The socioeconomic consequences of young women's childbearing: Reconciling disparate evidence*

By: David C. Ribar

David C. Ribar. "The Socioeconomic Consequences of Young Women's Childbearing: Reconciling Disparate Evidence," *Journal of Population Economics* 12:4 (1999), 547-65. doi:10.1007/s001480050113

Abstract:
Recent studies have begun to examine rigorously the links between early childbearing and subsequent socioeconomic status. Prominent in this literature has been a set of analyses that have used sibling fixed effects models to control for omitted variables bias. These studies report that the siblings difference procedure leads to smaller estimates of the effects of teen fertility than does standard regression analysis. While it is well known that the siblings fixed effects procedure makes strong assumptions regarding the type of omitted variables and is not necessarily robust to alternative assumptions, the assumptions of the procedure have not been explicitly examined. This paper uses 1979-1992 data from the National Longitudinal Survey of Youth to compare estimates of the income and education consequences of teenage and young adult fertility from standard regression and siblings fixed effects models with estimates from more general, alternative siblings models.
JEL classification: J13
Key words: Fertility, siblings models

Article:
1. Introduction
Adolescent childbearing is widely thought to represent a severe social, economic, and health problem in the United States. Research indicates that early fertility is associated with numerous consequences for young women including lower educational attainment, reduced labor supply, diminished earnings, and an increased risk of impoverishment and welfare dependency. Recently, studies have begun to examine the statistical relationship between early childbearing and subsequent socioeconomic status more rigorously. These studies have applied a variety of techniques including sibling fixed effects (Bennett et al. 1995; Geronimus and Korenman 1992; Hoffman et al. 1993a), quasi-experimental methods (Bronars and Grogger 1994; Hotz et al. 1997; Olsen and Farkas 1989), instrumental and other endogenous variables procedures (Angrist and Evans 1996; Klepinger, Lundberg and Plotnick 1995a, b; Marini 1984; Ribar 1994; Rindfuss et al. 1980), and survival analysis and other dynamic procedures (Bennett et al. 1995; Ribar 1996a, b; Upchurch and McCarthy 1990) and generated a considerable range of estimated effects. Although this research has been very successful in demonstrating that estimates of fertility consequences are sensitive to alternative statistical assumptions, the range of results and lack of reconciliation has limited any one study's or approach's policy usefulness.

At first glance, the variation in results is not surprising given the heterogeneity in estimation methods. However, even if the particular approaches differ, more uniformity might still be expected because many of these methods address a similar statistical issue —namely, bias from the potential correlation between fertility and other

* An earlier version of this paper was presented at the annual meetings of the Population Association of America, May 1994. The author gratefully acknowledges research support from the William T. Grant Foundation and from the National Institute of Child Health and Human Development under grant no. 1 P30 HD28263. He thanks Charlene Kalenkoski and Julie Kraut for excellent research assistance. He also thanks Bryan Boulier, Cliff Clogg, Jeff Grogger, Saul Hoffman, Joe Hotz, Don Kenkel, Sanders Korenman, Robert Mare, Mark Wilhelm, Ken Wolpin, Larry Wu, and an anonymous referee for helpful discussions. Responsible editor.• John F. Ermisch.

unobserved determinants of socioeconomic status. Unfortunately, the statistical procedures employed to date are not robust to alternative assumptions regarding the source of such correlation.

Simple sibling comparisons of the type employed by Geronimus and Korenman (1992) and Hoffman et al. (1993a) address biases that arise from shared unobserved family characteristics but do not address other potentially severe biases that arise from omitted person-specific factors or endogenous fertility. Instrumental variables (IV) procedures, which use exogenous variation in the observed determinants of childbearing to identify the consequences of fertility behavior, are a seemingly attractive alternative because these methods represent a general solution to the problems of both endogeneity and omitted variables. Practical difficulties arise, however, in finding instruments with sufficient predictive power that are also truly unrelated to other un-measured determinants of socioeconomic status.

This study develops and estimates a general model that incorporates key features of both of the above approaches and nests these features in a single framework. The nested approach is useful for clarifying the underlying assumptions of the different methodologies and for illustrating the effects these assumptions have on the resulting estimators. The study shows that by adopting some plausible, general assumptions, it can greatly narrow the set of possible estimates. Thus, the model can be used not only to replicate previous findings but also to reconcile them.[1]

The particular model this study considers is a multiple equation regression model of young sister's incomes and educations in which estimates of the effects of teenage childbearing are identified using alternative covariance restrictions on the unobserved determinants of these outcomes across sisters. One set of covariance restrictions forces the own and cross-sister correlations between the unobserved determinants of fertility and subsequent socioeconomic status to be equal and captures the key assumptions of the family fixed effects approach. Another set (that restricts the cross-sister correlations to be zero but leaves the own correlation terms unrestricted) replicates the assumptions of an IV procedure in which each sister's fertility behavior acts as an instrument for the other's childbearing. More general models that leave both the own and cross-sister correlation terms unrestricted are unidentified and therefore not estimable. Instead, the study develops a specification in which the unknown relationship between own and cross-sister correlations is described by a single parameter. With this specification, the study can characterize the estimated effects of fertility over a wide range of own and cross-sister correlation combinations. The various specifications are estimated using data on pairs of sisters from the 1979-1992 panels of the National Longitudinal Survey of Youth (NLSY).

The rest of this article is organized as follows. Section 2 reviews economic explanations for why teenage childbearing might be negatively associated with young women's socioeconomic well-being and motivates the subsequent empirical analysis. It also discusses statistical issues associated with modeling the effects of fertility. The study's econometric model is described in Sect. 3. Section 4 describes the analysis data set. Estimation results and specification comparisons are reported and discussed in Sect. 5. Section 6 presents conclusions as well as directions for future research.

2. Modeling issues

Theoretical considerations

Household production theory (Becker 1965; Gronau 1973) offers one explanation for a direct negative effect of fertility on young women's economic status. According to this theory, the responsibilities associated with the care of young children increase the opportunity cost of participating in job market activities and thereby decrease employment and earnings. For women who do work in the market, home responsibilities may increase absenteeism and otherwise detract from the level of effort on a particular job; such a decrease in productivity would also reduce earnings. As children age, the time required for their care may decrease suggesting that the direct economic consequences of fertility may only be temporary.

