Estimating the Effects of Family Background on the Return to Schooling

Olivier DESCHÊNES
Department of Economics, University of California, Santa Barbara, CA 93106 (*olivier@econ.ucsb.edu*)

This article examines the causal link between family background characteristics—parental education and family size—and returns to schooling. I implement a model of schooling and earnings with heterogeneous returns to education using data from the Occupational Change in a Generation Survey. I find that men raised in larger families have substantially lower returns to education, whereas the combined effects of parental education are more modest. In addition, like other "supply-side" instrumental variables studies of the causal effect of education, I find two-stage least squares estimates that are larger than the corresponding ordinary least squares estimates. The results suggest an alternative explanation for this phenomenon: constant marginal return to schooling, combined with a negative absolute ability bias and a positive comparative advantage bias.

KEY WORDS: Ability bias; Family background; Return to education; Self-selection.

1. INTRODUCTION

Economists have long been interested in the effects of family environment on the subsequent labor market success of individuals (see, e.g., Becker 1967; Taubman 1977; Griliches 1979). Part of this interest stems from the strong correlation between the educational attainment of parents and children, which may contribute to the transmission of socioeconomic status and inequality across generations. In recent years, this issue has drawn even more attention due to the changing nature of the American family and the increasing role of education as a determinant of economic well-being.

Recent studies of the causal association between schooling and earnings have emphasized the heterogeneity in the economic return to an additional year of education across otherwise comparable individuals (see, e.g., Heckman and Vytlacil 1998; Card 1999). Despite increased attention to the possibility of heterogeneous returns to education across individuals, there is still considerable uncertainty about the mechanism generating this heterogeneity. Part of this uncertainty is attributable to the absence of a formal model that explicitly recognizes the possibility that the causal return to schooling varies with observable characteristics, such as family background variables.

This article examines the relationship between family background characteristics and the return to schooling subsequently received by individuals in the labor market. The article begins by documenting several features of the relationship among family background factors, educational attainment, and earnings. Using a large sample from the 1973 Occupational Change in a Generation Survey, I find that men raised by better educated parents acquire more schooling and have higher earnings, whereas those raised in larger families are less educated and have lower earnings. Next, I show that the negative effect of family size varies with the gender composition of the sibship. In particular, holding family size and parental education constant, I find that men raised with more sisters have substantially lower schooling and earnings. These patterns are robust to a wide variety of specifications.

The contribution of this article is to develop and implement a formal model of schooling and earnings to interpret these patterns. In light of the recent instrumental variables studies of the causal effect of education, the return to schooling is allowed to vary across individuals, and, in particular, with the observable characteristics of the family. This distinguishes the current article from most of the literature, which typically assumes that the return to schooling is constant across the population or is a single random variable. A key implication of the model is that family background can potentially affect both the payoff to an additional year spent in school *and* the level of acquired schooling. Therefore, a complete assessment of the link between family background *and* the return to schooling must examine the effect of family background on both the marginal benefit and the marginal cost of schooling.

An extensive literature has clearly established that the identification of the causal relationship between schooling and earnings requires an exogenous source of variation in educational choices. It follows naturally that the identification of the parameters describing the costs and benefits of schooling requires two types of exclusion restrictions. The identification of the parameters in the marginal benefit function requires the existence of an observable variable affecting schooling choices only through its effect on the cost of schooling (i.e., an instrumental variable for schooling). Similarly, the identification of the parameters from the marginal cost function requires the existence of an observable variable affecting schooling choices only through its effect on the benefit of schooling. In this study, measures of school quality are used as variables that affect the benefit, but not the cost of schooling, conditional on family background characteristics. Then, following Butcher and Case (1994), I exploit the randomness embodied in the gender composition among siblings, holding family size constant, as a variable influencing only the cost of schooling. These two exclusion restrictions allow the estimation of the average causal effect of education, and of the parameters describing the effect of family background on the return to schooling. Moreover, because the effects of gender composition on educational attainment are presumably larger for poorer families (conditional on family size), it is possible to test the implicit assumption that sibling gender composition has no independent effect on earnings.

