National Longitudinal Surveys

U.S. Department of Labor
Bureau of Labor Statistics

Discussion Paper

The Intergenerational Effects of Fatherlessness on Educational Attainment and Entry-Level Wages

Jeff Grogger
Nick Ronan
University of California

September 1995

Report: NLS 96-30

The National Longitudinal Surveys (NLS) program supports many studies designed to increase understanding of the labor market and methods of data collection. The *Discussion Papers* series incorporates some of the research coming out of the program. Many of the papers in this series have been delivered as final reports commissioned as part of an extramural grant program, under which grants are awarded through a competitive process. The series is intended to circulate the findings to interested readers within and outside the Bureau of Labor Statistics.

Persons interested in obtaining a copy of any Discussion Paper, other NLS reports, or general information about the NLS program and surveys should contact the Bureau of Labor Statistics, Office of Economic Research, Washington, DC 20212-0001, or call (202) 606-7405.

The Intergenerational Effects of Fatherlessness on Educational Attainment and Entry-level Wages

Jeff Grogger
Nick Ronan

Department of Economics
University of California
Santa Barbara, CA 93106
(805) 893-4533
jeff@econ.ucsb.edu

Final Report to the Bureau of Labor Statistics,
September 1, 1995.

We thank workshop participants at the BLS for helpful comments. The opinions expressed in this report are solely those of the authors. They do not necessarily represent the positions of the Bureau of Labor Statistics, the Department of Labor, or any other government agency.

Executive Summary

In 1960, only five percent of all births occurred out of wedlock, and only 13 percent of all children lived in a single-parent family. By 1990, increasing divorce and out-of-wedlock childbearing had raised these figures substantially: more than one-fourth of all births were to unmarried mothers, and 27 percent of all children lived with only one parent. Bumpass (1984) predicted that, as a result of rising divorce and non-marital fertility, 50 percent of all children born as early as 1980 would spend part of their childhood in a fatherless family.

Assuming that workers enter the labor market at age 20, and that the workforce turns over every 40 years, these demographic trends imply that, within the next 20 years or so, roughly one-fourth of the labor force will have spent part of its childhood in a single-parent home. Put differently, this means that within the next couple of decades, there will be as many workers who grew up fatherless as workers who hold college degrees. The fraction of the labor force that was raised in a single-parent family will undoubtedly exceed the fraction that is unionized.

These trends in family structure may bode ill for tomorrow's labor force, at least if the predictions of economic theory are correct. In Becker's (1981) model, for example, children raised in families with fewer resources tend to have lower human capital. Thus economic theory would predict that, all else equal, the next generation of workers will enter the labor market with less human capital than the last.

The objective of this study is to estimate the effects of fatherlessness on the children's educational attainment and entry-level wages. We consider an important methodological issue not addressed by previous researchers: unobserved heterogeneity across families. One can imagine that families vary greatly in a number of ways that are unobservable to the analyst. Moreover, many of these unobservable family characteristics are likely to be correlated both with the probability of divorce and with the well-being of the children. Thus a cross-sectional regression of children's educational attainment on a measure of their childhood family structure fails to identify the effects of living in a fatherless family, because the effects of fatherlessness are confounded with the effects of the family-specific unobservables. We would generally expect such unobserved heterogeneity to lead to exaggerated estimates of the true effects of fatherlessness.

We adjust for family-specific unobservables by making within-family comparisons. Drawing on previous research, we specify a child's human capital to depend on the number of years she spends in a single-parent family. Because children enter and leave the family at different times, the duration of a spell of fatherlessness generally will vary among siblings. To eliminate the effects of family-specific unobservables, we difference the data within families, relating differences in human capital to differences in the duration of the fatherless spell.

If spells were measured accurately, then differencing within families would provide valid estimates of the effects of a year of fatherlessness. Our data on childhood living arrangements are measured retrospectively, however, and there is evidence of a fair amount of measurement error, particularly in the differenced data. Under standard assumptions, measurement error causes the estimated regression coefficients to be biased downward.

The usual solution to this problem, instrumental variables estimation, performs poorly in this case. The approach we adopt instead is method-of-moments estimation. We implement this approach by using sibling comparisons to estimate the extent of the measurement error in our retrospective data.

The data are taken from the NLSY. This rich longitudinal survey has several important features without which our analysis would be impossible. First, roughly half of its members have siblings who also took part in the survey. It therefore offers samples of siblings that are large enough for meaningful analyses. Second, it has detailed information on the childhood living arrangements of its respondents. Finally, its participants have, for the most part, completed their education. Thus we can analyze the effect of fatherlessness on children's ultimate educational attainment, rather than intermediate measures such as high school graduation.

For whites we find that fatherlessness has a negative effect on educational attainment, regardless how we estimate the models. Moreover, although the estimates vary somewhat, all are significant, at least at the 10 percent level of confidence. Thus for whites, the evidence is clear: longer spells in a single-parent family lower educational attainment. The only question concerns the precise magnitude of this negative effect. We find some evidence that is consistent with the unobserved heterogeneity hypothesis, but we also find evidence of measurement error. On the basis of a number of tests, we cannot rule out the explanation that these two countervailing specification errors simply cancel each other out. Our best estimate is that each additional year of fatherlessness reduces the child's educational attainment by six-tenths of a year. Since the typical fatherless spell lasts about nine years, we conclude that the typical white child in a single-parent family will acquire about one-half year less education than she would have had her parents remained together.

The picture is similar for Hispanics: additional years of fatherlessness lead to lower educational attainment, and the estimates are largely similar regardless how they are computed. Because our samples of Hispanics are smaller than our samples of whites, we must necessarily be somewhat more cautious about our conclusions. To a great extent, however, the typical Hispanic child who spends time in a single-parent family faces disadvantages similar to those of his white counterparts. On average, Hispanic children who grow up fatherless acquire one-half year less education than they would have if they had they lived with both parents.

For blacks the results are most surprising. Our unadjusted estimates indicate the typical pattern: longer spells of fatherlessness lead to lower educational attainment. When we control for family-specific unobservables, however, the estimate changes sign. Furthermore, it is difficult to attribute this occurrence to chance, since the adjusted estimates are statistically significant. Thus on the surface, our estimates suggest that black children who live in single-parent homes actually acquire more education than they would if they lived with both parents. This result is puzzling, and calls for further study. Indeed, until it is confirmed by future research, it best to view this finding as tentative.

Because fatherlessness reduces educational attainment, at least for whites and Hispanics, and because education is an important determinant of adult wages, we expect that fatherless would contribute adversely to the children's adult earnings as well. When we analyze the effects of fatherlessness on wages directly, however, the evidence is mixed. Although the unadjusted estimates are typically negative, estimates that adjust for family effects are either positive or negative but very small. Adding further to the difficulty in interpreting these results, the adjusted estimates typically are insignificant.

We suspect that these mixed results stem from the nature of the wage data available in the NLSY. By necessity, the NLSY includes only entry-level wages, which for a number of reasons may be rather noisy compared with the wages of prime-age workers. We conclude that it would be best to revisit this issue with data on older workers.

I. Introduction

In 1960, only five percent of all births occurred out of wedlock, and only 13 percent of all children lived in a single-parent family. By 1990, increasing divorce and out-of-wedlock childbearing had raised these figures substantially: more than one-fourth of all births were to unmarried mothers, and 27 percent of all children lived with only one parent. Bumpass (1984) predicted that, as a result of rising divorce and non-marital fertility, 50 percent of all children born as early as 1980 would spend part of their childhood in a fatherless family.

Assuming that workers enter the labor market at age 20, and that the workforce turns over every 40 years, these demographic trends imply that, within the next 20 years or so, roughly one-fourth of the labor force will have spent part of its childhood in a single-parent home. Put differently, this means that within the next couple of decades, there will be as many workers who grew up fatherless as workers who hold college degrees. The fraction of the labor force that was raised in a single-parent family will undoubtedly exceed the fraction that is unionized.

These trends in family structure may bode ill for tomorrow's labor force, at least if the predictions of economic theory are correct. In Weiss and Willis' (1985) model, for example, an absent father invests less in his children because it is difficult for him to monitor how the mother allocates his financial contributions. Becker's (1981) model also predicts that children raised in families with fewer resources will tend to have lower human capital. Thus economic theory would predict that, all else equal, the next generation of workers will enter the labor market with less human capital than the last.

Against this backdrop, it is surprising that so few economists have studied the intergenerational effects of family structure. Nevertheless, previous research, conducted primarily by sociologists, largely has confirmed the predictions of economic theory. Researchers have found that childhood fatherlessness decreases educational attainment and adult wages, and has adverse effects on a number of other socioeconomic outcomes.[1]

An important issue not addressed by previous researchers, however, is unobserved heterogeneity across families. One can imagine that families vary greatly in a number of ways that are unobservable to the analyst. Moreover, many of these unobservable characteristics are likely to be correlated both with the probability of divorce and with the well-being of the children. Thus a cross-sectional regression of children's educational attainment on a measure of their childhood family structure fails to identify the effects of living in a fatherless family, because the effects of fatherlessness are confounded with the effects of the family-specific unobservables. We would generally expect such unobserved heterogeneity to lead to exaggerated estimates of the true effects of fatherlessness.

