restriction of equation (4) implies either an upper bound (if $\sigma_{XZ} < 0$ and $\sigma_{XU} > 0$) or a lower bound (if $\sigma_{XZ} < 0$ and $\sigma_{XU} < 0$) while the IV estimator provides the lower and upper bound, respectively.

Because the correlation between teenage childbearing and the occurrence of a miscarriage is negative we already have two-sided bounds using the occurrence of a miscarriage as an instrument. The question becomes whether age at menarche helps in tightening these bounds. This depends on whether the upper bound using age at menarche as an instrument is smaller than the upper bound using the occurrence of a miscarriage as an instrument.

## 3.2 Assumptions on the Correlation between the Instruments and the Unobserved Heterogeneity

Nevo and Rosen (2008) replace the assumption of no correlation between the instrument and the error term by the weaker assumption that this correlation has to be smaller in absolute terms than the correlation between the endogenous regressor and the error term. However, one needs to make an assumption about the sign of this correlation. In this section we discuss this assumption, and in addition, we provide key summary statistics for the observed variances and covariances of the endogenous regressor and the instruments which are needed to calculate the bounds.

We summarize the information about our instruments and our assumptions about key correlations in Tables 1 and 2. Given our assumptions on the correlation between the unobserved factors and our instruments, one can derive two-sided bounds using the occurrence of a miscarriage as an instrument, whereas one can estimate upper bounds using age at menarche (or its negative) as an instruments.

We assert that there is a positive correlation between the error term and early childbearing. In our example, a dummy for not completing high school is the outcome variable and teenage childbearing is the endogenous regressor. Unobserved factors such as a disadvantaged family background may make a teenager both more likely to have children very early and to drop out of high school. The covariances between the instruments and the dummy for early childbearing have the sign one would expect. Women who have experienced a miscarriage are less likely to have experienced a live-birth as teenagers, and teenagers with an early onset of their menarche are more likely to become pregnant and bear children as teenagers.

Furthermore, we assume that the occurrence of a miscarriage, henceforth $Z_1$, is also positively correlated with the error term. In the full sample of all women, miscarriages are positively correlated with pregnancies and to the extent that teenagers from disadvantaged families have earlier pregnancies one would expect a positive correlation. In addition, the occurrence of miscarriages may still be correlated with certain risk factors such as smoking or drinking (Hotz, McElroy, and Sanders 1999) or a very young age at conception. Problems with the validity of the instrument remain if one cannot control for all of those risk factors. Furthermore, Ashcraft and Lang (2006) argue that the option of obtaining an abortion invalidates the instrument. They show that women obtaining an abortion come from more advantaged backgrounds. If some abortions occur before a potential miscarriage, the sample of women experiencing a miscarriage is not a random but comes, on average, from a more disadvantaged background. Similarly, Fletcher and Wolfe (2009) examine the role of unobserved community level effects which are which are correlated with miscarriages.

For the second instrument, age at menarche ($Z_2$), the direction of the correlation is more open for argument. Weil (2007) uses age at menarche as a health indicator in his work and reports that in poorer countries the age at menarche is later. If poor living conditions are associated with a late age at menarche one could expect that $\rho_{Z_2U} > 0$. Notice however, that he uses this indicator to compare developed with developing countries. In our sample, we only have a comparison within the United States, and we assume that $\rho_{Z_2U} < 0$. Freedman et al. (2002) present evidence that obesity and an early age at menarche are positively correlated. Obesity is a marker of a disadvantaged family background which would indicate that women from disadvantaged family background may have a lower age at menarche.

The assumptions on the correlation structure are the same for the more general version of the model where there are additional covariates $W$. In this case, one regresses both the outcome and the endogenous regressor on these covariates and obtains residuals from these regressions. Let $\widetilde{X}, \widetilde{Y}$ denote these residuals. Using these residuals one can estimate a bivariate model. Alternatively, one can also just use 2SLS estimation with additional covariates to obtain the same estimates.

## 3.3 Estimating Confidence Intervals Covering the True Parameter

Woutersen (2008) allows for the regular bootstrap to be used for confidence intervals. That is, we sample individuals with replacement and get an estimate for the lower and upper bound for each subsample. We generate 5000 subsamples. We then put all the estimates of the upper and lower bound in a ordered vector (with length 10,000) and calculate the 2.5% and 97.5% percentile. This confidence interval has the correct coverage and the asymptotic refinement of the bootstrap

# 4 Results

Table 6 presents the estimates of the bounds of the effect on early childbearing both with and without additional covariates for the complete sample and for subsamples of women who have ever been pregnant, women who have become pregnant as teenagers, and women who currently have children in their household. In addition, we present the OLS estimates of the coefficient on teenage childbearing in these samples. The bounds estimates use the occurrence of a miscarriage, negative of age at menarche, and the transformed $Z^*$ as instruments. Using these regression estimates, we can derive bounds for the effect of teenage childbearing on high school completion.