Journal of Business & Economic Statistics
July 2007, Vol. 25, No. 3
DOI 10.1198/073500106000000567

The results of a series of specification tests provide no evidence against the hypothesis that, conditional on family size, sibling gender composition is an exogenous determinant of schooling.

The results can be summarized as follows. First, men raised in larger families have significantly lower returns to education. This finding is entirely attributable to the lower benefits per year of education received by individuals raised with more siblings (i.e., it is not related to differences in the costs of schooling). The *combined* effects of parental education on the returns to schooling are more modest. Men who were brought up by better educated fathers have higher marginal returns to schooling, while those with better educated mothers have lower marginal returns to schooling. In addition to their opposite signs, the analysis suggests that these effects of parental education operate through distinct mechanisms relating familial environment and returns to education. Father's education is associated with higher benefits per year of education, whereas mother's education is associated with lower costs per year of education, consequently raising education levels but lowering the marginal return.

This article is organized as follows. Section 2 describes the data and provides a preliminary descriptive analysis. Section 3 presents a model of schooling and earnings, emphasizing the contribution of family background characteristics to the heterogeneity in the returns to schooling across the population. Section 4 presents the identification and estimation strategies used in this article. Section 5 discusses the empirical results. Section 6 concludes.

2. PRELIMINARY ANALYSIS AND DATA DESCRIPTION

An ideal dataset for the study of the effects of family background on the return to schooling would provide detailed information on current labor market outcomes of individuals, as well as information on the characteristics of their families during the childhood years. The data in this article are taken from the 1973 Occupational Changes in a Generation survey (OCG). This dataset provides a unique source of family background information (number of siblings, education of both parents, family income at age 16, state of birth, etc.). Whereas other datasets such as the NLSY and the PSID contain similar family background information, small samples and missing-data problems limit the interpretation of any result derived from them. The data from the OCG are drawn from a large and representative sample of the adult male population in the United States. The survey was carried out as an eight-page mailout–mailback supplemental questionnaire to individuals in the sampling frame for the 1973 March Current Population Survey (CPS). The target population consisted of civilian males aged 20–65. The present analysis is based on the sample of 32,986 males aged 20–65 from the March CPS population.

This study focuses on a sample of men aged 24–65 who were born in the United States excluding Hispanics. Table 1 presents some summary statistics for the baseline sample and the subsample of prime-aged workers. All statistics reported in this article are weighted by the OCG sample weights. As column 1 shows, the OCG sample provides a nationally representative sample of the population of men aged 20–65. The family background information reveals that the average sibship size was about 4 for these cohorts of men. Individuals were also asked to report the education of their parents, and the entries in Table 1 show that mothers are slightly better educated than fathers. A small fraction of individuals did not report their parents' education. Observations with missing data on parental education were imputed with the predicted values from separate regressions of father and mother's education on other measures of family background and children's education. The statistical models reported later always include dummy variables indicating whether either parent's education was imputed, and the inclusion or exclusion of such observations does not affect the results presented below. Column 2 reports the characteristics of individuals aged 24–65 who worked full time in 1972 and earned more than $60 per week (which corresponds to the weekly earnings of those working 40 hours per week at the 1972 federal minimum wage). The analyses in this article will be performed on the subsample of 17,300 observations in column 2. Comparisons of the mean characteristics in columns 1 and 2 indicate no important differences between the baseline sample and the subsample of workers.

Table 1. Summary statistics

	(1) Full OCG sample	(2) Full-time workers
Age	40.0	41.2
Years of education	11.75	12.15
Percent black	.10	.10
Weekly wage	180.9	229.3
Labor market experience	20.8	21.3
Number of siblings	4.01	3.87
Number of brothers	2.02	1.94
Number of sisters	1.99	1.93
Father's education	8.50	8.20
Mother's education	9.00	8.72
Lived on a farm open country at age 16	.19	.19
Lived with both parents at age 16	.83	.83
Born in the South	.33	.35
Living in the South in 1973	.31	.31
Living in an SMSA in 1973	.69	.72
Pupil–teacher ratio	—	30.26
Relative teacher salary	—	1.07
Observations	32,032	17,300

NOTE: (1) All observations in the baseline sample of men aged 20–65. (2) Sample of men aged 24–65, earning at least $60 per week, working at least 48 weeks last year, not in school during the reference week, and with nonmissing information on size of the sibship.