The goal of this study is to distinguish the effects of observable spells of fatherlessness from the effects of unobservable family-specific characteristics. We do this by making within-family comparisons. Drawing on previous research, we specify a child's human capital to depend on the number of years he spends in a single-parent family. Because children enter and leave the family at different times, the duration of a spell of fatherlessness generally will vary among siblings. To eliminate the effects of family-

---

1 McLanahan, 1985, 1988; Hogan and Kitigawa, 1985; Krein, 1986; Keith and Finlay, 1988; Krein and Beller, 1988; McLanahan and Bumpass, 1988; Astone and McLanahan, 1991; Li and Wojtkiewicz, 1992; Wojtkiewicz, 1992; and Haveman and Wolfe, 1994.

specific unobservables, we difference the data within families, relating differences in human capital to differences in the duration of the fatherless spell.

If spells were measured accurately, then differencing within families would provide consistent estimates of the effects of a year of fatherlessness. Our data on childhood living arrangements are measured retrospectively, however, and there is evidence of a fair amount of measurement error, particularly in the differenced data. Under standard assumptions, measurement error causes the estimated regression coefficients to be biased downward.

The usual solution to this problem, instrumental variables estimation, performs poorly in this case, for reasons we discuss below. The approach we adopt instead is method-of-moments estimation. We implement this approach by using sibling comparisons to estimate the extent of the measurement error in our retrospective data.

The results are somewhat mixed. For whites, we find that the estimated effects of fatherlessness on educational attainment are quite robust to changes in the specification. For blacks and Hispanics, simple OLS estimates are negative, but the estimates that account for possible model misspecification are generally quite variable.

II. Data

We use data from the National Longitudinal Survey of Youth (NLSY), a national panel study of 12,686 youths who were 14 to 22 years old in 1979. Survey respondents have been interviewed annually since 1979, and asked questions about their living arrangements, education, and earnings. Our study makes use of two special features of the NLSY: its subsample of siblings and its 1988 retrospective on respondents' childhood living arrangements.

Our primary measure of child human capital is the respondent's level of education at age 27. Measuring educational attainment at this age has two benefits. First, most people have finished their education by then, so education at age 27 essentially measures completed education. Second, all NLSY respondents had turned 27 by 1992, the last year for which education data were available. Thus we are able to use as many observations as the survey provides.

We also study entry-level wages. Here we use average wages, exploiting the longitudinal nature of the survey. Specifically, we average all valid wage data beginning three years after the respondent left school, and extending until age 27.[2] Using average wages should add to the precision of our regression estimates, which is particularly beneficial due to the noisiness of entry-level earnings data.

In 1988 the survey asked respondents about their living arrangements from birth through age 19. From this retrospective we constructed our primary measure of family structure: the number of years that the respondent lived in a fatherless household.[3] Actually, this definition is slightly misleading, because our "years fatherless" variable is the total amount of time during childhood that the respondent spent living in **[XX back to original]** a mother-only, father-only, mother-stepfather, or father-stepmother household. For the most part, however, this variable captures true fatherlessness, since time in a father-only or father-stepmother family accounts for less than 10 percent of the total number of years spent in the absence of at least one biological parent. Note that by this

2 For those individuals who had completed fewer than 10 years of schooling, we average valid wage data from age 18 to age 27.

3 Throughout the paper, we use the terms "fatherless household" and "single-parent household" synonymously, even though they are not quite the same thing, and neither term is exactly accurate.

definition, fatherlessness can arise due to an out-of-wedlock birth, a divorce or separation, or the death of a parent.

We also constructed a second family structure measure as the number of years during childhood spent living without either parent. For the most part, these spells in other, typically non-family living arrangements, were spent in foster homes, detention centers, children's homes, with grandparents, or with other relatives. We separate these relatively unusual living arrangements from the more common family-type arrangements because our primary focus is on the effects of single-parent families that arise due to either a divorce of out-of-wedlock birth.[4]

To estimate the effects of fatherlessness, we exploit the sibling structure of the NLSY. In the original wave of interviews, households were the primary sampling unit, and all youths who met the age restrictions in each household were drawn into the survey. As a result, about half of the survey respondents have siblings who are respondents as well. To the extent that siblings share a common family environment, sibling comparisons can be used to control for unobservable characteristics of their family.

Means of the various family structure variables are presented in Table 1, along with the sample means of several other variables that we include in the regression models below.[5] The first four columns present data from the full sample. We see that whites obtain the most education on average, followed by blacks and Hispanics in that order.

[4] While we recognize that some of our "other" living arrangements, particularly with grandparents or other relatives, may come about as a result of an out-of-wedlock birth or a divorce, we exclude them from our main measure of fatherlessness because they may come about for different reasons as well. Because there are relatively few sample members in this category, our results are largely insensitive to our choice of how to classify them.

[5] Appendix I explains our specific sample inclusion criteria and how we constructed the variables.

Blacks, however, spend the most time in a fatherless household, followed by Hispanics and whites. This bivariate evidence thus suggests that that the relationship between family structure and education may vary by race. The last four columns of Table 1 present data from the subsample of respondents from multi-sib households. For the most part, this subsample, which will be the basis for most of our estimates, appears comparable to the full sample.

III. Estimation

A. The Model

We study the effects of fatherlessness on educational attainment and adult wages using the regression model

$$y_{fi} = \gamma x_{fi} + Q_{fi}\beta + \mu_f + \varepsilon_{fi}, \qquad f = 1,\ldots,n;\ i = 1,\ldots,T_f, \tag{1}$$

where, y_{fi} is the outcome variable (educational attainment or adult wages) of the *i*th child in the *f*th family, x_{fi} is the number of years spent in a fatherless household, Q_{fi} is a vector of background characteristics such as family size, maternal education, and birth order. The variable μ_f is the family-specific unobservable, ε_{fi} is a zero-mean, i.i.d. disturbance term, and γ and β are parameters to be estimated. Specifically, γ measures the effect of a year of fatherlessness on y_{fi}.

This specification merits some discussion. It may seem reasonable to posit that time spent in different living arrangements would have different effects on the human capital of children, and that time in a fatherless household might have different effects depending on the age of the child. In equation (1), however, we have constrained the effects of all family-type living arrangements (besides the traditional two-parent family) to be the same, and have ruled out any interactions between fatherlessness and age.

The support for these restrictions comes from Wojtkiewicz (1992), who studied the relationship between family structure and high school graduation using data from the NLSY. Wojtkiewicz explicitly allowed for different types of living arrangements to have different effects on the likelihood of graduation, and for age-dependence as well. On the basis of a number of specification tests, he concluded that a simple specification that constrained the effects of all non-traditional family-type living arrangements to be the same, and constrained age interactions to be zero, provided the same information as the more complex specifications. We have substantially replicated Wojtkiewicz's results, and concur with his conclusion.[6]

This restricted specification greatly simplifies our analysis. Moreover, in the presence of age interactions, our simple differencing scheme would no longer solve the unobserved heterogeneity problem. Indeed our approach to the measurement error problem would fail as well.

The first problem for estimation is that x_{fi} and μ_f are likely to be correlated. In this case, OLS estimates of equation (1) are inconsistent. To eliminate the family-specific unobservable, we difference equation (1) within families, obtaining

$$\Delta y_{fi} = \Delta x_{fi}\gamma + \Delta Q_{fi}\beta + \Delta\varepsilon_{fi} \qquad (2)$$

[6] Wojtkiewicz's classification of spell types differs from ours slightly in that he included spells with grandparents and other relatives in his main measure of non-traditional living arrangements (akin to our years fatherless measure), whereas we include such spells in our secondary ("other") category. Conceptually, our classification scheme better suits our primary focus, as discussed above. As a practical matter, however, the two classification schemes yield nearly identical results. Details are provided in Appendix II, along with our replication of Wojtkiewicz's results.

where $\Delta y_{fi} = y_{fi} - y_{fj}$ for some j≠i, and so on. Provided that ε_{fi} is uncorrelated with x_{fi}, which amounts to assuming that child-specific unobservable determinants of education do not influence the duration of the child's spell of fatherlessness, OLS applied to the sibling-differenced data yields consistent estimates. Note that equation (2) can only be fitted to the subsample with multiple siblings in the household.