We use Woutersen's (2008) method to calculate the 95% confidence intervals for the set estimates employing the bootstrap. In addition, we also estimated 95% confidence intervals using the method of Imbens and Manski (2004). The resulting confidence intervals only differed marginally and can be found in the appendix.

The OLS results in the last column confirm previous findings of the literature showing a clear association of teenage childbearing with lower educational attainment. Notably, the association is weakest in the teenage pregnancy sample indicating maybe self-selection of individuals with low educational prospects into early pregnancy.

Turning to the negative of age at menarche as an instrument, we can only derive upper bounds for the effect of teenage childbearing. Panel A shows the results for the complete sample, where the upper bound is estimated using 2SLS with the negative of age at menarche as an instrument. If one believes that this instrument is valid, then this is just the conventional point estimator of the causal effect of early childbearing. It is 0.218 when using no covariates and 0.146 when using covariates. These estimates of the upper bound of the effect are well below the OLS estimates as expected if the correlation between teenage childbearing and the structural error term is positive. However, they are very imprecisely estimated resulting in 95% confidence intervals for the true parameter that are not very informative. In all the other samples, a similar picture emerges.

Turning to miscarriage, we present estimates of the upper and lower bound using the occurrence of a miscarriage or the transformed $Z^*$ as instruments. Using the occurrence of a miscarriage as an instrument, it is possible to derive the lower bound for the causal effect of early childbearing. The upper bound is estimated using the transformed $Z^*$ which is a linear combination of teenage childbirth and the occurrence of a miscarriage.

In panel A we again present the results for the complete sample where concerns about the validity of the instrument are most serious. In this sample, the estimated upper bound for the causal effect of early childbearing is well below the OLS estimate. Even considering the uncertainty associated with these estimates, the 95% confidence intervals using the occurrence of a miscarriage as instruments do not cover the corresponding OLS estimates.

Even if one uses the upper end of the confidence interval which is closest to OLS, one would find a smaller adverse effect of early childbearing on educational attainment. Notice that the estimates of the bounds are consistent with positive effects of early childbearing on educational attainment, and they are certainly consistent with a zero effect.

Turning to the sample of women who have ever been pregnant in panel B, we estimate a lower bounds for the effect of early childbearing of around zero. But the upper bound indicates in this sample a greatly increased risk of dropping out of high school for teenage mothers of comparable size to the OLS estimates. Turning to the sample of women who experienced a pregnancy as teenagers in panel C using the instruments leads to qualitatively different conclusions than the corresponding OLS estimates. Judging by the estimates of the lower bounds one cannot exclude the possibility that early childbearing is beneficial for educational achievement. And even when one is more conservative and looks only at the upper bounds one finds estimates of the upper bound that are well below the OLS estimates.

In the sample of women who have children, a similar picture emerges as for the sample of ever pregnant women. Again, the upper bound of the estimate of the effect of early childbearing is quantitatively close to the OLS estimate.

In this particular case, age at menarche does not add much new information to sharpen the estimates because the associated standard errors are large. But somewhat similar to an overidentifiying assumptions test, the presence of a second instrument is useful for a specification test. Nevo and Rosen (2008) suggest inspecting the overlap of the set estimates using the different instruments. If there is no overlap the model would be misspecified. In all our cases, the set estimates overlap, so that we cannot reject the specification using this test.

# 5 Robustness checks

In this section, we estimate various semiparametric models to show the robustness of the estimation results. First, we extend Robinson's (1988) model to allow for a endogenous regressor. Let $K(w)$ be a kernel as in Robinson (1988, definition 1). Let $K_{ij} = K(\frac{W_i - W_j}{a})$ where $a$ is the bandwidth and let $\mathcal{G}_{\mu}^{\alpha}$ be a class of functions $\mathbb{R}^q \to \mathbb{R}$ as in Robinson (1988, definition 2). Let

$$\begin{aligned}
\hat{Y}_i &= 1(\frac{\sum_j K_{ij}}{Na^q} > \eta)\frac{\sum_j K_{ij} Y_i}{\sum_j K_{ij}} \\
\hat{X}_i &= 1(\frac{\sum_j K_{ij}}{Na^q} > \eta)1(\frac{\sum_j K_{ij} X_i}{\sum_j K_{ij}}) \\
\hat{Z}_i &= 1(\frac{\sum_j K_{ij}}{Na^q} > \eta)1(\frac{\sum_j K_{ij} Z_i}{\sum_j K_{ij}}).
\end{aligned}$$

where $\frac{\sum_j K_{ij}}{Na^q}$ is an estimate of the density of $W$ at $W_i$. So we would trim the data if this density is estimated to be very low (we did not need to trim the data in the application). Define

$$\hat{\beta}_{\text{Partial linear}} = \frac{\sum_i (Z_i - \hat{Z}_i)(Y_i - \hat{Y}_i)}{\sum_i (Z_i - \hat{Z}_i)(X_i - \hat{X}_i)}.$$

The following lemma closely follows Robinson (1988) but relaxes the exogeneity assumption on $X$ and introduces the instruments $Z$.