The data from the OCG samples are supplemented by characteristics of public schools in each state for the years 1918–1968. In particular, semiannual data from the Biennial Survey of Education covering the years 1918–1958 and annual data from the Digest of Education Statistics starting in 1960 provide information about statewide enrollment, number of teachers, teacher salary, and term length. These data have been used in previous studies, notably by Card and Krueger (1992), from whom I draw the school characteristics data used to supplement the OCG data. Based on state and year of birth, I assigned the average elementary and secondary school quality that was potentially available to each individual if he were to complete the first

12 years of schooling. Focusing on "potential" school quality rather than actual school quality leaves the endogeneity of educational attainment with school quality aside. In this study, I focus on two measures of school quality: the pupil–teacher ratio and the average teacher salary, relative to the average wages in each state. As Card and Krueger (1992) and Heckman, Layne-Farrar, and Todd (1996) documented, other characteristics of public schools, such as term length, are only weakly related to returns to schooling and do not vary as much across cohorts. Despite this evidence, expenditure-based measures of school quality, such as teacher salary, remain controversial (see, e.g., Hanushek 1986).

The analysis begins by examining regressions of educational attainment and log earnings on measures of family background and school quality. All models include cohort dummies (for men born between 1910 and 1919, 1920 and 1929, 1930 and 1939, and 1940 and 1949), a race dummy, three dummies for the region of birth, three dummies for the region of residence in 1973, and an indicator for residence in a Standard Metropolitan Statistical Area (SMSA) in 1973. In addition, the models include other measures of family structure at age 16: a dummy indicating if the respondent lived with both parents at age 16 and a dummy indicating if the respondent lived on a farm. Table 2 presents a variety of reduced-form regressions for years of completed education. The models in columns 1–3 add an increasing set of family background characteristics: parental education (column 1) and number of siblings (column 2). Column 3 adds a simple measure of sibling gender composition: an indicator variable for the presence of any sisters among a person's siblings. Finally, column 4 adds the two measures of school quality. As shown by the F statistics in row 7 and the t statistics, the family background characteristics are always individually and jointly significant at the 5% level.

Rows 1–3 show that men with better educated parents complete more years of education, whereas those from larger families acquire less schooling. Row 4 confirms this, but also shows that the effect of the number of siblings varies with the gender composition of the sibship: The estimates in columns 3 and 4 indicate that holding the number of siblings and family characteristics constant, men with at least one sister have .2 fewer years of education than otherwise comparable men (p value = .001). School quality, as measured by the pupil–teacher ratio and relative teacher pay, is strongly correlated with years of education, as indicated by an F statistic of 73.61 (p value = .001). In columns 5–8 the models are estimated separately for four different birth cohorts (for men born between 1910 and 1919, 1920 and 1929, 1930 and 1939, and 1940 and 1949). The within-cohort analysis is motivated by the important changes in family structure and school quality for the cohorts born between 1910 and 1949. Those important changes might not be fully captured by the cohort dummies included in models 1–4. As indicated by the F statistics in rows 7 and 8, family background and school quality are jointly significant for each of the four birth cohorts. Column 9 provides the F statistics for testing the equality of the effects across the four cohorts. The results indicate that the effect of father's education on schooling is similar across cohorts (p value = .625). The effects of family size and gender composition are also similar across cohorts (p values = .922 and .620, respectively). Analogous results are also found for the effect of school quality on educational attainment.