In the case where the length of the fatherless spell is measured with error, the situation is more complicated. Suppose that, rather than observing x_{fi} directly, we observe $z_{fi} = x_{fi} + v_{fi}$, where v_{fi} is an i.i.d. zero-mean measurement error that is uncorrelated with x_{fi}, Q_{fi}, and ε_{fi}. In levels, the model is now

$$y_{fi} = \gamma z_{fi} + Q_{fi}\beta + \mu_f + \eta_{fi}, \qquad (3)$$

where $\eta_{fi} = \varepsilon_{fi} - \gamma v_{fi}$, and in differences, we have

$$\Delta y_{fi} = \Delta z_{fi}\gamma + \Delta Q_{fi}\beta + \Delta \eta_{fi}. \qquad (4)$$

Because z_{fi} is a function of v_{fi}, and v_{fi} appears in the composite error term η_{fi}, OLS applied to the model in levels, equation (3), would yield inconsistent estimates even if years of fatherlessness and the family-specific unobservable were uncorrelated. Similarly, OLS applied to the differenced model in equation (4) will result in inconsistent estimates due to correlation between Δz_{fi} and $\Delta \eta_{fi}$.

B. Instrumental Variables Estimation

The usual solution to the measurement error problem is instrumental variables estimation. In fact in the case of panel data, Griliches and Hausman (1986) have shown that, under plausible assumptions, no external instruments are needed. In levels, we could simply use one sib's report of z_{fi} as an instrument for her sibling's report. Provided the measurement errors were uncorrelated across siblings, this would yield consistent estimates so long as the family effect and the fatherless spell were uncorrelated.

If x_{fi} and μ_f are correlated, however, then we must difference the data to eliminate μ_f. In this case, the IV approach is easiest to illustrate in the case of a family with exactly three children in the sample. After differencing, the first observation will include data on $z_{f3} - z_{f2}$, the second will include $z_{f2} - z_{f1}$, and the third will include $z_{f3} - z_{f1}$, say. The variable z_{f1} is a valid instrument for $z_{f3} - z_{f2}$, z_{f3} is a valid instrument for $z_{f2} - z_{f1}$, and z_{f2} is a valid instrument for $z_{f3} - z_{f1}$.[7] In principle, the panel structure of the data itself provides all the necessary instruments.

In practice, however, these instruments perform poorly in this particular application. The reason for this is simple: the length of the third child's fatherless spell is only weakly correlated with the difference between the spell lengths reported by the first and second children. In Griliches and Hausman's model, the panel dimension of the data was time, so autocorrelation in the z's generally would imply that the regressor (in differences) and the instrument (in levels) would be reasonably highly correlated. When the panel dimension of the data stems from the presence of siblings, in contrast, the correlation between differences and levels can be quite low. In this case, IV estimation produces unsatisfactory results.

B. Method-of-Moments Estimation

[7] Alternatively, under the homogeneity assumption that the variance of the measurement error is the same for all siblings, one could use z_{f3} as an instrument for z_{f3}-z_{f2}, z_{f1} as an instrument for z_{f2}-z_{f1}, and z_{f2} as an instrument for z_{f3}-z_{f1}.

For this reason we chose an alternative approach, method-of-moments-estimation, which is based on a simple fact. Assuming that z_{fi} is the only mis-measured regressor in the model and μ_f is uncorrelated with x_{fi}, the probability limit of the OLS estimate of γ in equation (3) is given by

$$p\lim(\hat{\gamma}_l) = \gamma\left[1 - \frac{V(v_{fi})}{V(z_{fi})(1-R_{l,a}^2)}\right] \tag{5}$$

where $V(v_{fi})$ is the variance of v_{fi}, $V(z_{fi})$ is the variance of z_{fi}, and $R_{l,a}^2$ is the R^2 from an auxiliary regression of z on Q. Of course, differencing allows us to relax the assumptions that the unobserved family effect is uncorrelated with the fatherless spell. The probability limit of the OLS estimate of γ from equation (4), the model in differences, is given by

$$p\lim(\hat{\gamma}_d) = \gamma\left[1 - \frac{2V(v_{fi})}{V(\Delta z_{fi})(1-R_{d,a}^2)}\right], \tag{6}$$

where $V(\Delta z_{fi})$ is the variance of Δz_{fi} and $R_{d,a}^2$ is the R^2 from an auxiliary regression of Δz on ΔQ. The term $V(\Delta z_{fi})$ and the auxiliary R^2's can be estimated readily from the data. If $V(v_{fi})$ can be estimated consistently as well, then one can construct the method-of-moments (MOM) estimators

$$\tilde{\gamma}_l = \hat{\gamma}_l\left[1 - \frac{\hat{V}(v_{fi})}{\hat{V}(z_{fi})(1-R_{l,a}^2)}\right]^{-1} \tag{7}$$

and

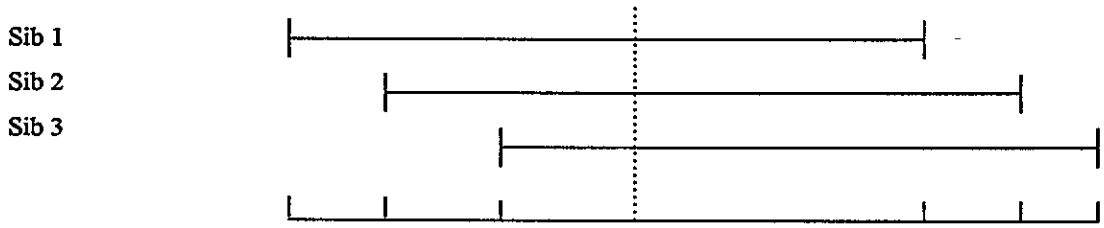
$$\tilde{\gamma}_d = \hat{\gamma}_d\left[1 - \frac{2\hat{V}(v_{fi})}{\hat{V}(\Delta z_{fi})(1-R_{d,a}^2)}\right]^{-1}. \tag{8}$$

Implementing the MOM estimator clearly hinges on obtaining a consistent estimate of $V(v_{fi})$, the variance of the measurement error.

The usual approach to this problem is based on replicated measures of the variable of interest. Suppose that z_{f1} and z_{f2} were both noisy measures of the same latent variable x_f, so $z_{f1}=x_f+v_{f1}$ and $z_{f2} = x_f + v_{f2}$, where v_{f1} and v_{f2}, the measurement errors, are independent of each other and x_f. In this case, $\operatorname{cov}(z_{f1}, z_{f2}) = V(x_f)/\sqrt{V(z_{f1})V(z_{f2})} = V(x_f)/V(z_{f1})$, where the last equality holds under the homogeneity assumption that $V(v_{f1}) = V(v_{f2})$. Since the measurement errors are independent of x_f, we have $V(v_{f1}) = V(v_{f2}) = [1-\operatorname{cov}(z_{f1}, z_{f2})]V(z_{f1})$, which can be estimated from the data.

A complication arises in the family structure problem, however, because z_{f1} and z_{f2} do not measure the same thing. This is because siblings whose parents divorce, or whose parents were unmarried to begin with, generally will spend different amounts of time in the resulting single-parent household. To see this consider the three-child family depicted in Figure 1. For purposes of illustration, we assume that the children were born in 1957, 1959, and 1962, that the parents divorced in 1965, and that all three children remained in their family of origin until age 18.

Figure 1: A Three-Child Household



Year	1957	1959	1962	1965	1975	1977	1980

Because the children are different ages at the time of the divorce, the durations of their spells of fatherlessness differ. For the first child, the true length of the fatherless spell is x_1 = 1975-1965=10 (the f subscript is dropped for clarity), for the second child, x_2 = 12, and for the third child, x_3 = 15. The children's actual reports of their fatherless spells, the z_i terms, may differ from the actual spells, however. In general, we would not expect the report of the first child to equal the report of the second child even on average, because in the case of divorce, earlier-born children generally experience shorter spells of fatherlessness.[8] Thus the simple approach for estimating the variance of the measurement error outlined above will not work in this case.

The problem can be solved once we recognize that the middle child spends his entire childhood in the presence of at least one of his siblings. Thus for any subperiod of sib 2's childhood, both sib 2 and one of his siblings will provide reports of the amount of time that they spent in a single-parent household during that subperiod. With a judicious choice of the particular subperiod, we can ensure that the actual time in a single-parent family was the same for both children. Thus we have replicated measures which, under some assumptions, allow us to estimate the variance of the measurement error over the children's entire childhood.

To see this, note that sibs 1 and 2 are present together in the household between 1959 and 1975, and sibs 2 and 3 are present together in the household between 1962 and

[8] In the case of out-of-wedlock childbearing, earlier-born children generally will experience longer spells.

1977. During the subperiod from 1959 to 1970, say, sibs 1 and 2 experienced the same spell of fatherlessness, lasting for 5 years starting in 1965 and ending in 1970. During the subperiod from 1971 to 1977, sibs 2 and 3 experienced the same spell of fatherlessness, lasting the entire 6 years. Assuming homogeneity, and provided that the measurement errors in the children's reports of subperiod-specific spells of fatherlessness are independent across subperiods, the sum of subperiod measurement error variances is equal to the variance of the measurement error over sib 2's entire childhood. By homogeneity, the variance of the measurement error is the same for all siblings.