**Lemma**

Let the following conditions hold: (i) $(W_i, X_i, Y_i, Z_i), i = 1, 2, ...,$ are independent and distributed as $(W, X, Y, Z)$; (ii) $Y = X\beta + \theta(W) + U$; (iii) $U$ is independent of $W, Z$; (iv) $E(U^2) = \sigma^2 < \infty$; (v) $E|X|^4 < \infty$ and $E|Z|^4 < \infty$; (vi) $W$ admits a pdf $\mathcal{G}_\lambda^\infty$, for some $\lambda > 0$; (vii) $\zeta \in \mathcal{G}_v^4$ for some $\mu > 0$; (viii) $\theta \in \mathcal{G}_v^4$ for some $v > 0$; (ix) as $N \to \infty$, $Na^{2q}b^4 \to \infty, Na^{2\min(\lambda+1,\mu)+2\min(\lambda+1,v)}b^{-4} \to 0$, $a^{\min(\lambda+1,2\lambda,\mu,v)}b^{-2} \to 0, b \to 0$; (x) k$\in \mathcal{K}_{\max(l+m-1,l+n-1),}$ for integers l,m,n such that $l - 1 < \lambda \leq l, m - 1 < \mu \leq n$; (xi) $E[\{Z - E(Z|W)\}\{X_i - E(X|W)\}] = \delta \neq 0$. Then

$$N^{1/2}(\hat{\beta} - \beta) \to N(0, \frac{\sigma^2}{\delta^2}).$$

Proof: See appendix.

We use the lemma to derive upper and lower bounds for the causal effect of early childbearing on high school completion in the same way that we used the two stage least squares estimator before. The estimation results of the partial linear model with endogenous regressors are very close to the estimation results of the linear model, which is reassuring. We also estimated the bivariate probit model, which exploits the fact that both childbearing and finishing high school are binary events. The estimation results are well inside the confidence interval of the linear model. In order to test the validity of the distributional assumptions, we use the semiparametric estimators of Lewbel (2000) and Magnac and Maurin (2008). These estimators can be used to calculate the ratio of parameters and, thereby, test the validity of the MLE estimates. These semiparametric estimators do not reject the likelihood model. We also used Lewbel (2000) and Magnac and Maurin (2008) to calculate the marginal effects[3]. The marginal effects were inside the confidence interval for the marginal effects we found using the linear model or the probit model for most specifications.

[3]by extending the assumptions, see the appendix for details.

## 6 Conclusion

In this paper we estimated bounds for the causal effect of early childbearing on educational attainment. We have found informative bounds for this parameter. The upper bound of the set estimate is well below the OLS estimate for all the instruments, therefore leading to different qualitative conclusions. The estimated bounds would even be consistent with no or even positive effects of early childbearing on educational attainment. Being conservative and only considering the upper bound of the effect one finds an adverse effect of early childbearing which is somewhat smaller than the OLS estimates suggest. However, especially when using age at menarche as an instrument, there is some estimation uncertainty associated with it. Combining both instruments one can tighten the set estimates even further because the upper bound estimated with age at menarche is smaller than the upper bound using the occurrence of a miscarriage as an instrument. However, one would still prefer the 95% confidence intervals only using the occurrence of a miscarriage as instrument because these bounds are more precisely estimated. But this overlap can also be seen as a specification test. If there was no overlap, then one would question whether the model is correctly specified. Constructing the overlap can also be seen as a way to reconcile different estimates when using alternative instruments. In our case, it is easy to reconcile the estimates because the confidence intervals for the parameters using each instrument individually overlap. Substantively, we find that the adverse effects of early childbearing are smaller than one would expect based on the usual OLS estimates confirming earlier empirical results. This lends more credibility to those results because one obtains them here under much weaker assumptions about the validity of instruments. Geronimus and Korenman (1993) note "We continue to recognize the limitations of currently available methods and data for accounting for unobserved heterogeneity and selectivity." We cannot change the selectivity (self selection) but this paper shows how to use weaker identifying assumptions in order to derive informative bounds for the the causal effect of childbearing.