Indirect costs of childbearing, which might arise from an effect of fertility on women's investments in human capital (Becker 1993), are also potentially important. Again looking at the opportunity costs of time, early fertility may detract from a young woman's schooling as well as her work effort. To the extent that schooling and work experience are associated with better economic opportunities later in life, diminished investment in these activities will decrease subsequent economic success. Teen childbearing has also increasingly come to mean out-of-wedlock childbearing. The presence of children from previous relationships may decrease young women's marriage opportunities, and unions for those who do wed may be unstable or involve partners with poor economic prospects. In either case, factors related to marriage may lead to diminished family incomes.

Economic theory can also be used to examine the behavioral component of fertility. Simply stated, theory posits that women apply their personal preferences to balance the direct and opportunity costs of childbearing with its benefits (see Becker 1981 or Montgomery and Trussell 1986 for detailed theoretical discussions). If women are forward-looking, the economic approach also implies that they consider both the short- and long-run implications of potential decisions. In this framework, women with poor schooling or employment prospects may face diminished relative costs of childbearing and be more likely to bear children at an early age. If this is the case, economic opportunities might well drive childbearing rather than the other way around. While negative associations between fertility and socioeconomic outcomes are consistent with economic theory, positive associations are also possible.

The theory can be extended to incorporate additional hypotheses regarding how teenage and young adult women come to be productive in schools, work, and household activities as well as theories from outside economics about how preferences and values are formed (e.g., how the home, neighborhood, or school environment while growing up affect attitudes). These extensions generally do not firm up the predictions of the model; they do, however, enlarge the set of variables that might be considered relevant to an empirical analysis and further alert us to the possibility of problems associated with omitted variables.

Empirical considerations

The goal of the empirical literature on the consequences of adolescent childbearing has been to estimate relationships along the lines of

$$Y_i = \beta F_i + \psi' Z_i + \varepsilon_i \tag{1}$$

where Y_i is a socioeconomic outcome of interest for woman *i* at a particular point in time, F_i is an indicator for early childbearing, Z_i is a vector of other observed determinants of socioeconomic status, ε_i reflects the unobserved determinants, and β and ψ are coefficients to be estimated. Estimates of β describe the conditional association between fertility and socioeconomic status.

A substantial number of studies have used simple OLS and qualitative dependent variable methods to regress outcomes such as annual family income or educational attainment on indicators for early fertility and other observed characteristics. These studies have generally found that adolescent fertility has strong negative effects on these outcomes (see Hofferth 1987 for a review).[3] The statistical complications associated with this approach, however, are well-recognized. Simple regression techniques lead to biased estimates of β if the unobserved determinants of Y_i are correlated with fertility.

From the preceding theoretical discussion, there are numerous determinants that could give rise to such a correlation (e.g., factors related to work-and family-role attitudes, school and job abilities). Thus, bias in the coefficient estimates and, by extension, much of the existing literature seems very likely. However, because the theory is ambiguous about the predicted effects of these variables, the direction and magnitude of bias are open questions.

Indirect variables. To mitigate potential biases, most empirical research has been careful to include extensive sets of indirect controls for relevant omitted variables. While this strategy is intuitive and straightforward to implement, it is still an incomplete solution. To the extent that indirect controls leave some portion of ε_i correlated with fertility, substantial bias may remain.

Family fixed effects. A recent set of studies including Geronimus and Korenman (1992), Bennett et al. (1995) and Hoffman et al. (1993a) has applied siblings difference methods as a solution to omitted variables bias. To illustrate this method in terms of the model given above, decompose the error ε_i so that it consists of a factor ϕ_i that is also a determinant of fertility and another part e_i that is unrelated to fertility (let $\varepsilon_i = \phi_i + e_i$). These analyses posit that ϕ_i is a family-specific factor which does not vary across siblings and estimate variants of regression equation (1) that are differenced across sisters. Differencing sweeps out ϕ_i and eliminates the source of bias. The siblings studies find that the estimated consequences of early fertility are much smaller than those reported in standard analyses.

Properties of sibling estimators were reviewed by Griliches (1979) who cautioned that simple siblings comparisons are sensitive to alternative assumptions regarding the omitted factor ϕ_i. Importantly, bias is eliminated only if ϕ_i is identical across siblings. Although this assumption seems unlikely, the siblings difference approach might nevertheless be preferred if it reduces bias in cases where ϕ_i is highly but imperfectly correlated across siblings. Unfortunately, Griliches showed that bias in such instances may be exacerbated if ϕ_i is less highly correlated across siblings than the other unobserved determinants of fertility.

Instrumental variables. An alternative remedy to the problem of omitted variables bias involves the use of instrumental variables. The most common application of the IV approach involves directly instrumenting fertility using one or more variables which are strongly correlated with childbearing but otherwise unrelated to subsequent socioeconomic status.[4] Practical difficulties arise, however, in locating variables which satisfy both of these properties. More often than not, measures that are convincingly unrelated to the outcomes of interest end up being only modest predictors of fertility.[5] IV procedures have been used by numerous researchers to examine the effects of early fertility on outcomes such as birth weight (Rosenzweig and Schultz 1983; Grossman and Joyce 1990) and educational attainment (Angrist and Evans 1996; Klepinger et al. 1995a; Marini 1984; Olsen and Farkas 1989; Ribar 1994; Rindfuss et al. 1980) and have generated mixed evidence regarding these effects.

3. Econometric model

This study examines the standard regression, family fixed effects and IV estimation approaches in the context of a general siblings model. To develop the siblings model, it is useful to alter the notation slightly. Let N denote the number of sister pairs; let i ($\in$ 1, N) index families, and let j ($\in$ 1, 2) index sisters within families. Also, let f_{ij} and y_{ij} denote deviations from the means (taken across all sisters and families) of fertility and subsequent socioeconomic status. A simple two-equation model for each sister's fertility and subsequent socioeconomic status can be written

$$\begin{aligned} f_{ij} &= v_{ij} \\ y_{ij} &= \beta f_{ij} + \varepsilon_{ij} = \beta v_{ij} + \varepsilon_{ij} \end{aligned} \qquad \text{for } i = 1, N \quad \text{and} \quad j = 1, 2. \qquad (2)$$