To formalize this idea, we divide sib 2's childhood not into two periods but rather into three, corresponding to the dates at which various children enter or leave the family, writing the length of time that he spends in a fatherless household as $x_2 = x_2^1 + x_2^2 + x_2^3$ (again we drop the f subscripts for clarity). The 1 superscript denotes the period between the births of sib 2 and sib 3, when sibs 1 and 2 are the only children present (i.e., 1959 to 1962 in figure 1). The 2 superscript denotes the period between the birth of sib 3 in 1962 and the time when sib 1 leaves home, in 1975. The 3 superscript denotes the period between 1975 and 1977, when sib 2 leaves the nest. For sib 1, we have $x_1 = x_1^0 + x_1^1 + x_1^2$, and for sib 3 we have $x_3 = x_3^2 + x_3^3 + x_3^4$. Superscript 0 refers to the time when sib 1 is the only child present, and superscript 4 denotes the period when sib 3 is the only child left at home.

Measured spells of fatherlessness are given by $z_2 = z_2^1 + z_2^2 + z_2^3$ and accordingly for the other children, where $z_i^j = z_i^j + v_i^j$. Within any subperiod, the true fatherless spell is the same for all children, so $x_i^j = x_l^j$ for all $i \neq l$. The independence and homogeneity assumptions invoked in estimating $V\left(v_{fi}\right)$, where $v_{f2} = v_{f2}^1 + v_{f2}^2 + v_{f2}^3$ and so forth, are given by

$$E\left[v_{fi}^t, v_{fj}^s\right] = 0, \qquad \text{for all f, i, j, s, and t, and} \qquad (9a)$$

$$V\left(v_{fi}^t\right) = V\left(v^t\right) \qquad \text{for all f, i, and t.} \qquad (9b)$$

Together, (9a) and (9b) imply

$$V\left(v_{fi}\right) = V\left(v^t\right) + V\left(v^{t+1}\right) + V\left(v^{t+2}\right). \qquad (9c)$$

Assumption (9a) implies that: (i) subperiod measurement errors are uncorrelated across subperiods, both within and between siblings; and (ii) same-subperiod measurement errors are uncorrelated across siblings. Assumption (9b) says that the variances of subperiod-specific measurement errors are the same for all sibs. These imply (9c), which says that the variance of the total measurement error is equal to the sum of the variances of the subperiod-specific measurement errors.

To estimate the subperiod-specific variances, difference the reported sub-period-specific spells across siblings, for example, $\Delta z_{f2}^1 = z_{f2}^1 - z_{f1}^1$, $\Delta z_{f2}^2 = z_{f2}^2 - z_{f1}^2$, and $\Delta z_{f3}^3 = z_{f3}^3 - z_{f2}^3$. Because the truth is the same within sub periods, $V\left(\Delta z_{f2}^1\right) = 2V\left(v^1\right)$, $V\left(\Delta z_{f2}^2\right) = 2V\left(v^2\right)$, and $V\left(\Delta z_{f3}^3\right) = 2V\left(v^3\right)$. All of these quantities can be estimated from the subsample of families with three respondent children.[9] Thus our estimator of $V\left(v_{fi}\right)$ is $\hat{V}(v_{fi}) = \frac{1}{2}\left[\hat{V}\left(\Delta z_f^1\right) + \hat{V}\left(\Delta z_f^2\right) + \hat{V}\left(\Delta z_f^3\right)\right]$.

[9] We are not limited to using only three-person households to calculate the variance of the measurement error. In each family with more than 2 respondents, there exist $n_f - 2$ interior siblings whose reports can be used to estimate this variance. In each of these families, we can compare the responses of

IV. Results

A. Educational Attainment

Table 2 presents OLS estimates by race from the regression of educational attainment on years of fatherlessness and a set of background variables which included the number of years in a non-familial living arrangement, the child's birth order, the mother's education, the size of the family, and a sex dummy. Chow tests failed to reject pooling by sex, although they strongly rejected pooling by race. The first three columns in Table 2 present results from the full sample, whereas the last three present estimates from the subsample of multi-sib households.

Years of fatherlessness are negatively related to educational attainment, and for all races, the relationship is significant. The coefficient for whites is larger than the coefficient for blacks, which is similar to the results of Krein and Beller (1986). The coefficient for Hispanics is also smaller than whites. To our knowledge, we are the first to report separate estimates for Hispanics.

Time in non-familial living arrangements also has a negative coefficient, as does birth order, at least for whites and blacks. The finding that later-born sibs acquire less education on average has been reported elsewhere (Hanushek 1992). Family size is strongly negatively related to education, consistent with Becker's (1981) model of the trade-off between child quality and quantity. The education of the mother is strongly related the education of her children, although the effect is stronger for whites than for blacks or Hispanics. In general, the results from the multi-sib subsample are similar to those from the full data set.

---

the interior siblings with the responses of the two adjacent siblings over the 3 periods as defined above.

Let us now focus on the effect of fatherlessness. Table 3 presents various estimates of the effect of fatherlessness on educational attainment based on the subsample of multi-sib households. In each case the regression models include the full set of background variables in addition to years of fatherlessness. The first row reports OLS estimates, and so simply replicates the estimates in the last three columns of the top row of Table 2. For whites, each additional year of fatherlessness reduces educational attainment by 0.064 years, on average. For blacks, the figure is substantially smaller, although still statistically significant. The estimate for Hispanics lies between the estimates for whites and blacks, and is also significantly different from zero.

The next row of the table provides estimates of the reliability of the years fatherless regressor. The reliability is inversely related to the variance of the measurement error, and is given by

$$r_l = 1 - \frac{V(v_{fi})}{V(z_{fi})}. \tag{10}$$

The estimated reliabilities are fairly high. For example, the reliability for whites indicates that only 13 percent of the variance of z_{fi} is attributable to measurement error, a proportion that varies only little by race.

In the next row of the table we present method-of-moment estimates that use these reliabilities to correct for measurement error. These estimates would be consistent if measurement error were the only source of misspecification, that is, if there were no correlation between family-specific unobservables and years of fatherlessness. Although we do not consider this a particularly plausible hypothesis, we present these estimates for

---

We use these comparisons to estimate the variance of the measurement error.

sake of completeness. We see that the MOM estimates are only slightly larger than their OLS counterparts, which is to be expected given the modest amount of measurement error.

The estimates in the next row are from OLS applied to the model in equation (4), in which the data have been differenced across siblings. The estimates for whites are quite a bit smaller than the OLS estimates based on equation (3), providing evidence of substantial correlation between family-specific unobservables and years of fatherlessness. An additional year of fatherlessness is now estimated to reduce educational attainment by only 0.035 years among whites. Although this estimate is just more than half the size of the estimate in the first row, it is still significantly different from zero, at least at the ten percent level. For whites, there seems to be little doubt that fatherlessness has a negative effect of educational attainment. The only question concerns the magnitude of the effect.

For blacks, however, controlling for family-specific unobservables has a much more surprising effect. The OLS estimate based on sibling differences is positive rather than negative, a counterintuitive result. Moreover, it is difficult to attribute this result to chance, because the estimate is significant at conventional levels.

For Hispanics, controlling for unobserved heterogeneity across families has little effect on the estimates. It does lower the precision of the estimates, however. Taken at face value, both OLS based on the sibling differences and OLS based on levels suggests that for Hispanics, each additional year of fatherlessness reduces education by about 0.04 years.

Row 5 presents estimates of the reliabilities of within-family differences in years of fatherlessness. This measure too is a function of the variance of the measurement error, and is given by

$$r_d = 1 - \frac{2V(v_{fi})}{V(\Delta z_{fi})}. \tag{11}$$

The reliabilities of the differenced family structure variable are substantially lower than the reliabilities of the levels. From equation (11), we can see why. First, the numerator of the second term on the right-hand side of (11) is twice its counterpart in equation (10). Second, provided that fatherless spells are positively correlated within families, which indeed they must be, the variance of the within-family difference is less than twice the variance of the levels. Intuitively, the differences have more measurement error than the levels because differencing removes more signal than noise.

The greater measurement error in the differenced data suggests another reason why OLS applied to sibling differenced data yields smaller estimates than OLS applied to levels. Rather than accounting for correlation between family structure and family-specific unobservables, differencing the data within families may simply be exacerbating the problem of measurement error, accentuating the usual attenuation bias.

The estimates in the sixth row of Table 3 reinforce this notion. These are the method-of-moment estimates from the differenced data, which account both for family-specific unobserved heterogeneity and measurement error. As expected, accounting for measurement error yields larger coefficients. Indeed for whites and Hispanics, the coefficients on years of fatherlessness are now more negative than the original OLS estimates based on levels. For blacks, accounting for measurement error in the differenced data yields a larger positive coefficient.

The different approaches to estimation thus yield substantially different estimates of the effect of fatherlessness on educational attainment. For whites, the largest coefficient is greater than the smallest by more than a factor of three. For Hispanics, the largest and smallest estimates differ by more than a factor of four, although only the OLS estimate based on levels is significant. For blacks, the estimates vary as to sign.