# References


Ashcraft, Adam, and Kevin Lang (2006): "The Consequences of Teenage Childbearing," NBER Working Paper 12485.

Chernozhukov, Victor, Han Hong, and Elie Tamer (2007): "Parameter Set Inference in a Class of Econometric Models," *Econometrica*, 75(5), 1243–1284.

Fletcher, Jason M., and Barbara L. Wolfe (2009): "Education and Labor Market Consequences of Teenage Childbearing. Evidence Using the Timing of Pregnancy Outcomes and Community Fixed Effects," *Journal of Human Resources*, 44(2), 303–325.

Freedman, David S., Laura Kettel Khan, Mary K. Serdula, William H. Dietz, Sathanur R. Srinivasan, and Gerald S. Berenson (2002): "Relation of Age at Menarche to Race, Time Period, and Anthropometric Dimensions: The Bogalusa Heart Study," *Pediatrics*, 110(4), 1–7.

Geronimus, Arline T., and Sanders Korenman (1992): "The Socioeconomic Consequences of Teen Childbearing Reconsidered," *The Quarterly Journal of Economics*, 107(4), 1187–1214.

——— (1993): "The Socioeconomic Costs of Teenage Childbearing: Evidence and Interpretation," *Demography*, 30(2), 281–290.

Hoffman, Saul D., E. Michael Foster, and Frank F. Furstenberg Jr. (1993): "Reevaluating the Costs of Teenage Childbearing: Response to Geronimus and Korenman," *Demography*, 30(2), 291–296.

Horowitz, Joel L. (2001): "The Bootstrap," in *Handbook of Econometrics*, ed. by J. Heckman, and E. Leamer. Amsterdam, North-Holland: Elsevier.

Horowitz, Joel L., and Charles F. Manski (1998): "Censoring of Outcomes and Regressors due to Survey Nonresponse: Identification and Estimation Using Weights and Imputations," *Journal of Econometrics*, 84, 37–58.

Hotz, V. Joseph, Susan Williams McElroy, and Seth G. Sanders (1997): "The Costs and Consequences of Teenage Childbearing for Mothers and Government," in *Kids Having Kids*, ed. by R. A. Maynard. Washington D.C.: Urban Institute Press.

——— (1999): "Teenage Childbearing and its Life Cycle Consequences. Exploiting a Natural Experiment," NBER Working Paper 7397.

——— (2005): "Teenage Childbearing and its Life Cycle Consequences," *Journal of Human Resources*, XL(3), 683–715.

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

Imbens, Guido W., and Charles F. Manski (2004): "Confidence Intervals for Partially Identified Parameters," *Econometrica*, 72(6), 1845–1857.

Lewbel, Arthur (2000): "Semiparametric qualitative response model estimation with unknown heteroskedasticity or instrumental variables," *Journal of Econometrics*, 97, 145–177.

Magnac, Thierry, and Eric Maurin (2008): "Partial Identification in Monotone Binary Models: Discrete Regressors and Interval Data," *Review of Economic Studies*, 75, 835–864.

Nevo, Aviv, and Adam Rosen (2008): "Identification with Imperfect Instruments," cemmap Working Paper 16/08.

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

Robinson, P. M. (1988): "Root-N-Consistent Semiparametric Regression," *Econometrica*, 56(4), 931–954.

Weil, David N. (2007): "Accounting for the Effect of Health on Economic Growth," *Quarterly Journal of Economics*, 122(3), 1265–1306.

Woutersen, Tiemen (2008): “A Simple Way to Calculate Confidence Intervals for Partially Identified Parameters,” Mimeo. Johns Hopkins University.

Table 1: Summary Statistics

| | Mean | Variance | $\sigma_{X,Z}$ |
|---|---|---|---|
| Teenage childbearing | 0.14 | 0.12 | |
| -Age at menarche | -12.57 | 2.82 | 0.05 |
| Occurrence of miscarriage | 0.11 | 0.10 | -0.01 |

Table 2: Identifying Assumptions

| | $\rho_{XU}$ | $\sigma_{X,Z}$ | $\rho_{ZU}$ |
|---|---|---|---|
| Teenage childbearing | + | | |
| -Age at menarche | + | 0.05 | + |
| Occurrence of miscarriage | + | -0.01 | + |

Note: $\sigma_{X,Z}$ is estimated from data.