For purposes of illustration, the present model abstracts from other observed determinants of fertility and socioeconomic status. Instead, the variation in fertility depends only on a random variable, v_{ij} and the variation in socioeconomic status depends only on changes in fertility and the random variable, ε_{ij} It is straightforward to incorporate other observed controls, and these are included in later analyses.[6]

By construction, the mean for each of the random variables is zero. Denote the variances and covariances of the random variables as follows:

$$\begin{aligned}&\text{Var}(v_{ij}) = \sigma_v^2, \quad \text{Var}(\varepsilon_{ij}) = \sigma_\varepsilon^2, \quad \text{Cov}(v_{ij}, \varepsilon_{ij}) = \sigma_{v\varepsilon} \quad \text{for}\\ &i = 1, N \quad \text{and} \quad j = 1, 2\\ &\text{Cov}(v_{ij}, v_{ik}) = \tilde{\sigma}_v^2, \quad \text{Cov}(\varepsilon_{ij}, \varepsilon_{ik}) = \tilde{\sigma}_\varepsilon^2, \quad \text{Cov}(v_{ij}, \varepsilon_{ik}) = \tilde{\sigma}_{v\varepsilon} \quad \text{for}\\ &i = 1, N; \quad j, k = 1, 2 \quad \text{and} \quad j \neq k.\end{aligned} \tag{3}$$

The variances and covariances for siblings' fertility and socioeconomic status can then be expressed

$$\begin{aligned}&\text{Var}(f_{ij}) = \sigma_v^2\\ &\text{Cov}(f_{ij}, y_{ij}) = \beta\sigma_v^2 + \sigma_{v\varepsilon} && \text{for } i = 1, N \text{ and } j = 1, 2\\ &\text{Var}(y_{ij}) = \beta^2\sigma_v^2 + 2\beta\sigma_{v\varepsilon} + \sigma_\varepsilon^2\\ &\text{Cov}(f_{ij}, f_{ik}) = \tilde{\sigma}_v^2\\ &\text{Cov}(f_{ij}, y_{ik}) = \beta\tilde{\sigma}_v^2 + \tilde{\sigma}_{v\varepsilon} && \text{for } i = 1, N; j, k = 1, 2 \text{ and } j \neq k.\\ &\text{Cov}(y_{ij}, y_{ik}) = \beta^2\tilde{\sigma}_v^2 + 2\beta\tilde{\sigma}_{v\varepsilon} + \tilde{\sigma}_\varepsilon^2.\end{aligned} \tag{4}$$

Method of Moments (MoM) estimators for the parameters of the model can be constructed by setting the theoretical moments from (4) equal to the corresponding sample moments for the siblings' fertility and socioeconomic status and solving for the theoretical parameters. Note, however, that it is not possible to identify all of the parameters in this particular specification because there are more parameters (seven) than moment conditions (six). Restrictions on the parameters are necessary for identification.

This study considers several versions of the above model that impose alternative sets of parameter restrictions. It starts with restrictions that capture key properties of the standard regression, siblings difference, and IV methods. It then returns to a variant of the general specification and shows how each of the estimators can be related.[7]

Standard regression specification. The key assumption of the standard regression approach is that the random determinants of each individual's fertility and socioeconomic status are uncorrelated, i.e., that $\sigma_{\varepsilon v} = 0$. When this restriction is applied in specification (4), there is a unique solution for $\beta, \beta = \frac{Cov(f_{ij}, y_{ij})}{Var(f_{ij})}$. Let s_{fy} denote the sample individual-specific covariance between fertility and socioeconomic status, and let s_f^2 denote the sample variance of fertility. Then the MoM estimator is the familiar expression $\tilde{\beta}_R = s_{fy}/s_f^2$.

Approximate family fixed effects specification. Another specification of the model captures key assumptions of the family fixed effects approach. Unlike the standard regression model, the family fixed effects model allows for some degree of correlation between ε_{ij} and v_{ij}. In particular, it specifies the individual and cross-sibling covariances between ε_{ij} and the random determinants of fertility to be equal such that $\sigma_{\varepsilon v} = \tilde{\sigma}_{\varepsilon v}$. This restriction is nearly equivalent to assuming that the unobserved determinants of socioeconomic status can be decomposed into afamily-specific random variable related to early fertility and another independent random component (as described in Sect. 2).[8] With this restriction, there is once again a unique solution for β,

$$\beta = \frac{\text{Cov}(f_{ij}, y_{ij}) - \text{Cov}(f_{ij}, y_{ik})}{\text{Var}(f_{ij}) - \text{Cov}(f_{ij}, f_{ik})} \tag{5a}$$

and a resulting MoM estimator

$$\hat{\beta}_{FE} = \frac{s_{fy} - \tilde{s}_{fy}}{s_f^2 - \tilde{s}_f^2}. \tag{5b}$$

This estimator is equivalent to applying OLS to a version of the second equation from specification (2) in which the socioeconomic and fertility variables are differenced across sisters.

<u>Sisters IV specification.</u> The sisters IV specification restricts the cross-sibling correlation $\tilde{\sigma}_{\varepsilon v}$ to be zero but leaves the individual-specific correlation $\sigma_{\varepsilon v}$ unrestricted. In this specification, each sister's fertility effectively serves as an instrument for the other's childbearing behavior. With this restriction, the solution for β is β = $\frac{Cov(f_{ij}, y_{ik})}{Cov(f_{ij}, f_{ik})}$, and the resulting MoM estimator is $\tilde{\beta}_{IV} = \frac{\tilde{s}_{fy}}{\tilde{s}_f^2}$.

<u>General specification.</u> To obtain a tractable expression for $\tilde{\beta}$ under more general conditions, reparameterize the cross-sibling covariance $\tilde{\sigma}_{\varepsilon v}$ as a proportion of the individual-specific covariance such that $\tilde{\sigma}_{v\varepsilon} = \rho\sigma_{\varepsilon v}$ With this change, the MoM estimator becomes

$$\hat{\beta} = \frac{\rho s_{fy} - \tilde{s}_{fy}}{\rho s_f^2 - \tilde{s}_f^2}. \tag{6}$$

Table 1. Own and cross-sibling correlations and covariances for fertility and socioeconomic outcomes

Outcome	Mean	Variance	Own teenage birth		Sister's teenage birth	
			Correlation	Covariance	Correlation	Covariance
Birth before age 20	0.2351	0.1800	–	–	0.1758	0.0316
Log family income-to-needs ratio	0.6490	0.7332	−0.3252	−0.1181	−0.2457	−0.0892
Log family income	9.9647	0.7070	−0.2078	−0.0741	−0.2066	−0.0737
Completed years of education[a]	12.8793	4.3051	−0.4204	−0.3721	−0.2724	−0.2411

Note: Statistics calculated using sister-pairs data from the 1979–1992 panels of the NLSY. Except for statistics involving years of education, statistics are based on 1174 individual (587 sister-pair) observations.
[a] Statistics calculated using 1268 individual (634 sister-pair) observations.