It is clearly important to choose from among these various estimates, to determine which provides the best representation of the data. A simple approach is to use a set of Hausman tests. Under the null hypothesis of no model misspecification, OLS based on levels provides consistent and asymptotically efficient estimates. The other estimators also should be consistent, but in general will be less efficient. Under the alternative, however, OLS applied to levels will be inconsistent. In the presence of both measurement error and family-specific unobservables, only the method-of-moments estimator applied to sibling differences will yield consistent results.

The first row of Hausman statistics provides tests of the null of no misspecification against the alternative that measurement error is present in levels, but there is no unobserved family effect correlated with years of fatherlessness. The next row provides tests against the alternative that there are family-specific unobservables, but no measurement error. The third row provides tests against the composite alternative that allows for both unobserved heterogeneity and measurement error.

Consider first the results for whites. Neither the test against measurement error nor the test against unobserved family effects yields significant evidence against the null hypothesis, although the test against family effects comes close. The test against the composite alternative fails to reject as well, which is not surprising given the large standard error of the MOM estimate applied to sibling differences.

We therefore conclude that, for whites, the data provide no significant evidence against the null of no misspecification, in which case OLS applied to levels provides the best estimate of the effects of fatherlessness on educational attainment. Disregarding significance levels for the moment, the various estimates tell a plausible story of how this could happen. The estimated reliabilities suggest that some measurement error indeed is present, even in levels. Under standard assumptions, measurement error causes OLS estimates to be too small in absolute value. The OLS estimates based on sibling differences, likewise, suggest that family-specific unobservables may be present, because we would generally expect OLS applied to levels to be biased upward (in absolute value) in the presence of such unobserved heterogeneity. Thus both types of misspecification may be present to some extent, although each offsets the other.

For blacks the story is different. Although measurement error has little effect on the estimates in levels, there is strong evidence of family-specific heterogeneity that is correlated with spells of fatherlessness. The surprising finding, of course, is that accounting for family effects actually yields positive and significant coefficients.

There is some evidence in the literature that the adverse consequences of growing up fatherless are smaller for blacks than for whites, and indeed our results are consistent with this general finding (Krein and Beller 1986). To our knowledge, however, these are the first significant estimates to suggest that fatherlessness might actually have beneficial effects for blacks. Clearly this puzzling result calls for further study.

For Hispanics the results of the Hausman tests are qualitatively similar to whites. Individually, there is little evidence of either measurement error or family effects. Although the MOM coefficient based on sibling differences is much larger than the other estimates, because its standard error is large, we fail to reject against the composite alternative as well. The Hausman tests thus fail to reject the null, in which case the OLS applied to levels provides the best estimate of the effects of fatherlessness.

In summary, the specification tests lead us to conclude that the OLS estimates, which are akin to the estimates found in the previous literature, provide the best representation of the effect of fatherless on educational attainment, at least for whites and Hispanics. For whites there is some insignificant evidence of both family effects and measurement error, but these two types of misspecification seem largely to offset each other. For Hispanics, the evidence of misspecification is slight.

For whites, our best estimate is that each year of fatherlessness leads to a decrease of 0.06 years of education; for Hispanics the decrease is 0.04 years. The average spell of fatherlessness lasts 9.4 years for whites who experience fatherlessness; for Hispanics, the conditional mean is 11.3 years. Thus for both whites and Hispanics, the average child growing up in a fatherless family would acquire about one-half year less education than his counterpart from a two-parent home. Presumably, such a reduction in education would affect his adult earnings as well.

B. Entry-level Wages

Unfortunately, the only data at our disposal with which to test this proposition are data on entry-level wages. Entry-level wages are potentially problematic, because relative wages early in the life-cycle may be only weakly related to relative wages during the prime earnings years. If high-skill workers take jobs with lower starting pay but with higher prospects for earnings growth, for example, then the effect of fatherlessness on entry-level wages might appear different than its effect on wages over the full life-cycle.

Nevertheless, to get a sense for the effect of fatherlessness on earnings, we present in Table 4 results based on the same assortment of estimators used to study the effect of wages on educational attainment.[10] Before examining these estimates, however, it is instructive to determine the magnitudes that these coefficients should have if the sole effect of fatherlessness on wages were due to its effect on education. If a year of schooling increases wages by 10 percent on average, and each year of fatherlessness reduces education by 0.06 years, then each year of fatherlessness should reduce wages by 0.06 percent. This is a small effect, and given the sample sizes at our disposal, it may be difficult to estimate an effect of this magnitude very precisely.

The dependent variable is the average wage described in section II. For whites, the OLS coefficient based on levels is statistically significant, and is roughly double the magnitude we would expect if the only effect of fatherlessness on wages was due to its effects on education. We cannot reject the hypothesis that the true coefficient is equal to 0.006, however. For blacks and Hispanics the OLS coefficients based on levels are also negative, though neither is significant. The coefficient for Hispanics differs from zero only at the fifth decimal place.

Correcting for measurement error alone has only a negligible effect on the estimates. In contrast, correcting for family effects by themselves changes the sign of the

---

10 In addition to the set of regressors included in the education models, the wage models included age and age squared as well as indicators for region of residence, urbanicity, and the regional unemployment rate.

coefficients for both blacks and Hispanics. As one would expect, correcting for both measurement error and family effects produces estimates that are larger than the estimates based on OLS applied to sibling differences, but with the same sign.

Although the pattern of the point estimates across different estimators is different in the case of wages than it was in the case of education, the conclusions based on the formal specification tests are largely the same. For the most part, the null hypothesis of no misspecification cannot be rejected. The exception concerns blacks, for whom the estimates based on sibling differences are positive and significant.

Thus our conclusions based on the formal hypothesis tests are the same for the wage models as for the education models. With the exception of blacks, OLS based on levels provides the best estimate of the effects of fatherlessness on entry-level wages.. The estimates for whites indicate that the wages of the average fatherless worker are about 12 percent lower than the wage of the average worker who grew up in a two-parent family. For Hispanics, there is no evidence that fatherlessness lowers entry-level wages.

We note that, for a number of reasons, we view the conclusions regarding these wage models as more tentative than the conclusions we drew from the education models. In the first place, for whites, the test against family effects only narrowly fails to reject. More generally, the estimates from the wage model are less precise than the estimates from the education model. As a results, researchers with different points of view could justifiably draw the conclusion that fatherlessness has no effect on entry-level wages. If one's prior belief were that family effects and measurement error were present, so that the only basis for inference were the MOM estimates based on sibling differences, then one would fail to reject the null of no effect. Moreover, the specification tests could not reject the maintained null of general misspecification.

Therefore, although the data are consistent with the notion that fatherlessness has negative effects on entry-level wages, at least for whites, we view this finding as less conclusive than our results regarding the effects of fatherlessness on educational attainment. Much of the inconclusiveness undoubtedly stems from the noisiness of the entry-level wage data, and the relatively small sample sizes at our disposal. It would be desirable to revisit this question with data on a larger sample of prime-age workers.

V. Conclusions

Past research, conducted primarily by non-economists, has suggested that children who grow up fatherless acquire less human capital than children in traditional two-parent homes. Our analysis indicates that, for whites, this general conclusion is quite robust. Our best estimates suggest that on average, white children in single-parent families obtain one-half year less schooling than their counterparts in two-parent families. Our results also indicate that fatherlessness is likely to lead to lower educational attainment among Hispanics.

For blacks, however, accounting for unobserved family effects yields estimates that are positive and statistically significant. Taken at face value, this indicates that black children from single-parent homes actually fare better than blacks who live with both parents. It is hard to take such a surprising result at face value, however; this finding clearly calls for further research. It also serves as a methodological warning against pooling data across the races to study the effects of family structure.

Since fatherlessness reduces education, at least for whites and Hispanics, one might expect that workers who grew up in single-parent families would earn less on the labor market. Although our wage analyses yield results that are consistent with this notion, at least for whites, the evidence is rather mixed. It may be that, for a number of reasons, entry-level wages mask the true effect. It would be desirable to revisit this issue with data on older workers.

References

Astone, Nan Marie and Sara S. McLanahan. "Family Structure, Parental Practices, and High School Completion." *American Sociological Review*. 56, (1991): 309-20.

Becker, Gary S. *A Treatise on the Family*. Cambridge, Mass.: Harvard University Press (1981).

Bumpass, Larry L. 1984 "Children and Marital Disruption: A Replication and Update. "*Demography*. 21: 71-82.

Griliches, Zvi and Jerry Hausman. 1986. "Errors in Variables in Panel Data." *Journal of Econometrics*. 31: 93-118.

Hanushek, Eric A. 1992. "The Trade-off between Child Quantity and Quality." *Journal of Political Economy*. 100: 84-117.

Haveman, Robert, and Barbara Wolfe. 1994. *Succeeding Generations On the Effects of Investments in Children*. Russell Sage Foundation, New York.

Hogan, Dennis P. and Evelyn M. Kitigawa. "The Impact of Social Status, Family Structure, and Neighborhood on the Fertility of Black Adolescents." *American Journal of Sociology*. 90, (1985): 825-55.