| LPM | Probit | Semiparametric Binary Choice | 2SLS miscarriage | 2SLS age at menarche | 2SLS both | Bivariate Probit miscarriage | Bivariate Probit age at menarche | Bivariate Probi both |
|---|---|---|---|---|---|---|---|---|
| Panel A: Complete Sample, N=6443 | | | | | | | | |
| 0.283*** | 0.275*** | 0.211*** | -0.458** | 0.105 | -0.186 | -0.106*** | 0.009 | -0.067* |
| (0.017) | (0.018) | (0.075) | (0.210) | (0.184) | (0.138) | (0.027) | (0.061) | (0.037) |
| Panel B: Ever Pregnant Sample, N=4810 | | | | | | | | |
| 0.257*** | 0.266*** | 0.221*** | -0.064 | 0.102 | -0.006 | -0.080* | -0.013 | -0.042 |
| (0.018) | (0.019) | (0.046) | (0.123) | (0.170) | (0.099) | (0.046) | (0.074) | (0.049) |
| Panel C: Teenage Pregnancy Sample, N=1284 | | | | | | | | |
| 0.171*** | 0.195*** | 0.097* | -0.116 | 0.288 | -0.110 | -0.127* | 0.135 | -0.123* |
| (0.030) | (0.031) | (0.054) | (0.085) | (0.815) | (0.085) | (0.072) | (0.391) | (0.072) |
| Panel D: Have Child, N=3910 | | | | | | | | |
| 0.236*** | 0.246*** | 0.216*** | -0.200 | -0.072 | -0.147 | -0.127*** | -0.104* | -0.107** |
| (0.020) | (0.021) | (0.075) | (0.185) | (0.219) | (0.141) | (0.044) | (0.059) | (0.045) |

Note: LPM and 2SLS models are estimated using age at interview, race, religion, dummies for mother's and father's education, a dummy for an intact family, and a dummy for being foreign born. We include a third order polynomial in age at interview interacted with all dummy variables. In addition, we interact race with all other dummy variables approximating a saturated model.

Table 3: Set Estimates and Bootstrapped 95% Confidence Intervals using Each Instrument Individually.

| | Age at menarche | | Miscarriage Bootstrapped CI | | OLS |
|---|---|---|---|---|---|
| | Lower Bound | Upper Bound | Lower Bound | Upper Bound | |
| Panel A: Complete Sample. N=6443 | | | | | |
| 2SLS | | | | | |
| No covariates | $-\infty$ | 0.218 | -0.501 | 0.310 | 0.368*** |
| 95% CI | | 0.464 | -0.973 | 0.345 | (0.017) |
| With covariates | $-\infty$ | 0.146 | -0.428 | 0.239 | 0.291*** |
| 95% CI | | 0.439 | -0.827 | 0.274 | (0.017) |
| With covariates saturated | $-\infty$ | 0.105 | -0.458 | 0.228 | 0.283*** |
| 95% CI | | 0.354 | -0.754 | 0.257 | (0.017) |
| Non-parametric IV | | | | | |
| NN match | -$\infty$ | 0.173 | -0.442 | 0.283 | |
| 95% CI | | 0.445 | -0.884 | 0.317 | |
| Panel B: Ever pregnant sample. N=4810 | | | | | |
| No covariates | $-\infty$ | 0.108 | 0.052 | 0.306 | 0.338*** |
| 95% CI | | 0.347 | -0.158 | 0.344 | (0.018) |
| With covariates | $-\infty$ | 0.089 | -0.039 | 0.229 | 0.263*** |
| 95% CI | | 0.366 | -0.251 | 0.267 | (0.018) |
| With covariates saturated | $-\infty$ | 0.102 | -0.064 | 0.220 | 0.257*** |
| 95% CI | | 0.331 | -0.226 | 0.249 | (0.018) |
| Non-parametric IV | | | | | |
| NN match | $-\infty$ | 0.039 | 0.010 | 0.273 | |
| 95% CI | | 0.373 | -0.203 | 0.312 | |
| Panel C: Teenage pregnancy sample. N=1284 | | | | | |
| No covariates | $-\infty$ | -0.048 | -0.117 | 0.156 | 0.253*** |
| 95% CI | | 2.172 | -0.267 | 0.215 | (0.028) |
| With covariates | $-\infty$ | 0.114 | -0.115 | 0.094 | 0.174*** |
| 95% CI | | 2.244 | -0.261 | 0.155 | (0.029) |
| With covariates saturated | $-\infty$ | 0.166 | -0.116 | 0.091 | 0.171*** |
| 95% CI | | 1.808 | -0.235 | 0.141 | (0.030) |
| Non-parametric IV | | | | | |
| NN match | $-\infty$ | 0.259 | -0.080 | 0.150 | |
| 95% CI | | 2.006 | -0.233 | 0.212 | |
| Panel D: Have kids sample. N=3910 | | | | | |
| No covariates | $-\infty$ | 0.050 | -0.036 | 0.286 | 0.320*** |
| 95% CI | | 0.307 | -0.317 | 0.330 | (0.020) |
| With covariates | $-\infty$ | -0.096 | -0.165 | 0.206 | 0.243*** |
| 95% CI | | 0.251 | -0.487 | 0.249 | (0.020) |
| With covariates saturated | $-\infty$ | -0.072 | -0.200 | 0.196 | 0.236*** |
| | | 0.307 | -0.449 | 0.231 | (0.020) |
| Non-parametric IV | | | | | |
| NN match | $-\infty$ | -0.157 | -0.136 | 0.249 | |
| 95% CI | | 0.359 | -0.490 | 0.294 | |