Because ρ is unknown, the estimator is unidentified. Expression (6) is nevertheless useful because it indicates what $\tilde{\beta}$ would be given any alternative assumption for ρ. For instance, the expression simplifies to the fixed effects estimator if ρ = 1, the sisters IV estimator if p = 0, and the standard regression estimator as $|\rho| \to \infty$. More generally, plausible restrictions on ρ might help to bound β.

To examine the implications of different restrictions on ρ, consider variances and covariances of sisters' teenage childbearing and young adult log income-to-needs ratios calculated from the National Longitudinal Survey of Youth. Figures from Table 1 indicate that the sample variance of early fertility outcomes (s_f^2) is 0.1800, the covariance of fertility outcomes across sisters ($\tilde{s}_f^2$) is 0.0316, the individual-specific covariance between fertility and the income-to-needs measure (s_{fy}) is —0.1181, and the covariance between fertility and income-to-needs across sisters ($\tilde{s}_{fy}$) is —0.0892. When we enter these numbers into the estimation formulas, we see that the alternative methods lead to dramatically different estimates of the effect of early fertility on subsequent socioeconomic status. The standard regression estimate of this effect is —0.66. The family fixed effects estimate is considerably smaller at —0.19, while the sisters IV estimate is considerably larger at —2.82.

Figure 1 graphs estimates of β calculated for all ρ from —1 to 2. The most striking feature of Fig. 1 is the asymptote at ρ = 0.18. The asymptote occurs at the point where the denominator of expression (6) is zero (where ρ equals the correlation in fertility outcomes across sisters). As ρ approaches 0.18 from below, the estimate of β falls farther below the standard regression estimate; as p approaches 0.18 from above, the opposite occurs. Clearly, the standard regression, family fixed effects and sisters IV estimators do not lie along a continuum in ρ.

MoM estimators for $\sigma_{v\varepsilon}$ and $\tilde{\sigma}_{\varepsilon v}$ can also be formed conditional on ρ. The sample data are consistent with three conditions on these terms. For all ρ less than zero, the individual-specific covariance is positive and the cross-sibling covariance is negative. For ρ between zero and 0.18, both covariances are

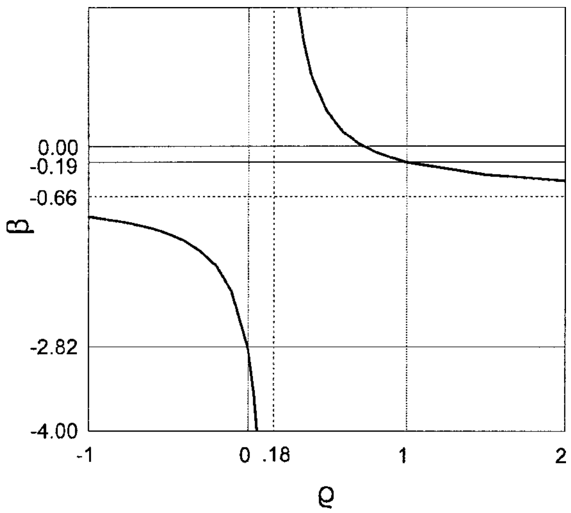

Fig. 1. Alternative estimates of β conditional on ρ

positive, while for ρ greater than 0.18, both covariances are negative. Returning to Fig. 1, these conditions imply that sign restrictions on the covariance terms can be used to bound the estimate of β. For instance, if we assume that the unobserved family-specific determinants of fertility and the income-to-needs ratio are negatively related ($\tilde{\sigma}_{v\varepsilon} < 0$), the sisters IV estimate of —2.82 is a lower bound for $\tilde{\beta}$. If we instead assume that the unobserved individual-specific determinants of fertility and the income-to-needs ratio are negatively related ($\sigma_{v\varepsilon} < 0$), the standard regression estimate of —0.66 becomes the lower bound. Finally, if we assume that both sets of determinants are negatively related but that the individual-specific determinants are more strongly negatively related than the family-specific determinants ($\sigma_{v\epsilon} \leq \tilde{\sigma}_{v\varepsilon} < 0$), the family fixed effects estimate becomes the lower bound. While this last restriction might seem arbitrary, it is, in fact, plausible. The restriction holds if the unobserved determinants of fertility can be grouped into a single variable that is (a) negatively associated with socioeconomic status and (b) positively but perhaps imperfectly correlated across siblings.

As this discussion shows, p conveniently parameterizes alternative assumptions regarding the influence of unobserved individual- and family-specific factors. Restrictions on this parameter (and hence on the underlying assumptions) can be used to bound estimates of the effects of teenage fertility. While this article focuses on a few restrictions in the context of teenage childbearing, the estimation approach can be applied to other situations where siblings data or short panels have been used to account for biases from omitted variables.

4. Data

The primary data for this analysis come from the 1979-1992 panels of the National Longitudinal Survey of Youth (Center for Human Resource Research 1994). The NLSY is a national sample of 12,686 individuals who were 14 to 21 years old in 1979.[9] Individuals have been re-interviewed annually since 1979. The survey contains detailed longitudinal demographic and economic information including data on fertility, schooling, and family and individual incomes. Personal and family background data are also available. Considerable effort has been directed toward minimizing sample attrition; consequently, retention through the 1992 panel is roughly 90%.

For each household sampled in 1979, interviews were conducted and relationship codes, recorded for every appropriately aged individual in the household. Thus, the NLSY supports construction of a moderate-sized sub-sample of near-age siblings.[10] From the 1979 panel, there are 775 households with two or more sisters present.[11] After excluding observations with missing information on fertility and the exogenous variables, the sample size is reduced to 724 households (excluding observations with missing socioeconomic outcome information further reduces the sample). For households with more than two sisters, the paper follows Geronimus and Korenman (1992) and examines the two oldest sisters.[12]

Three variables are used to describe each woman's socioeconomic wellbeing as a young adult —total annual family income, the income to needs ratio (income divided by the poverty level for the woman's reported family size), and years of educational attainment. To maximize the consistency of the measures across women, age 24 data have been selected.[13] Because these data extend across several years, dollar-denominated outcomes have been deflated to 1992 values using the Personal Consumption Deflator.