Table 2. Reduced-form regressions of educational attainment on family background and school quality

	(1)	(2)	(3)	(4)	(5) Born 1910–1919	(6) Born 1920–1929	(7) Born 1930–1939	(8) Born 1940–1949	(9) F statistics: equality of (5)–(8)
1. Father's education	.1737	.1583	.1578	.1576	.1693	.1702	.1546	.1524	.59
	[.0072]	[.0071]	[.0071]	[.0070]	[.0161]	[.0129]	[.0123]	[.0110]	(.625)
2. Mother's education	.2175	.1858	.1856	.1852	.1913	.1766	.2144	.1641	3.32
	[.0079]	[.0078]	[.0078]	[.0078]	[.0168]	[.0132]	[.0130]	[.0122]	(.019)
3. Number of siblings	—	−.1699	−.1571	−.1542	−.1607	−.1590	−.1566	−.1468	.16
		[.0071]	[.0078]	[.0078]	[.0176]	[.0149]	[.0152]	[.0150]	(.922)
4. Sibling composition	—	—	−.2066	−.2127	−.1001	−.2693	−.2940	−.1672	.59
(=1 if any sisters)			[.0534]	[.0532]	[.1431]	[.1072]	[.1033]	[.0893]	(.620)
5. Pupil–teacher ratio	—	—	—	−.0624	−.0485	−.0370	−.0330	−.0342	.23
				[.0061]	[.0108]	[.0105]	[.0132]	[.0159]	(.873)
6. Relative teacher salary	—	—	—	.5883	.6766	.3672	.3500	.2187	1.92
				[.0911]	[.1501]	[.1190]	[.1582]	[.2106]	(.125)
F statistics									
7. Family background main effects	994.15	936.34	738.40	778.20	184.51	271.31	294.50	220.76	—
	(.001)	(.001)	(.001)	(.001)	(.001)	(.001)	(.001)	(.001)	
8. School quality main effects	—	—	—	73.61	24.01	12.33	5.65	3.47	—
	—	—	—	(.001)	(.001)	(.001)	(.004)	(.031)	
9. R^2	.30	.32	.32	.33	.31	.31	.32	.29	—

NOTE: Sample size is 17,300. Standard errors in brackets, p values in parentheses. All models include a race dummy, three cohort dummies, three region of birth dummies, three indicators for region of residence in 1973, SMSA status in 1973, and imputation dummies. Other family background controls are indicators for living with both parents at age 16 and for living on a farm at age 16.

Table 3. Reduced-form regressions of log earnings on family background and school quality

	(1)	(2)	(3)	(4)	(5) Born 1910–1919	(6) Born 1920–1929	(7) Born 1930–1939	(8) Born 1940–1949	(9) F statistics: equality of (5)–(8)
1. Father's education	.0114	.0104	.0104	.0104	.0083	.0151	.0130	.0064	4.57
	[.0011]	[.0011]	[.0011]	[.0011]	[.0026]	[.0021]	[.0020]	[.0018]	(.003)
2. Mother's education	.0159	.0139	.0139	.0143	.0182	.0140	.0172	.0081	5.39
	[.0012]	[.0013]	[.0013]	[.0013]	[.0027]	[.0021]	[.0021]	[.0020]	(.001)
3. Number of siblings	—	−.0109	−.0090	−.0084	−.0078	−.0122	−.0097	−.0043	1.92
		[.0011]	[.0013]	[.0013]	[.0028]	[.0024]	[.0024]	[.0024]	(.124)
4. Sibling composition (=1 if any sisters)	—	—	−.0310	−.0312	−.0474	−.0293	−.0318	−.0243	.25
			[.0086]	[.0086]	[.0229]	[.0172]	[.0165]	[.0143]	(.863)
5. Pupil–teacher ratio	—	—	—	−.0033	−.0048	−.0028	−.0010	.0065	4.88
				[.0010]	[.0017]	[.0017]	[.0021]	[.0026]	(.002)
6. Relative teacher salary	—	—	—	.1381	.1215	.0848	.1638	.1904	4.37
				[.0149]	[.0240]	[.0191]	[.0254]	[.0337]	(.005)
F-statistics									
7. Family background main effects	200.20	179.58	151.96	150.22	37.16	70.73	70.65	19.18	—
	(.001)	(.001)	(.001)	(.001)	(.001)	(.001)	(.001)	(.001)	
8. School quality main effects	—	—	—	50.36	19.64	12.00	21.19	17.31	—
				(.001)	(.001)	(.001)	(.001)	(.001)	
9. R^2	.18	.18	.18	.19	.17	.21	.24	.11	—

NOTE: Sample size is 17,300. Standard errors in brackets, p values in parentheses. All models include a race dummy, three cohort dummies, three region of birth dummies, three indicators for region of residence in 1973, SMSA status in 1973, and imputation dummies. Other family background controls are indicators for living with both parents at age 16 and for living on a farm at age 16.