Note: 5000 bootstrap replications were used, 1000 bootstrap replications for saturated model.

# 7 Appendix

## 7.1 Construction of the upper bound when miscarriage is used as an instrument

Here we derive the upper bound as we do in the case of the occurrence of a miscarriage. Suppose we observe $\{X_i, Y_i, Z_i\}$ where $i = 1, ..., N$. Let the realizations be identically and independently distributed. Also, let

$$\begin{aligned} Y_i &= X_i\beta + \varepsilon_i, \text{ and} \\ \rho_{X\varepsilon} &\geq \rho_{Z\varepsilon} \end{aligned} \tag{5}$$

where $\beta$ is a scalar. This yields

$$\frac{Cov(X, \varepsilon)}{\sigma_X \sigma_\varepsilon} \geq \frac{Cov(Z, \varepsilon)}{\sigma_Z \sigma_\varepsilon}$$

$$\frac{Cov(X, Y - X\beta)}{\sigma_X \sigma_\varepsilon} \geq \frac{Cov(Z, Y - X\beta)}{\sigma_Z \sigma_\varepsilon}$$

so that

$$\frac{Cov(X, Y) - \beta \cdot Var(X)}{\sigma_X} \geq \frac{Cov(Z, Y) - \beta \cdot Cov(Z, X)}{\sigma_Z}$$

and

$$\beta \cdot \{Cov(Z, X)\frac{\sigma_X}{\sigma_Z} - Var(X)\} \geq \{Cov(Z, Y)\frac{\sigma_X}{\sigma_Z} - Cov(X, Y)\}.$$

Thus,

$$\beta \leq \frac{Cov(Z, Y)\frac{\sigma_X}{\sigma_Z} - Cov(X, Y)}{Cov(Z, X)\frac{\sigma_X}{\sigma_Z} - Var(X)}.$$

since $Var(X) - Cov(Z, X)\frac{\sigma_X}{\sigma_Z} = \sigma_X \left(\sigma_X - Cov(Z, X)/\sigma_Z\right) > 0$.

Replacing

$$\frac{Cov(Z, Y)\frac{\sigma_X}{\sigma_Z} - Cov(X, Y)}{Cov(Z, X)\frac{\sigma_X}{\sigma_Z} - Var(X)}$$

with its empirical counterparts one obtains an upper bound. See Adam and Rosen (2008) for a more general version of this derivation.

## 7.2 Construction of the lower bound when miscarriage is used as an instrument

Here we derive the lower bound when using miscarriage as an instrument. The probability limit of the standard IV estimator is given by

$$\beta_Z^{IV} = \beta + \frac{\sigma_{ZU}}{\sigma_{XZ}}$$

The expression $\frac{\sigma_{ZU}}{\sigma_{XZ}}$ is negative because $\sigma_{XZ} < 0$ and $\sigma_{ZU} > 0$. Hence, $\beta_Z^{IV}$ provides a lower bound when using the occurrence of a miscarriage as an instrument.

## 7.3 Construction of the upper bound when age at menarche is used as an instrument

Here we derive the upper bound when using the negative of age at menarche as an instrument. The probability limit of the standard IV estimator is given by

$$\beta_Z^{IV} = \beta + \frac{\sigma_{ZU}}{\sigma_{XZ}}$$

The expression $\frac{\sigma_{ZU}}{\sigma_{XZ}}$ is positive because both the denominator and the numerator are positive. Hence the conventional IV estimator provides an upper bound.