A binary variable indicating whether the woman experienced a birth prior to age 20 is used as the study's measure of early fertility. While this simple indicator has limitations (e.g., it does not distinguish between women who had one teenage birth and multiple teenage births), it is easily interpreted and comparable to measures used in other studies.

Standard explanatory variables available from the NLSY include the woman's age in 1979 as well as her ethnic origin. Detailed family background data have also been collected. Among these data are measures of each parents' educational attainment (variables for total years of completed schooling, years of post-secondary schooling, and an indicator for missing information), family structure (number of siblings and an indicator for residence in a non-intact household), and an indicator for whether anyone in the household received magazines.

The NLSY also includes longitudinal geographic data and descriptors for local economic conditions. The paper uses 4- and 5-year averages (from ages 16 to 19 and ages 20 to 24) of annual rural urban dummy variables to approximate the percentage of time women spent in metropolitan areas as teenagers and young adults.[14] A similar 5-year average of county-level joblessness rates is used to describe women's employment opportunities as young adults. Geographic identifiers are used to link women with external county-level data on average earnings for retail workers (U.S. Bureau of Economic Analysis 1994). The 5-year average for this variable is taken to capture local wage opportunities.

Means and standard deviations of the analysis variables for the sister pairs sample are listed in Appendix A. The statistics, which are unweighted, evidence the effects of oversampling in the NLSY —women of African and Hispanic origin are over-represented as is the incidence of teenage fertility.

Table 2. Own and cross-sibling correlations and covariances for fertility and socioeconomic residuals

Outcome	Variance	Own teenage birth		Sister's teenage birth	
		Correlation	Covariance	Correlation	Covariance
Birth before age 20	0.1549	–	–	0.0795	0.0123
Log family income-to-needs ratio	0.5798	−0.2304	−0.0690	−0.1292	−0.0387
Log family income	0.6040	−0.1214	−0.0371	−0.1069	−0.0327
Completed years of education[a]	3.0777	−0.3321	−0.2285	−0.1484	−0.1021

Note: Statistics calculated using residuals from regressions of listed variables on the basic and family background control variables described in Appendix A. Except for statistics involving years of education, statistics are based on 1174 individual (587 sister-pair) observations.

[a] Statistics calculated using 1268 individual (634 sister-pair) observations.

This suggests that caution should be applied in generalizing the results from this analysis. Comparisons (not shown) with statistics from a general sample of young women from the NLSY indicate that, beyond some obvious characteristics like the number of siblings, there are few differences between the samples. [15]

5. Empirical findings

To show the general relationships among the analysis variables, means, variances, individual-specific correlations and covariances, and cross-sibling correlations and covariances for the fertility and socioeconomic measures are calculated and reported in Table 1. All of the correlation coefficients in Table 1 are significantly different from zero in the anticipated directions. Among the individual correlations, teenage childbearing is negatively related to education and the two income measures. The corresponding cross-sibling coefficients are weaker but also negative.

One surprise from these results is the relatively weak correlation between sisters' fertility outcomes. Although the correlation coefficient is significantly greater than zero, it is smaller in absolute terms than any of the cross-sibling correlations between fertility and socioeconomic status. At a minimum, the small correlation indicates that family-specific effects account for only a portion of the variation in fertility. It also indicates that sister's fertility may be a weak instrument for own fertility in the subsequent analyses.

Table 2 reports variances and covariance for residuals of the fertility and socioeconomic measures that have been purged of correlation with the exogenous controls listed in Appendix A. Specifically, each outcome variable was regressed against both the individual's and sister's measures of the control variables; residuals were then obtained from these initial regressions.[16] As one might expect, the variances and covariances among the residuals are all closer to zero than the corresponding figures from Table 1. The directions of the relationships, however, remain unchanged.

Table 3. Alternative estimates of the socioeconomic effects of teenage childbearing

	OLS with basic controls	OLS with basic and family background controls	Family fixed effects	Sisters IV
	Dependent variable: LOG FAMILY INCOME-TO-NEEDS RATIO			
Estimated effect of a teenage birth	−0.546***	−0.446***	−0.213***	−3.146***
	(0.055)	(0.056)	(0.074)	(1.209)
R^2	0.182	0.251	0.384	–
	Dependent variable: LOG FAMILY INCOME			
Estimated effect of a teenage birth	−0.316***	−0.240***	−0.031	−2.655**
	(0.057)	(0.058)	(0.079)	(1.142)
R^2	0.096	0.158	0.281	–
	Dependent variable: COMPLETED YEARS OF EDUCATION			
Estimated effect of a teenage birth	−1.953***	−1.486***	−0.890***	−8.695***
	(0.127)	(0.120)	(0.139)	(3.072)
R^2	0.202	0.364	0.606	–

Note: Models estimated using sister-pairs data from the 1979–1992 panels of the NLSY. Regressions also include intercepts and coefficients for basic and family background control variables as indicated in the column headings; for lists of the control variables see Appendix A. Estimated standard errors appear in parentheses.
* Significant at 0.10 level.
** Significant at 0.05 level.
*** Significant at 0.01 level.

Table 3 displays results from alternative regression specifications estimated using the NLSY sisters data. Like the study by Geronimus and Korenman (1992), estimates are reported from OLS regressions with basic controls (first column), regressions that add family background controls (second column), and regressions with fixed effect controls for family background (third column). Unlike previous studies, the table also includes results from IV models in which sister's teenage childbearing is used as an instrument for fertility (fourth column). To conserve space, the table displays only the coefficients on the teenage fertility variable and fit statistics for each specification.[17]

The results from the first set of models indicate that teenage childbearing is significantly negatively associated with the income-to-needs ratios, family incomes, and educational attainments of young women after controlling for race, birth-year cohort, and local labor market conditions. All of the coefficients are substantively large. They indicate that teenage fertility is associated with a 42% drop in the income-to-needs ratio, a 27% drop in family incomes, and a two year decrement in schooling.