Keith, Verna M. and Barbara Finlay. "The Impact of Parental Divorce on Children's Educational Attainment, Marital Timing, and Likelihood of Divorce." *Journal of Marriage and the Family*. 50, (1988): 797-809.

Krein, Sheila Fitzgerald. 1986. "Growing Up in a Single-Parent Family: The Effect on Education and Earnings of Young Men." *Family Relations*. 35: 161-68.

Krein, Sheila Fitzgerald and Andrea H. Beller. 1988. "Educational Attainment of Children From Single-Parent Families: Differences by Exposure, Gender, and Race." *Demography*. 25: 221-34.

Li, Jiang Hong and Roger Wojtkiewicz. "A New Look at the Effects of Family Structure on Status Attainment." *Social Science Quarterly*. 73, (1992): 581-95.

McLanahan, Sara. 1985. "Family Structure and the Reproduction of Poverty." *American Journal of Sociology*. 90: 873-901.

McLanahan, Sara S. "Family Structure and Dependency: Early Transitions to Female Household Headship." *Demography*. 25, (1988): 1-16.

McLanahan, Sara S. and Larry Bumpass. "Intergenerational Consequences of family Disruption." *American Journal of Sociology*. 94, (1988): 130-52.

Weiss and Willis. **XX Jeff.**

Wojtkiewicz, Roger A. 1992 "Simplicity and Complexity in the Effects of Parental Structure on High School Graduation." *Demography*. 30: 701-17.

Table 1: Sample Means, by race

Variable name	Full sample Total	White	Black	Hispanic	Multi-sib households Total	White	Black	Hispanic
Education	12.74	13.08	12.49	12.00	12.80	13.21	12.44	12.19
	(2.39)	(2.46)	(1.99)	(2.52)	(2.31)	(2.47)	(1.92)	(2.23)
Real wage	8.29	8.79	7.22	8.24	8.28	9.04	6.93	8.24
	(5.44)	(5.70)	(4.73)	(5.27)	(5.21)	(5.61)	(4.21)	(5.02)
Years fatherless	3.40	2.33	5.75	3.32	3.00	1.72	5.32	2.85
	(5.99)	(4.89)	(7.40)	(5.92)	(5.76)	(4.22)	(7.30)	(5.54)
Years other	0.43	0.26	0.79	0.44	0.19	0.12	0.32	0.20
	(2.20)	(1.57)	(3.10)	(2.23)	(1.35)	(0.88)	(1.90)	(1.37)
Birth order	2.98	2.73	3.40	3.18	3.13	2.80	3.60	3.30
	(2.25)	(1.93)	(2.61)	(2.50)	(2.30)	(1.87)	(2.67)	(2.64)
Female	0.51	0.51	0.50	0.50	0.48	0.49	0.48	0.46
	(0.50)	(0.50)	(0.50)	(0.50)	(0.50)	(0.50)	(0.50)	(0.50)
Only child	0.03	0.03	0.03	0.02	N.A.	N.A.	N.A.	N.A.
	(0.17)	(0.17)	(0.17)	(0.14)				
One sibling	0.13	0.16	0.08	0.09	0.08	0.12	0.04	0.05
	(0.33)	(0.37)	(0.26)	(0.29)	(0.28)	(0.33)	(0.19)	(0.21)
Mother's education	10.39	11.37	10.15	7.39	10.64	11.75	10.44	7.48
	(3.76)	(3.17)	(3.51)	(4.35)	(3.77)	(2.72)	(2.87)	(3.95)
Sample size:	9660	5516	2559	1585	4579	2428	1384	767
Number of individuals with valid wage data:	9179	5297	2375	1507	4345	2267	1229	714

Notes: Unweighted sample means. Standard deviations in parentheses.

Table 2: OLS Coefficients from the Regression of Educational Attainment at Age 27 on Years Fatherless and Other Control Variables, by race

Dependent variable: Completed schooling at age 27

	Full sample			Multi-sib households		
Variable name	White	Black	Hispanic	White	Black	Hispanic
Years fatherless	-0.060	-0.027	-0.022	-0.064	-0.023	-0.044
	(0.006)	(0.005)	(0.010)	(0.010)	(0.007)	(0.014)
Years other	-0.087	-0.026	-0.013	-0.113	-0.022	0.002
	(0.018)	(0.011)	(0.028)	(0.048)	(0.026)	(0.056)
Birth order	-0.066	-0.038	0.023	-0.027	-0.022	0.052
	(0.017)	(0.016)	(0.027)	(0.025)	(0.020)	(0.032)
Female	0.097	0.438	0.188	0.250	0.560	0.361
	(0.057)	(0.073)	(0.119)	(0.084)	(0.097)	(0.154)
Only child	0.449	0.720	0.335	N.A.	N.A.	N.A.
	(0.176)	(0.216)	(0.416)			
One sibling	0.452	0.409	1.106	0.651	0.019	1.179
	(0.082)	(0.142)	(0.215)	(0.134)	(0.265)	(0.370)
Mother's education	0.411	0.241	0.187	0.420	0.229	0.148
	(0.012)	(0.015)	(0.016)	(0.018)	(0.022)	(0.021)
Constant	8.561	10.033	10.433	8.383	10.002	10.836
	(0.167)	(0.198)	(0.199)	(0.248)	(0.281)	(0.245)
Sample size	5516	2559	1585	2428	1384	767

Notes: Standard errors are in parentheses. Flags were created to accommodate missing values in the variables reported in the table. Separate flags were set equal to one for the observations with a missing value in the birth order or mother's education variables, and they were set equal to zero otherwise. The missing value in the variable was then set to zero. Also, separate flags were set equal to one if we observed the respondent's education at age 28 or age 26, they were set to zero otherwise. The regressions for the white samples included an indicator of membership in the poor white subsample.

Table 3: Alternative Estimates of the Effect of Years in a Fatherless Household on Educational Attainment at Age 27, by race

Dependent variable: Completed schooling at age 27

		White	Black	Hispanic
(1)	Ordinary least squares	-0.064	-0.023	-0.044
		(0.010)	(0.007)	(0.014)
(2)	Estimated reliability (r_l)	0.87	0.83	0.90
(3)	Method of moments (levels)	-0.074	-0.027	-0.049
		(0.012)	(0.008)	(0.015)
(4)	Sibling differences	-0.035	0.021	-0.041
		(0.019)	(0.011)	(0.029)
(5)	Estimated reliability (r_d)	0.32	0.33	0.24
(6)	Method of moments (sibling differences)	-0.118	0.066	-0.182
		(0.062)	(0.035)	(0.129)
Hausman test statistics:				
(7)	(1) vs. (3)	2.27	1.07	0.86
(8)	(1) vs. (4)	3.22	26.88	0.01
(9)	(1) vs. (6)	0.78	6.74	1.16
Sample sizes:				
	Levels	2428	1384	767
	Sibling differences	1673	1013	548

Notes: Standard errors are in parentheses. Each regression includes the control variables reported in the body of Table 2 as well as the flags listed in the notes to Table 2. The Hausman statistics are calculated using the difference in the variable of interest to increase the power of the test. The critical value for the 95 percent level of the chi-squared test is 3.84. The sample sizes differ between the OLS and sibling-difference specifications because we lose one observation from all two-respondent families.

Table 4: Alternative Estimates of the Effect of Years in a Fatherless Household on the Logarithm of the Average Real Wage, by race for males only (own educational attainment not included as a regressor)

Dependent variable: The logarithm of the average real wage rate of males

		White	Black	Hispanic
(1)	Ordinary least squares	-0.011	-0.004	-0.005
		(0.004)	(0.003)	(0.005)
(2)	Estimated reliability (r_l)	0.87	0.83	0.90
(3)	Method of moments (levels)	-0.012	-0.005	-0.005
		(0.005)	(0.004)	(0.006)
(4)	Sibling differences	0.011	0.011	-0.008
		(0.009)	(0.006)	(0.014)
(5)	Estimated reliability (r_d)	0.32	0.33	0.24
(6)	Method of moments (sibling differences)	0.048	0.042	-0.067
		(0.036)	(0.023)	(0.121)
Hausman test statistics:				
(7)	(1) vs. (3)	0.11	0.14	0.00
(8)	(1) vs. (4)	7.45	8.33	0.05
(9)	(1) vs. (6)	2.72	4.07	0.26
Sample sizes:				
	Levels	727	428	267
	Sibling differences	436	279	164

Notes: Standard errors are in parentheses. The regressions reported in this table do not include the respondent's educational attainment as a regressor. Each regression contains the control variables reported in the body of Table 2, the flags listed in the notes to Table 2, and indicators for region of residence, urbanicity, and the regional unemployment rate. The Hausman statistics are calculated using the difference in the variable of interest to increase the power of the test. The critical value for the 95 percent level of the chi-squared test is 3.84. The sample sizes differ between the OLS and sibling-difference specifications because we lose one observation from all families having only two male respondents.