## 7.4 Imbens and Manski CI, not intended for publication

Table 4: Set Estimates and 95% Confidence Intervals using Each Instrument Individually.

| | Age at menarche | | Miscarriage<br>Imbens & Manski CI | | OLS |
|---|---|---|---|---|---|
| | Lower Bound | Upper Bound | Lower Bound | Upper Bound | |
| Panel A: Complete Sample. N=6443 | | | | | |
| No covariates | $-\infty$ | 0.218 | -0.501 | 0.310 | 0.368*** |
| 95% CI | | 0.466 | -0.896 | 0.346 | (0.017) |
| With covariates | $-\infty$ | 0.146 | -0.428 | 0.239 | 0.291*** |
| 95% CI | | 0.434 | -0.775 | 0.275 | (0.017) |
| Panel B: Ever pregnant sample. N=4810 | | | | | |
| No covariates | $-\infty$ | 0.108 | 0.052 | 0.306 | 0.338*** |
| 95% CI | | 0.344 | -0.151 | 0.343 | (0.018) |
| With covariates | $-\infty$ | 0.089 | -0.039 | 0.229 | 0.263*** |
| 95% CI | | 0.359 | -0.240 | 0.266 | (0.018) |
| Panel C: Teenage pregnancy sample. N=1284 | | | | | |
| No covariates | $-\infty$ | -0.048 | -0.117 | 0.156 | 0.253*** |
| 95% CI | | 1.209 | -0.261 | 0.216 | (0.028) |
| With covariates | $-\infty$ | 0.114 | -0.115 | 0.094 | 0.174*** |
| 95% CI | | 1.204 | -0.255 | 0.155 | (0.029) |
| Panel D: Have kids sample. N=3910 | | | | | |
| No covariates | $-\infty$ | 0.050 | -0.036 | 0.286 | 0.320*** |
| 95% CI | | 0.315 | -0.312 | 0.327 | (0.020) |
| With covariates | $-\infty$ | -0.096 | -0.165 | 0.206 | 0.243*** |
| 95% CI | | 0.258 | -0.467 | 0.248 | (0.020) |

Note: Confidence intervals based on robust standard errors.

## 7.5 Woutersen CI, not intended for publication

Table 5: Set Estimates and 95% Confidence Intervals using Each Instrument Individually.

| | Age at menarche | | Miscarriage Woutersen CI | | OLS |
|---|---|---|---|---|---|
| | Lower Bound | Upper Bound | Lower Bound | Upper Bound | |
| Panel A: Complete Sample. N=6443 | | | | | |
| No covariates | $-\infty$ | 0.218 | -0.501 | 0.310 | 0.368*** |
| 95% CI | | 0.466 | -0.896 | 0.346 | (0.017) |
| With covariates | $-\infty$ | 0.146 | -0.428 | 0.239 | 0.291*** |
| 95% CI | | 0.434 | -0.775 | 0.275 | (0.017) |
| Panel B: Ever pregnant sample. N=4810 | | | | | |
| No covariates | $-\infty$ | 0.108 | 0.052 | 0.306 | 0.338*** |
| 95% CI | | 0.344 | -0.151 | 0.345 | (0.018) |
| With covariates | $-\infty$ | 0.089 | -0.039 | 0.229 | 0.263*** |
| 95% CI | | 0.359 | -0.240 | 0.267 | (0.018) |
| Panel C: Teenage pregnancy sample. N=1284 | | | | | |
| No covariates | $-\infty$ | -0.048 | -0.117 | 0.156 | 0.253*** |
| 95% CI | | 1.209 | -0.261 | 0.216 | (0.028) |
| With covariates | $-\infty$ | 0.114 | -0.115 | 0.094 | 0.174*** |
| 95% CI | | 1.204 | -0.255 | 0.155 | (0.029) |
| Panel D: Have kids sample. N=3910 | | | | | |
| No covariates | $-\infty$ | 0.050 | -0.036 | 0.286 | 0.320*** |
| 95% CI | | 0.315 | -0.311 | 0.331 | (0.020) |
| With covariates | $-\infty$ | -0.096 | -0.165 | 0.206 | 0.243*** |
| 95% CI | | 0.258 | -0.466 | 0.251 | (0.020) |

Note: Confidence intervals based on robust standard errors.

Table 6: Set Estimates and Bootstrapped 95% Confidence Intervals using Each Instrument Individually.