When additional controls for family background are included in the second set of specifications, the estimated negative associations for teenage fertility are reduced, though they remain significant and substantively large (e.g., the coefficients indicate that teenage fertility is associated with a 36% drop in the income-to-needs ratio, a 21% drop in family incomes, and a 1½ years drop in schooling). The results from this second set of specifications are consistent with the findings of other standard regression analyses (see Hofferth's 1987 review). The difference in estimates between the models that do and do not include indirect controls is also similar to the pattern documented by Geronimus and Korenman (1992).

The third column in Table 3 reports results from sister-difference OLS models. Application of the fixed effects technique leads to large reductions in the magnitudes of estimated effects of teenage fertility. For the income-to-needs ratio and schooling models, the coefficients on teenage fertility are roughly half the size of the corresponding estimates from the previous column (i.e., implying a 19% reduction in the income-to-needs ratio and less than a year's reduction in schooling), though still substantively large. For the family income model, the estimated association for teenage fertility virtually disappears. These results confirm the general findings of Geronimus and Korenman (1992) and Hoffman et al. (1993a) that the use of family fixed effects greatly diminishes the estimated effects of teenage fertility but that some evidence of costs remains.

So far, the empirical analysis has simply replicated previous studies using a consistent data set (albeit with some improvements such as the measurement of outcomes at a consistent age) but with no fundamental changes to their estimation methodologies. The study now turns, in the fourth column, to results from siblings IV specifications. Estimates from the IV models are all much more strongly negative than the previous estimates and indicate that teenage childbearing has severe consequences for young women's income and schooling attainments. These results are similar to those reported by Klepinger et al. (1995a, b) who also found that an IV methodology led to substantially larger estimates of the consequences of early fertility. The results differ, however, from those of Ribar (1994), Olsen and Farkas (1989) and others who found that the IV approach led to weaker estimates.

A technical explanation for the strong negative IV results can be found in Table 2. When we move from the standard regression to the IV estimator, we replace the individual-specific covariance between the residuals for fertility and socioeconomic status in the numerator with the cross-sibling covariance and the individual-specific residual variance of fertility in the denominator with the cross-sibling covariance. The cross-sibling residual covariances between fertility and the income-to-needs, family income and education variables are 56, 88 and 45% of their respective individual-specific covariances while the cross-sibling residual covariance in fertility is only 8% of the individual-specific variance. The modest changes in the numerators coupled with the large changes in the denominators lead to the dramatic changes in the IV estimators.

Using formula (6), we can examine how the estimated effects of teenage fertility vary over a range of alternative assumptions regarding the relative strengths of the individual- and family-specific correlations in the unobserved determinants. As with the earlier illustrative example, I use graphs to present the estimation estimates. The graphs are shown in Fig. 2.

The first graph (a) in Fig. 2 displays alternative estimates of the effects of teenage fertility on the log income-to-needs variable; the second graph (b) displays estimates for the log family income variable, and the third graph (c) displays results for the education variable. All three graphs have the same general shape. All three have a vertical asymptote at $\rho = 0.0795$ and a horizontal asymptote corresponding to the standard regression estimate of β. For values of ρ less than 0.0795, the estimated effect of fertility is more negative than the standard regression estimate; for values of ρ greater than 0.0795, the

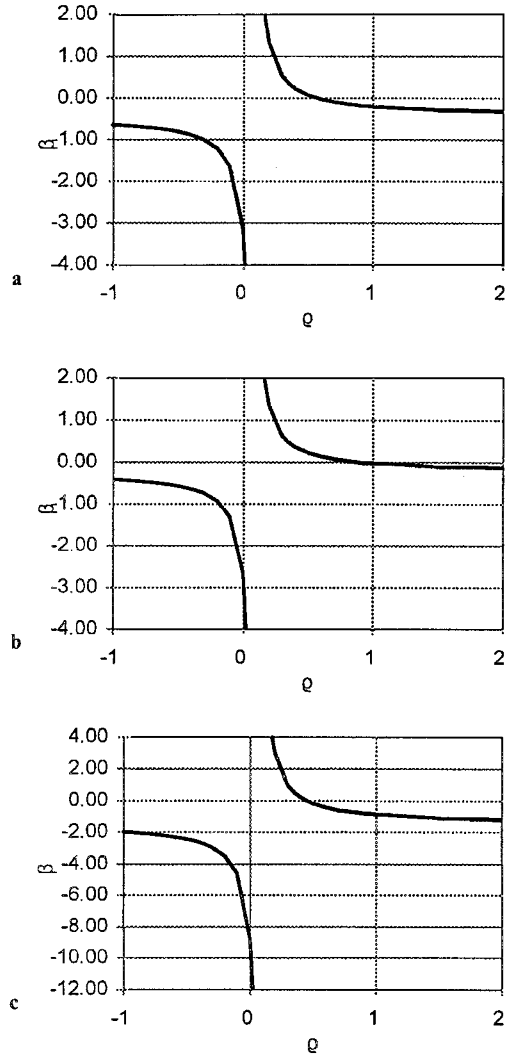

Fig. 2a–c. Alternative estimates of the effect of teenage fertility on socioeconomic status. **a.** Log family income-to-needs ratio; **b.** Log family income; **c.** Completed years of education

estimated effect of fertility is higher. If we carefully read the graphs we see that each reproduces the sisters IV estimate at $\rho = 0$ and the family fixed effects estimate at $\rho = 1$.

As in the illustrative example, restrictions on the individual- and family-specific correlation terms ($\sigma_{v\epsilon}$ and $\tilde{\sigma}_{v\varepsilon}$) can be used to bound the estimates. If we assume that the individual-specific correlations between the unobserved determinants of teenage fertility and the measures of subsequent socioeconomic status are negative, estimates of β are bounded from below by the standard regression estimates. Among other things, this restriction rules out the strongly negative sisters IV estimates.

If we further assume that the person-specific correlations are stronger than the family-specific correlations, the lower bounds are the family fixed effects estimates. Note that these estimates are still consistent with substantial costs of childbearing. The smallest (in absolute terms) estimate —the coefficient for teenage fertility in the log family income regression — is consistent with a 17% loss at the lower end of its 95% confidence interval. Nevertheless, given the plausibility of the restriction, the results do suggest that the fixed effects estimates reported by Geronimus and Korenman (1992) be treated as lower bounds for the estimated effects of teenage fertility. They also suggest the results of Ribar (1994), Hotz et al. (1997), and others who report smaller consequences than Geronimus and Korenman be given greater credence.