Table 5: Alternative Estimates of the Effect of Years in a Fatherless Household on the Logarithm of the Average Real Wage, by race for males only (own educational attainment included as a regressor)

Dependent variable: The logarithm of the average real wage rate of males

		White	Black	Hispanic
(1)	Ordinary least squares	-0.009	-0.003	-0.005
		(0.004)	(0.003)	(0.005)
(2)	Estimated reliability (r_l)	0.87	0.83	0.90
(3)	Method of moments (levels)	-0.010	-0.003	-0.005
		(0.005)	(0.004)	(0.006)
(4)	Sibling differences	0.013	0.010	-0.005
		(0.008)	(0.006)	(0.014)
(5)	Estimated reliability (r_d)	0.32	0.33	0.24
(6)	Method of moments (sibling differences)	0.054	0.040	-0.046
		(0.035)	(0.022)	(0.132)
Hausman test statistics:				
(7)	(1) vs. (3)	0.11	0.00	0.00
(8)	(1) vs. (4)	10.08	6.26	0.00
(9)	(1) vs. (6)	3.28	3.89	0.10
Sample sizes:				
	Levels	727	428	267
	Sibling differences	436	279	164

Notes: Standard errors are in parentheses. Unlike Table 4, the regressions reported in this table contained the respondent's educational attainment as a regressor. Each regression also included the control variables reported in the body of Table 2, the flags listed in the notes to Table 2, age, age squared, and indicators for region of residence, urbanicity, and the regional unemployment rate. The Hausman statistics are calculated using the difference in the variable of interest to increase the power of the test. The critical value for the 95 percent level of the chi-squared test is 3.84. The sample sizes differ between the OLS and sibling-difference specifications because we lose one observation from all families having only two male respondents.

Table 6: Alternative Estimates of the Effect of Years in a Fatherless Household on the Logarithm of the Average Real Wage, by race for females only (own educational attainment not included as a regressor)

Dependent variable: The logarithm of the average real wage rate of females

		White	Black	Hispanic
(1)	Ordinary least squares	-0.013	-0.005	0.008
		(0.005)	(0.004)	(0.006)
(2)	Estimated reliability (r_l)	0.87	0.83	0.90
(3)	Method of moments (levels)	-0.015	-0.006	0.008
		(0.006)	(0.005)	(0.007)
(4)	Sibling differences	-0.012	0.011	0.027
		(0.013)	(0.009)	(0.017)
(5)	Estimated reliability (r_d)	0.32	0.33	0.24
(6)	Method of moments (sibling differences)	-0.043	0.043	0.176
		(0.047)	(0.036)	(0.102)
Hausman test statistics:				
(7)	(1) vs. (3)	0.36	0.11	0.00
(8)	(1) vs. (4)	0.01	3.94	1.43
(9)	(1) vs. (6)	0.41	1.80	2.72
Sample sizes:				
	Levels	675	354	195
	Sibling differences	417	215	117

Notes: Standard errors are in parentheses. The regressions reported in this table did not contain the respondent's educational attainment as a regressor. However, Each regression contains the control variables reported in the body of Table 2, the flags listed in the notes to Table 2, age, age squared, and indicators for region of residence, urbanicity, and the regional unemployment rate. The Hausman statistics are calculated using the difference in the variable of interest to increase the power of the test. The critical value for the 95 percent level of the chi-squared test is 3.84. The sample sizes differ between the OLS and sibling-difference specifications because we lose one observation from all families having only two female respondents.

Table 7: Alternative Estimates of the Effect of Years in a Fatherless Household on the Logarithm of the Average Real Wage, by race for females only (own educational attainment included as a regressor)

Dependent variable: The logarithm of the average real wage rate of females

		White	Black	Hispanic
(1)	Ordinary least squares	-0.010	-0.004	0.010
		(0.005)	(0.004)	(0.006)
(2)	Estimated reliability (r_l)	0.87	0.83	0.90
(3)	Method of moments (levels)	-0.011	-0.005	0.011
		(0.005)	(0.004)	(0.006)
(4)	Sibling differences	-0.009	0.006	0.017
		(0.013)	(0.009)	(0.016)
(5)	Estimated reliability (r_d)	0.32	0.33	0.24
(6)	Method of moments (sibling differences)	-0.031	0.027	0.132
		(0.046)	(0.038)	(0.120)
Hausman test statistics:				
(7)	(1) vs. (3)	0.26[1]	0.18[1]	0.00[1]
(8)	(1) vs. (4)	0.01	1.54	0.22
(9)	(1) vs. (6)	0.21	0.67	1.04
Sample sizes:				
	Levels	675	354	195
	Sibling differences	417	215	117

Notes: Standard errors are in parentheses. The regressions reported in this table included the respondent's educational attainment as a regressor, the control variables reported in the body of Table 2, the flags listed in the notes to Table 2, age, age squared, and indicators for region of residence, urbanicity, and the regional unemployment rate. The Hausman statistics are calculated using the difference in the variable of interest to increase the power of the test. The critical value for the 95 percent level of the chi-squared test is 3.84. The sample sizes differ between the OLS and sibling-difference specifications because we lose one observation from all families having only two female respondents.

(1) Rounding of the standard errors in the tables makes it appear these statistics can not be estimated.

Appendix I

Data Construction

It was necessary to exclude some of the original 12,686 individuals from the analysis. Because we look at the effects of family structure on educational attainment at age 27, we could not include individuals with missing data on either of these measures. We also excluded 9 biological siblings who report being of different races. The resulting data set had 9,660 individuals with complete childhood living arrangement and educational attainment data; 9,179 of whom had valid wage data as well.

In the initial interview in 1979, the NLSY identified the relationship of each respondent to all co-resident respondents and collected information on the number of siblings and the number of older siblings of each respondent. We used this information to identify biological siblings within households and to construct measures of family size and birth order.

We identified two respondents as biological siblings if, in 1979, both claimed the other as a biological brother or sister. Half-brothers and half-sisters did not qualify. There were 4,579 individuals who had a biological-sibling respondent with valid educational attainment information, of which 2,428 are white, 1,384 are black, and 767 are Hispanic. Of these, 4,210 individuals had a sibling respondent with valid wage data as well: 2,267 white, 1,229 black, and 714 Hispanic.

To construct our measures of family size, it was first necessary to correct for inconsistencies within families. To do this, we constructed an average family size from the responses to the number of siblings question. We then constructed two dummy variables to measure family size. First, we set a dummy variable equal to one if the average family size equaled one, and it was set to zero otherwise. Second, we constructed a dummy variable equal to one if the average family size was two, and it was set to zero otherwise.

To calculate the respondent's birth order, it was again necessary to correct for inconsistencies within families. We first constructed a birth-order variable by adding one to the individual's reported number of older siblings. We then checked for consistency between the sibling's birth dates and our estimate of their birth order. In families having inconsistencies between birth dates and birth order, we corrected discrepancies where proper order could be determined, otherwise birth order for all individuals in these families was set to missing.

We created flags for missing and changed values to include in the regressions. All missing values of control variables were set to zero and an indicator of this was constructed. We also constructed a flag equal to one if birth order was edited, and it was set to zero otherwise.

In 1988, the NLSY asked about the educational attainment of each respondent's mother. To correct for measurement error in this variable, we used the within family mean of these reports.

In each year, the NLSY determines the respondent's region of residence, whether the respondent lives in an urban or rural area, and the regional unemployment rate. We used the reports of these variables in the year the respondent turned 27 to control for local labor market characteristics that might affect wages. If this information was missing in the year the respondent turned 27, we used the educational-attainment algorithm reported in the text to construct these measures. Again, missing values of these variables were set to zero and an indicator of this was included in each regression.

Appendix II

Wojtkiewicz Replication

Wojtkiewicz (1992) used the family-structure retrospective in the NLSY to determine how experiences of parental structure affect high school graduation. We present the results of his Model 2.1, our replication of his model, and estimates of this model using our analysis samples in Table A1. The first column restates the results reported in Wojtkiewicz (1992). The second column shows our replication of these results. The third and fourth columns report the results of estimating his model 2.1 using our samples.

A comparison of the first two columns of the table show that we were able to replicate his results. In all cases, the estimated family-structure parameters are not significantly different, and in some cases, they are identical. Six of the seven birth cohort dummies have the same sign, and they are not significantly different. There is no statistical difference between the sex, race, and parental education variables in the two regressions as well.

A comparison of the results in columns three and four show the expected result that the parameter estimates do not vary between individuals in our full sample and those in the multi-sib households. Moreover, a comparison of the parameter estimates in columns three and four with those in columns one and two show little difference between the results obtained using our samples, Wojtkiewicz's results, or our replication of his results.