Table 3 presents reduced-form regressions (i.e., excluding schooling) of log weekly earnings on the same specifications of family background and school quality displayed in Table 2. The estimated coefficients have the same signs as those reported in Table 2, but they are smaller in magnitude. In the estimated models, all but one of the family background variables are individually significant at the 5% level and all are jointly significant. Consistent with the findings in Table 2, men with better educated parents have higher earnings, whereas those from larger families have lower earnings on average. Again, the relationship between sibship size and log earnings appears to depend on the gender composition: Holding family size constant, men with at least one sister earn 3% less than otherwise comparable men (p value = .001). These patterns of parental education and gender composition typically hold true for the within-cohort regressions as well, with smaller estimated effects for the men born between 1940 and 1949. The school quality indicators are always jointly significant, although for some of the cohorts the effects are individually insignificant. Nevertheless, the point estimates indicate that individuals who received better primary and secondary education have higher earnings on average.

In summary, the evidence contained in Tables 2 and 3 suggests three clear patterns for the effects of family background and school quality on educational attainment and earnings. First, individuals with better educated parents have higher educational attainment and earnings. Second, men from larger families have less schooling and lower earnings on average. Moreover, holding family size constant, those with at least one sister have further lower earnings and education. Third, individuals educated in states and cohorts with better school quality (lower pupil–teacher ratio and higher relative teacher salary) also have more schooling and higher earnings, although the relationship is not as strong as it is for family background. Most of these patterns hold true within and across cohorts and are robust to a wide variety of specifications. The next section will exploit and further analyze these patterns in order to identify the effect of family background on the return to education.

3. THEORETICAL FRAMEWORK

The purpose of this study is to identify the effect of family background on the return to education. With this objective in mind, I set out a formal model of schooling and earnings that explicitly specifies the connection among family background factors, schooling, earnings, and return to schooling. In light of recent studies of the causal effect of education, the model allows the return to schooling to vary across individuals. Unlike previous studies, however, the model considers the effect of family background on both the marginal benefit *and* the marginal cost of an additional year of education. Therefore, family background characteristics will affect the return to education in two distinct ways: First, the benefits of an additional year of education, can be directly affected for example, through the transmission of innate ability; and second, families can influence the amount of schooling acquired by individuals through higher benefits of lower costs, indirectly affecting the marginal return to schooling. A key point is that unless the marginal benefit of an additional year of education is constant, both effects must be considered in order to assess the effect of family background on the return to schooling. Consequently, the parameters of the family background gradients in the marginal benefit and marginal cost schedules must be estimated.

To proceed, I augment the simple causal model for earnings and schooling proposed by Heckman and Vytlacil (1998) and Card (1999) to allow family background to directly influence the level of earnings:

$$\log y_i = a_i + b_i S_i - .5k_1 S_i^2 + F_i'\gamma + \varepsilon_i, \tag{1}$$

where S_i represents years of completed education and F_i represents a vector of family background variables. Other determinants of log earnings, such as labor market experience and race, are ignored at this point to keep the presentation simple. The intercept a_i represents the level of ability of individuals, which does not interact with the level of schooling (i.e., the absolute advantage). The other ability factor b_i, is the heterogeneous component of the education slope, which interacts with the level of schooling and grants higher net returns to schooling to individuals with higher b_i (i.e., the comparative advantage). Both a_i and b_i represent an unspecified combination of individual specific abilities, influences of familial environment, and inherited skills. In the following specification, both ability factors will be allowed to be freely correlated with family background characteristics.