6. Conclusion

This study develops and estimates several specifications of an endogenous variable model of sisters' socioeconomic status and childbearing. The alternative specifications incorporate covariance restrictions that capture important properties of the standard regression, sister difference, and instrumental variables approaches. A more general specification that uses a single parameter to describe the range of alternative covariance restrictions is also estimated. The specifications are employed in an effort to replicate and reconcile previous estimates of the income and schooling consequences of early childbearing.

The empirical analysis replicates many results of both the early and recent literature. Specifically, standard regression models similar to those adopted by early studies produce large estimates of the consequences of fertility. As with the analyses by Geronimus and Korenman (1992) and Hoffman et al. (1993a), these effects are greatly reduced in models that use sibling difference controls for family-specific omitted variables, though some evidence of consequences remains. Application of instrumental variable methods that account for individual-specific omitted variables leads to much stronger negative estimates of the effects of teenage fertility.

Analysis of the general model reveals that assumptions on the directions and relative strengths of the individual- and family-specific covariances between the unobserved determinants of fertility and socioeconomic status can be used to bound the estimated effects of early childbearing. For instance, if we make the reasonable assumption that the unobserved determinants of fertility and socioeconomic status are negatively related, the strongly negative instrumental variable estimates are ruled out. If we further assume that the unobserved individual-specific factors are at least as strongly related as the unobserved family-specific factors, then the sibling difference estimates represent alower bound on the estimated effects of fertility. While the general model does not lead to a specific, preferred point estimate, it does suggest that the range of acceptable estimates can be considerably narrowed.

Appendix

Table A. Variable means

Variables	Mean	(Std. Dev.)
Fertility and socioeconomic outcome variables		
Birth before age 20	0.235	(0.424)
Ln(family income)	9.965	(0.841)
Ln(income-to-needs ratio)	0.649	(0.856)
Years of schooling[a]	12.879	(2.075)
Basic control variables		
Age in 1979	17.156	(1.985)
African origin	0.284	(0.451)
Hispanic origin	0.157	(0.364)
Urban residence (from age 20–24)	0.789	(0.379)
Retail earnings (000s, from age 20–24)	15.856	(1.932)
Unemployment rate (from age 20–24)	8.387	(2.899)
Family background variables		
Number of siblings	4.537	(2.666)
Nonintact family at age 14	0.274	(0.446)
Mother's education missing	0.043	(0.204)
Mother's total years of schooling	10.319	(3.753)
Mother's years of post-secondary schooling	0.440	(1.220)
Father's education missing	0.118	(0.322)
Father's total years of schooling	9.399	(5.173)
Father's years of post-secondary schooling	0.762	(1.769)
Household received magazines	0.532	(0.499)
Urban residence (from age 16–19)	0.778	(0.388)
Individual observations	1174	

Notes: Statistics calculated using on sister-pairs data from the 1979–1992 panels of the NLSY.
[a] Education figures based on 1268 individual observations.

Endnotes

1. 1 In their summary of a conference (and the then-existing literature) on the consequences of early fertility, Bachrach and Carver (1992, p. 21) commented about the need for a reconciliation study:

A critical comparison of methods used to control for unobserved heterogeneity in estimating the effects of a given behavior or condition is needed.... [U]nder nonexperimental conditions such heterogeneity between those who exhibit a behavior and those who do not is important, and needs to be accounted for. The methods used to accomplish this differ in their underlying assumptions and in the results they produce. Further work is needed to evaluate and improve these methods.

To the author's knowledge, there has only been one other study of early fertility that has nested the siblings fixed effects methodology with other approaches. Rosenzweig and Wolpin (1995) combined cross-sibling and cross-cousin comparisons to examine the consequences of early childbearing on children's birth outcomes.

2. The percentage of non-marital births among teenage mothers increased from 48% in 1980 to 69 percent in 1991 (Moore 1994).
3. Hofferth also describes studies that report positive effects on outcomes such as labor force participation and entry into marriage.
4. In terms of the error decomposition described earlier, another approach would be to instrument ϕ_i; by combining measures from two or more indirect variables which are, once again, conditionally unrelated to Y_i. For instance, Griliches (1979) and Bound et al. (1986) describe estimators which use indirect measures of ϕ_i; for different siblings as IVs. The principal drawback of this approach (and reason it is not employed in this study) is that it requires explicit assumptions regarding the identity of ϕ_i;.
5. A natural approach to nesting the family fixed effects and IV methodologies would be to instrument fertility within a siblings difference model (see, e.g., Ribar and Wilhelm 1999 who successfully incorporated IVs into a fixed effects model of states' welfare spending). The difficulty in the present application is the lack of predictive power in the potential instruments. Recent IV studies (e.g., Angrist and Evans 1996; Klepinger et al. 1995a, b; Ribar 1994) have relied in whole or in part on identification from local area controls for access to or the costs of reproductive health services, variables which do not vary greatly across sisters.
6. To incorporate observed controls, specification (2) can be modified so that the dependent variables are residuals from initial regressions of each sister's fertility and socioeconomic status on a set of controls, instead of simple deviations from means.
7. Griliches (1979) analyzes several siblings models in terms of their covariance restrictions. Models based on covariance restrictions across family members have been used to examine the returns to schooling (Bound et al. 1986; see also Card's 1994 review) and the effects of teen parenthood on birth outcomes (Rosenzweig and Wolpin 1995).
8. The factor-analytic specifications additionally imply that $0 \leq \tilde{\sigma}_v^2 < \sigma_v^2$. This restriction was always met in the study's data.
9. In the initial survey, blacks, Hispanics, disadvantaged white youth, and military personnel were oversampled. Weights (not used here) are available to make the data nationally representative.
10. Hoffman et al. (1993a, b) and Geronimus and Korenman (1993) provided extensive discussions of the properties of this sampling strategy.
11. The number of multiple-sibling families is smaller than the number reported by Geronimus and Korenman (1992). Their analysis appears to have used all females in the household, not just sisters.
12. Beyond the identification of siblings, there are other differences between the methods used to construct the paper's sample and the methods used by previous studies. Geronimus and Korenman (1992) restricted their samples to include mothers only and examined outcomes for a particular year. Hoffman et al. (1993a) also examined outcomes for a uniform year rather than a uniform age.
13. The family income measures, which were constructed from several separate income variables by staff at the Center for Human Resource Research, have a moderately high number of missing observations. For women with missing income information at age 24, the study substituted age 25 data if they were available. The explanatory variables for these women were adjusted to capture an additional year of information.
14. Women who were older than 16 in 1979 are assigned their 1979 urban rural status for earlier years.