| | Age at menarche | | Miscarriage Bootstrapped CI | | OLS |
|---|---|---|---|---|---|
| | Lower Bound | Upper Bound | Lower Bound | Upper Bound | |
| Panel A: Complete Sample. N=6443 | | | | | |
| 2SLS | | | | | |
| No covariates | $-\infty$ | 0.218 | -0.501 | 0.310 | 0.368*** |
| 95% CI | | 0.464 | -0.973 | 0.345 | (0.017) |
| With covariates | $-\infty$ | 0.146 | -0.428 | 0.239 | 0.291*** |
| 95% CI | | 0.439 | -0.827 | 0.274 | (0.017) |
| With covariates saturated | $-\infty$ | 0.105 | -0.458 | 0.228 | 0.283*** |
| 95% CI | | 0.354 | -0.754 | 0.257 | (0.017) |
| Non-parametric IV | | | | | |
| NN match | -$\infty$ | 0.173 | -0.442 | 0.283 | |
| 95% CI | | 0.445 | -0.884 | 0.317 | |
| Panel B: Ever pregnant sample. N=4810 | | | | | |
| No covariates | $-\infty$ | 0.108 | 0.052 | 0.306 | 0.338*** |
| 95% CI | | 0.347 | -0.158 | 0.344 | (0.018) |
| With covariates | $-\infty$ | 0.089 | -0.039 | 0.229 | 0.263*** |
| 95% CI | | 0.366 | -0.251 | 0.267 | (0.018) |
| With covariates saturated | $-\infty$ | 0.102 | -0.064 | 0.220 | 0.257*** |
| 95% CI | | 0.331 | -0.226 | 0.249 | (0.018) |
| Non-parametric IV | | | | | |
| NN match | $-\infty$ | 0.039 | 0.010 | 0.273 | |
| 95% CI | | 0.373 | -0.203 | 0.312 | |
| Panel C: Teenage pregnancy sample. N=1284 | | | | | |
| No covariates | $-\infty$ | -0.048 | -0.117 | 0.156 | 0.253*** |
| 95% CI | | 2.172 | -0.267 | 0.215 | (0.028) |
| With covariates | $-\infty$ | 0.114 | -0.115 | 0.094 | 0.174*** |
| 95% CI | | 2.244 | -0.261 | 0.155 | (0.029) |
| With covariates saturated | $-\infty$ | 0.166 | -0.116 | 0.091 | 0.171*** |
| 95% CI | | 1.808 | -0.235 | 0.141 | (0.030) |
| Non-parametric IV | | | | | |
| NN match | $-\infty$ | 0.259 | -0.080 | 0.150 | |
| 95% CI | | 2.006 | -0.233 | 0.212 | |
| Panel D: Have kids sample. N=3910 | | | | | |
| No covariates | $-\infty$ | 0.050 | -0.036 | 0.286 | 0.320*** |
| 95% CI | | 0.307 | -0.317 | 0.330 | (0.020) |
| With covariates | $-\infty$ | -0.096 | -0.165 | 0.206 | 0.243*** |
| 95% CI | | 0.251 | -0.487 | 0.249 | (0.020) |
| With covariates saturated | $-\infty$ | -0.072 | -0.200 | 0.196 | 0.236*** |
| | | 0.307 | -0.449 | 0.231 | (0.020) |
| Non-parametric IV | | | | | |
| NN match | $-\infty$ | -0.157 | -0.136 | 0.249 | |
| 95% CI | | 0.359 | -0.490 | 0.294 | |

Note: 5000 bootstrap replications were used, 1000 bootstrap replications for saturated model.

| Semiparametric Binary Choice | | |
|---|---|---|
| miscarriage | age at menarche | both |
| Panel A: Complete Sample, N=6443 | | |
| -0.300** | -0.288** | -0.290** |
| (0.146) | (0.152) | (0.143) |
| Panel B: Ever Pregnant Sample, N=4810 | | |
| -0.310*** | -0.289** | -0.279** |
| (0.106) | (0.144) | (0.159) |
| Panel C: Teenage Pregnancy Sample, N=1284 | | |
| -0.089 | -0.010 | -0.087 |
| (0.056) | (0.209) | (0.058) |
| Panel D: Have Child, N=3910 | | |
| -0.301** | -0.300** | -0.286** |
| (0.147) | (0.146) | (0.163) |

# 8 Calculation of Lewbel estimator

## 8.1 Exogenous case

## 8.2 Endogenous case

For each bootstrap sample, I do the following steps:

### 8.2.1 Magnac & Maurin

| Magnac and Maurin | Magnac and Maurin | bivariate probit | bivariate probit |
|---|---|---|---|
| miscarriage | age at menarche | miscarriage | age at menarche |
| Panel A: Complete Sample, N=6443 | | | |
| [-18.87, -13.18] | [-5.85, -3.78] | -39.80 (15.60) | 3.00 (24.35) |
| Panel B: Ever Pregnant Sample, N=4810 | | | |
| [-8.13, -5.63] | [-3.62, -2.22] | -11.73 (7.45) | -1.89 (12.24) |
| Panel C: Teenage Pregnancy Sample, N=1284 | | | |
| [-5.59, -4.51] | [-505.35, -77.67] | -28.93 (20.01) | 42.04 (149.91) |
| Panel D: Have Child, N=3910 | | | |
| [ -16.60, -10.45] | [-11.74, -7.60] | -13.25 (5.11) | -10.73 (7.03) |