Consistent with the quadratic structural earnings function in (1) are linear marginal benefit and marginal cost schedules with heterogeneous intercepts:

$$MB_i \equiv b_i - k_1 S_i,$$

$$MC_i \equiv r_i + k_2 S_i,$$

where k_1 and k_2 are positive constants representing the slopes of the schedules and r_i is a person-specific discount rate. Again, r_i will be allowed to be freely correlated with family background characteristics. Equating marginal benefit and marginal cost yields an expression for the optimal schooling level:

$$S_i = \frac{b_i - r_i}{k}, \tag{2}$$

where $k = k_1 + k_2$. Equation (2) illustrates that schooling is determined by comparing the marginal benefit and marginal cost of an additional year of schooling. In this model, family background affects schooling through its effects on b_i and r_i. I then decompose the heterogeneity factors a_i, b_i, and r_i into an average effect and components due to family background variables, denoted by F. The decomposition also embodies two types of exclusion restrictions, which will be key for identification later:

$$a_i = a_0 + \sum_{j=1}^{J} a_j^F (F_{ji} - \overline{F}_j) + v_{1i}, \tag{3.1}$$

$$b_i = b_0 + \sum_{j=1}^{J} b_j^F (F_{ji} - \overline{F}_j) + \sum_{k=1}^{K} b_k^Q (Q_{ki} - \overline{Q}_k) + v_{2i}, \tag{3.2}$$

$$r_i = r_0 + \sum_{j=1}^{J} r_j^F (F_{ji} - \overline{F}_j) + \sum_{l=1}^{L} r_l^Z (Z_{li} - \overline{Z}_l) + v_{3i}. \tag{3.3}$$

According to the model in (3), Z_{li}, $l = 1, \ldots, L$, are instrumental variables for schooling in (1): They are observable variables that affect schooling choice, but are uncorrelated with the log earnings heterogeneity factors a_i and b_i. Therefore, instrumental variables like Z_{li} can be used to identify the parameters of the marginal *benefit* equation. The model embodies another exclusion restriction: Conditional on measures of family background, the observable variables Q_{ki}, $k = 1, \ldots, K$, affect the marginal benefit of schooling but not the marginal cost. Therefore, these are not proper instruments for schooling because they directly influence the return to schooling. However, variables satisfying this kind of exclusion restriction will identify the parameters of the marginal cost equation.

Equations (1)–(3) provide a generalized version of the linear causal model for schooling and log earnings. In the standard model, the unobserved determinants of log earnings and schooling a_i, b_i, and r_i are treated solely as random variables, and the family background gradients are all equal to 0. In the present context, b_0 denotes the average causal effect of schooling on log earnings and r_0 is the average discount factor in the population. The objective of this article paper is to develop and implement a methodology to estimate the parameters b_0 and r_0 and the parameters b_j^F and r_j^F, which measure the effect of family background measure j on the marginal benefit and marginal cost of schooling. To facilitate the interpretation of the family background gradients in (3.1)–(3.3), consider the case of parental education. If better educated parents transmit their innate abilities and enrich the "learning environment" or reduce the "cost" of learning in their families, then associated parameters b^F and r^F will be positive and negative, respectively.

Without further assumptions, however, it remains difficult to fully understand the implications of the preceding model for the observed relationship between log earnings, schooling, and family background variables. In what follows, the assumption of linearity of the conditional expectations in (3.1)–(3.3) will be maintained: $\mathrm{E}[v_{1i}|F_i, Q_i, Z_i] = \mathrm{E}[v_{2i}|F_i, Q_i, Z_i] = \mathrm{E}[v_{3i}|F_i, Q_i, Z_i] = 0$. This assumption does not impose a particular dependence structure on the stochastic components v_{1i}, v_{2i}, and v_{3i}. Moreover, it allows the stochastic components of the marginal benefit and cost equations to be correlated with other variables of the model, in particular with schooling. In fact, the conventional absolute ability bias arises because of a correlation between v_{1i} and S_i, whereas the comparative advantage bias arise because of a correlation between v_{2i} and S_i.