15. Concerns about sample comparability arise because of the documented consequences of family size on children's socioeconomic attainment and recent evidence from Butcher and Case (1994) that the presence of sisters reduces schooling among women.
16. The residuals were purged of correlations with observed controls for both the individual and her sister to make the MoM estimates fully comparable with siblings difference estimates (Chamberlain 1984).
17. Complete results are available from the author upon request.

References

Angrist J, Evans W (1996) Schooling and Labor Market Consequences of the 1970 State Abortion Reforms. NBER working paper no. 5406

Bachrach C, Carver K (1992) Outcomes of Early Childbearing: An Appraisal of Recent Evidence. Summary of a conference convened by The National Institute of Child Health and Human Development in Bethesda, MD

Becker G (1965) A Theory of the Allocation of Time Economic Journal75:493±517

Becker G (1981) A Treatise on the Family Harvard University Press, Cambridge, MA

Becker G (1993) Human Capital. A Theoretical and Empirical Analysis with Special Reference to Education 3rd edn. The University of Chicago Press Chicago, IL

Bennett N, Bloom D, Miller C (1995) The Influence of Nonmarital Childbearing on the Formation of First Marriages. Demography 32:1-16

Bound J, Griliches Z, Hall B (1986) Wages, Schooling and IQ of Brothers and Sisters: Do the Family Factors Differ? International Economic Review 21:77-105

Bronars S, Grogger J (1994) The Economic Consequences of Teenage Childbearing: Results from a Natural Experiment. American Economic Review 84:1141-1156

Butcher K, Case A (1994) The Effect of Sibling Sex Composition on Women's Education and Earnings. Quarterly Journal of Economics 109(3):531-564

Center for Human Resource Research (1994) NLS Handbook 1994. Ohio State University, Columbus

Card D (1994) Earnings, Schooling, and Ability Revisited. NBER working paper no. 4832 Chamberlain G (1984) Panel Data. In: Griliches Z, Intriligator M (eds) Handbook of Econometrics, Vol. II. Elsevier, Amsterdam pp. 1247-1318

Geronimus A, Korenman S (1992) The Socioeconomic Consequences of Teen Childbearing Reconsidered. Quarterly Journal of Economics 107:1187-1214

Geronimus A, Korenman S (1993) The Costs of Teenage Childbearing: Evidence and Interpretation, Demography 30:281-290

Griliches Z (1979) Siblings Models and Data in Economics: Beginnings of a Survey. Journal of Political Economy 87:S37-S64

Gronau R (1973) The Effect of Children on the Housewife's Value of Time. Journal of Political Economic 81:S168-S199

Grossman M, Joyce T (1990) Unobservables, Pregnancy Resolutions, and Birth Weight Production Functions in New York City. Journal of Political Economy 98:983-1007

Hofferth S (1987) The Social and Economic Consequences of Teenage Childbearing. In: Hayes C, Hofferth S (eds) Risking the Future.• Adolescent Sexuality, Pregnancy, and Childbearing, Vol. II. National Academy Press, Washington, DC

Hoffman S, Foster EM, Furstenberg F (1993a) Re-evaluating the Costs of Teenage Childbearing. Demography 30:1-13

Hoffman S, Foster EM, Furstenberg F (1993b) Reevaluating the Costs of Teenage Childbearing: Response to Geronimus and Korenman. Demography 30:291-296

Hotz VJ, McElroy S, Sanders S (1997) The Costs and Consequences of Teenage Childbearing for Mothers. In: Maynard R (ed) Kids Having Kids. Economic Costs and Social Consequences of Teen Pregnancy. Urban Institute Press, Washington, DC

Klepinger D, Lundberg S, Plotnick R (1995a) Adolescent Fertility and the Educational Attainment of Young Women. Family Planning Perspectives 27:23-28

Klepinger D, Lundberg S, Plotnick R (1995b) How Does Adolescent Fertility Affect the Human Capital and Wages of Young Women? Paper presented at the Population Association of America Meetings in San Francisco, CA
Marini M (1984) Women's Educational Attainment and the Timing of Entry into Parenthood. American Sociological Review 49:491-511
Montgomery M, Trussell J (1986) Models of Marital Status and Childbearing. In: Ashenfelter O, Layard R (eds) The Handbook of Labor Economics. North-Holland, Amsterdam
Moore K (1994) Facts at a Glance Child Trends, Washington, DC
Olsen R, Farkas G (1989) Endogenous Covariates in Duration Models and the Effect of Adolescent Childbirth on Schooling. Journal of Human Resources 24:37-53
Ribar D (1994) Teenage Fertility and High School Completion. Review of Economics and Statistics 76:413-424
Ribar D (1996a) The Effects of Teenage Fertility on Young Adult Childbearing. Journal of Population Economics 9(2):197-218
Ribar D (1996b) A Longitudinal Analysis of Young Women's Fertility and Educational Advancement. Paper presented at the Population Association of America Meetings in New Orleans, LA
Ribar D, Wilhelm M (1999) The Demand for Welfare Generosity. Review of Economics and Statistics 81:96-108
Rindfuss, R, Bumpass L, St. John C (1980) Education and Fertility: Implications for the Roles Women Occupy. American Sociological Review 45:431-447
Rosenzweig M, Schultz TP (1983) Estimating a Household Production Function: Heterogeneity, the Demand for Health Inputs, and Their Effects on Birth Weight. Journal of Political Economy 91:723-746
Rosenzweig M, Wolpin K (1995) Sisters, Siblings, and Mothers: The Effect of Teen-age Child-bearing on Birth Outcomes in a Dynamic Family Context. Econometrica 63(2):303-326
US Bureau of Economic Analysis (1994) Regional Economic Information System. US Department of Commerce, Washington, DC
Upchurch D, McCarthy J (1990) The Timing of a First Birth and High School Completion. American Sociological Review 55:224-234