Table A1: Comparison of the Effect of Years In Parental Structure Type on High School Graduation, by sample

	(1)	(2)	(3)	(4)
			Analysis sample	
Variable:	Wojtkiewicz's results	Our replication	Full sample	Multi-sib households
Mother only	-0.037	-0.033	-0.032	-0.037
	(0.007)	(0.007)	(0.005)	(0.007)
Mother stepfather	-0.036	-0.036	-0.030	-0.041
	(0.011)	(0.011)	(0.008)	(0.013)
Father only	-0.102	-0.121	-0.105	-0.102
	(0.025)	(0.026)	(0.019)	(0.030)
Father stepmother	-0.034	-0.035	-0.024	0.051
	(0.029)	(0.029)	(0.023)	(0.048)
Grandparents	-0.051	-0.036	-0.029	-0.021
	(0.016)	(0.016)	(0.013)	(0.028)
Other relatives	-0.034	-0.044	-0.055	-0.031
	(0.035)	(0.035)	(0.027)	(0.058)
Other	-0.082	-0.101	-0.088	-0.140
	(0.024)	(0.017)	(0.020)	(0.046)
Birth cohort-1957	-0.063	0.031	0.280	0.297
	(0.128)	(0.126)	(0.117)	(0.198)
Birth cohort-1958	-0.107	-0.015	0.188	0.413
	(0.126)	(0.124)	(0.116)	(0.179)
Birth cohort-1959	-0.097	-0.133	-0.060	0.204
	(0.125)	(0.121)	(0.111)	(0.163)
Birth cohort-1960	-0.141	-0.139	-0.066	0.151
	(0.118)	(0.116)	(0.108)	(0.148)
Birth cohort-1961	-0.124	-0.103	-0.064	0.124
	(0.118)	(0.115)	(0.108)	(0.144)
Birth cohort-1962	0.005	0.002	-0.010	0.066
	(0.118)	(0.115)	(0.108)	(0.142)

Table A2: Comparison of the Effect of Years In Parental Structure Type (continued)

Variable:	(1) Wojtkiewicz's results	(2) Our replication	(3) Analysis sample: Full sample	(4) Analysis sample: Multi-sib households
Birth cohort-1963	-0.126	-0.068	-0.089	0.067
	(0.117)	(0.114)	(0.108)	(0.144)
Female	0.426	0.437	0.404	0.610
	(0.059)	(0.058)	(0.053)	(0.078)
Black	0.067	-0.027	0.037	-0.134
	(0.078)	(0.074)	(0.072)	(0.104)
Hispanic	-0.491	-0.524	-0.441	-0.469
	(0.081)	(0.077)	(0.076)	(0.114)
One sibling	0.271	0.177	0.199	0.345
	(0.106)	(0.118)	(0.106)	(0.213)
More than 3 siblings	-0.399	-0.376	-0.423	-0.349
	(0.068)	(0.084)	(0.075)	(0.116)
Parent failed to graduate from high school	-0.974	-0.988	-0.959	-0.762
	(0.070)	(0.068)	(0.061)	(0.091)
Parent some college	0.571	0.619	0.605	0.517
	(0.129)	(0.127)	(0.112)	(0.159)
Parent college graduate	1.435	1.378	1.312	1.299
	(0.163)	(0.153)	(0.133)	(0.195)
Missing parental education	-1.461	-1.466	-1.474	-1.127
	(0.134)	(0.133)	(0.124)	(0.188)
Poor-white subsample	N.A.	N.A.	-0.697	-0.999
			(0.081)	(0.137)
Sample Sizes:	8381	8382	9660	4579

Notes: Standard errors are in parentheses. Wojtkiewicz's results are from Table 2, Model 2.1, columns 1 and 2 in Wojtkiewicz (1992). Also, for a discussion of variable definitions see Wojtkiewicz (1992). The main difference between our samples and Wojtkiewicz's is that we include respondents from the poor-white subsample.

National Longitudinal Surveys (NLS)
Discussion Paper Series

Number	Author	Title
01	Michael R. Pergamit	How the Federal Government Uses Data from the National Longitudinal Surveys
02	Norman M. Bradburn Martin R. Frankel Reginald P. Baker Michael R. Pergamit	A Comparision of Computer-Assisted Personal Interviews (CAPI) with Paper-and Pencil Interviews (PAPI) in the National Longitudinal Survey of Youth
03	Saul Schwartz Robert Hutchens George Jakubson	Dynamic Models of the Joint Determination of Labor Supply and Family Structure
04	A. Colin Cameron R. Mark Gritz Thomas MaCurdy	The effects of Unemployment Compensation on the Unemployment of Youth
05	Henry S. Farber	Evaluating Competing Theories of Worker Mobility
06	Frank L. Mott Paula Baker	Evaluation of the 1989 Child-care Supplement in the National Longitudinal Survey of Youth
07	Audrey Light Manuelita Ureta	Gender Differences in the Quit Behavior of Young Workers
08	Lisa M. Lynch	The Impact of Private Sector Training on Race and Gender Wage Differentials and the Career Patterns of Young Workers
09	Evangelos M. Falaris H. Elizabeth Peters	Responses of Female Labor Supply and Fertility to the Demographic Cycle
10	Anne Hill June E. O'Neill	A Study of Intercohort Change in Women's Work Patterns and Earnings
11	Arleen Leibowitz Jacob Alex Klerman Linda Waite	Women's Employment During Pregnancy and Following Birth
12	Lee A. Lillard	Work Experience, Job Tenure, Job Separation, and Wage Growth
13	Joseph G. Altonji Thomas A. Dunn	Family Background and Labor Market Outcomes

National Longitudinal Surveys (NLS)
Discussion Paper Series

Number	Author	Title
92-01	Michael R. Pergamit	How the Federal Government Uses Data from the National Longitudinal Surveys
92-02	Norman M. Bradburn Martin R. Frankel Reginald P. Baker Michael R. Pergamit	A Comparision of Computer-Assisted Personal Interviews (CAPI) with Paper-and Pencil Interviews (PAPI) in the National Longitudinal Survey of Youth
92-03	Saul Schwartz Robert Hutchens George Jakubson	Dynamic Models of the Joint Determination of Labor Supply and Family Structure
92-04	A. Colin Cameron R. Mark Gritz Thomas MaCurdy	The effects of Unemployment Compensation on the Unemployment of Youth
92-05	Henry S. Farber	Evaluating Competing Theories of Worker Mobility
92-06	Frank L. Mott Paula Baker	Evaluation of the 1989 Child-care Supplement in the National Longitudinal Survey of Youth
92-07	Audrey Light Manuelita Ureta	Gender Differences in the Quit Behavior of Young Workers
92-08	Lisa M. Lynch	The Impact of Private Sector Training on Race and Gender Wage Differentials and the Career Patterns of Young Workers
92-09	Evangelos M. Falaris H. Elizabeth Peters	Responses of Female Labor Supply and Fertility to the Demographic Cycle
92-10	Anne Hill June E. O'Neill	A Study of Intercohort Change in Women's Work Patterns and Earnings
92-11	Arleen Leibowitz Jacob Alex Klerman Linda Waite	Women's Employment During Pregnancy and Following Birth
92-12	Lee A. Lillard	Work Experience, Job Tenure, Job Separation, and Wage Growth
92-13	Joseph G. Altonji Thomas A. Dunn	Family Background and Labor Market Outcomes

92-14	George J. Borjas Stephen G. Bronars Stephen J. Trejo	Self-Selection and Internal Migration in the United States
92-15	James J. Heckman Stephen V. Cameron Peter Z. Schochet	The Determinants and Consequences of Public Sector and Private Sector Training
93-16	R. Mark Gritz Thomas MaCurdy	Participation in Low-Wage Labor Markets by Young Men
94-17	Alan L. Gustman Thomas L. Steinmeier	Retirement in a Family Context: A Structural Model for Husbands and Wives
94-18	Audrey Light	Transitions from School to Work: A Survey of Research Using the National Longitudinal Surveys
94-19	Christopher J. Ruhm	High School Employment: Consumption or Investment
95-20	Mark Lowenstein James Spletzer	Informal Training: A Review of Existing Data and Some New Evidence
95-21	Jacob Alex Klerman	Characterizing Leave for Maternity: Modeling the NLSY Data
95-22	Jacob Alex Klerman Arleen Leibowitz	Employment Continuity Among New Mothers
95-23	Stephen G. Bronars Carol Moore	Incentive Pay, Information, and Earnings: Evidence from the National Longitudinal Survey of Youth
95-24	Donald O. Parsons	The Evolving Structures of Female Work Activities: Evidence from the National Longitudinal Surveys of Mature Women Survey, 1967-1989
95-25	Donald O. Parsons	Poverty Dynamics Among Mature Women: Evidence from the National Longitudinal Surveys, 1967-1989
95-26	Ann Bartel Nachum Sicherman	Technological Change and the Skill Acquisition of Young Workers
95-27	Audrey Light	High School Employment
95-28	Laura Branden R. Mark Gritz Michael R. Pergamit	The Effect of Interview Length on Attrition in the National Longitudinal Survey of Youth
96-29	Thomas Dunn Douglas Holtz-Eakin	Capital Market Constraints, Parental Wealth and the Transition to self-Employment among Men and Women

96-30	Jeff Grogger Nick Ronan	The Intergenerational Effects of Fatherlessness on Educational Attainment and Entry-level Wages