Substituting (3.2) and (3.3) in the schooling (2) yields an equation for the realized schooling levels as a function of family background, and the variables excluded from the marginal benefit and marginal cost schedules (Z_i and Q_i):

$$S_i = \frac{1}{k}\Bigg[(b_0 - r_0) + \sum_{j=1}^{J}(b_j^F - r_j^F)(F_{ji} - \overline{F}_j) + \sum_{k=1}^{K} b_k^Q (Q_{ki} - \overline{Q}_k) - \sum_{l=1}^{L} r_l^Z (Z_{li} - \overline{Z}_l) + v_{2i} - v_{3i}\Bigg].$$

It follows that the effect of a family background variable F_{ij} on the marginal return to schooling is given by

$$\frac{\partial MB_i}{\partial F_{ji}} = \frac{\partial b_i}{\partial F_{ji}} - k_1 \frac{\partial S_i}{\partial F_{ji}},$$

$$= b_j^F (1 - \omega) + r_j^F \omega, \qquad j = 1, \ldots, J,$$

where $\omega = k_1/k$. This expression shows that the heterogeneity in the return to schooling arises because of the variation in

b_i *and* r_i. Family background can affect the return to schooling through its "direct" effect on b_i (captured by b_j^F) and its "indirect" effect on amount of schooling acquired, via b_i and r_i [captured by $\omega(b_j^F - r_j^F)$]. Therefore the direct and indirect effects must be estimated to fully assess the effect of family background on the return to schooling. In the special case where the earnings function (1) is linear in schooling, the marginal benefit of schooling is constant (which implies $\omega = 0$), and it is sufficient to measure the effects of family background on b_i only. This is the case considered by Altonji and Dunn (1996) and Ashenfelter and Rouse (1998) who obtained mixed evidence on whether parental education affects the return to education. This article generalizes their studies to the case where family background is allowed to jointly influence education levels and return to education.

4. EMPIRICAL FRAMEWORK

4.1 Identification of the Family Background Gradients

In the empirical application of this model, I consider three measures of family background, two measures of school quality, and one instrumental variable. In that setting, the model in (1)–(3) implies the following reduced-form regression model for schooling:

$$\begin{aligned} S_i &= \mathrm{E}[S_i|F_i, Q_i, Z_i] + \xi_i \\ &= \pi_{10} + \sum_{j=1}^{J} \pi_{1j}^F (F_{ji} - \overline{F}_j) + \sum_{k=1}^{K} \pi_{1k}^Q (Q_{ki} - \overline{Q}_k) \\ &\quad + \sum_{l=1}^{L} \pi_{1l}^Z (Z_{li} - \overline{Z}_l) + \xi_i, \end{aligned} \tag{4}$$

where the reduced-form coefficients in (4) depend on the parameters from the marginal benefit and marginal cost equations (3.2) and (3.3). The complete correspondence between the reduced-form coefficients and the parameters of the model is given in the Appendix. The regression function for log earnings is given by

$$\begin{aligned} &\mathrm{E}[\log y_i|S_i, F_i, Q_i, Z_i] \\ &\quad = \mathrm{E}[a_i|S_i, F_i, Q_i, Z_i] + \mathrm{E}[b_i|S_i, F_i, Q_i, Z_i]S_i \\ &\qquad - .5k_1 S_i^2 + F_i'\gamma + Q_i'\delta. \end{aligned}$$

It follows from (4) and the assumption of linear conditional expectations embodied in (3.1)–(3.3) that the linear projections of v_{1i} and v_{2i} on (S_i, F_i, Q_i, Z_i) are given by

$$\mathrm{E}^*[v_{1i}|S_i, F_i, Q_i, Z_i] = \lambda_S \xi_i, \tag{5.1}$$

$$\mathrm{E}^*[v_{2i}|S_i, F_i, Q_i, Z_i] = \Psi_S \xi_i, \tag{5.2}$$

where $\mathrm{E}^*[Y|X]$ denotes the linear projection of Y on an intercept and X, and where λ_S and Ψ_S are the linear projections coefficients of v_{1i} and v_{2i} on ξ_i. Therefore, the conditional expectations of the unobserved heterogeneity factors a_i and b_i are given by

$$\begin{aligned} &\mathrm{E}[a_i|S_i, F_i, Q_i, Z_i] \\ &\quad = a_0 + \sum_{j=1}^{J} a_j^F (F_{ji} - \overline{F}_j) + \lambda_S \xi_i, \\ &\mathrm{E}[b_i|S_i, F_i, Q_i, Z_i] \\ &\quad = b_0 + \sum_{j=1}^{J} b_j^F (F_{ji} - \overline{F}_j) + \sum_{k=1}^{K} b_k^Q (Q_{ki} - \overline{Q}_k) + \Psi_S \xi_i. \end{aligned}$$

In this model, λ_S is the absolute ability bias due to the correlation between a_i and S_i, and Ψ_S is a comparative advantage bias due to the correlation between b_i and S_i. The parameter Ψ_S is proportional to the total variance in schooling outcomes attributable to the heterogeneity in the schooling slopes. Letting f denote this (unobservable) fraction, one can show that $\Psi_S = kf$, where k is the sum of the slopes from the marginal benefit and marginal cost equations. Substituting the conditional expectations of the unobservables a_i and b_i in the regression function, I obtain the following estimating equation:

$$\begin{aligned} \log y_i &= \pi_{20} + \pi_{21} S_i + \pi_{22} S_i^2 + \pi_{23} \xi_i + \pi_{24} S_i \xi_i \\ &\quad + \sum_{j=1}^{J} \pi_{2j}^F (F_{ji} - \overline{F}_j) S_i + \sum_{k=1}^{K} \pi_{2k}^Q (Q_{ki} - \overline{Q}_k) S_i \\ &\quad + F_i'\gamma + Q_i'\delta + e_i, \end{aligned} \tag{6}$$

where the regression coefficients in (6) depend on the parameters of the structural earnings function (1) and on the parameters from the marginal benefit and marginal cost equations (3.2) and (3.3). The complete correspondence between the regression coefficients and the parameters of the model is given in the Appendix. Given the two types of exclusion restrictions discussed previously, the regression coefficients in (4) and (6) identify all the parameters of interest in the model: the population averages b_0 and r_0, the family background gradients b_j^F and r_j^F, and the slopes of the schedules, k_1 and k_2. The inclusion of ξ_i and $\xi_i S_i$ as controls in the regression will eliminate any ability or endogeneity biases from the relationship between log earnings and years of education and will allow direct estimation of the parameters λ_S and Ψ_S.

This control function approach is due to Garen (1984) and has been further extended in several articles since (see Heckman and Vytlacil 1998; Card 1999; Woolridge 2000; Florens, Heckman, Meghir, and Vytlacil 2004). In particular, one could include higher order terms in ξ_i and $S_i\xi_i$ and relax the linearity assumptions in (5.1) and (5.2). Given the already large number of parameters to be estimated in the current setting, I elected not to pursue this possibility.

4.2 Estimation

This study will focus on the effects of three measures of family background on the return to schooling: father's education, mother's education, and number of siblings (denoted by F_1, F_2, and F_3 respectively). In most datasets, these three measures of familial environment are the strongest predictors of children's future outcomes. Two measures of school quality, the pupil–teacher ratio, and the relative teacher pay (denoted by Q_1 and Q_2, respectively), will be used as exclusion restrictions in the marginal cost equation. Therefore, a maintained assumption in this article is that holding family background and size constant, measures of school quality at the cohort and state level only influence the marginal benefit of schooling and have no effect on the marginal cost. This assumption would fail, for example, if an increase in school quality raises the perceived cost of schooling because of the higher effort required to progress through the academic curriculum. Following Butcher and Case (1994), measures of sibling gender composition, in particular the presence of any sisters in the sibship (conditional on family size), will be used as instrumental variables for schooling. This exclusion restriction identifies the parameters of the marginal benefit equation. The motivation for this approach is that gender composition of the sibship is random. Its drawbacks are discussed later.

The first step of the empirical analysis is to estimate the regression coefficients of the schooling and log earnings equations. This following system of two regressions is estimated jointly:

$$\begin{aligned} S_i = {} & \pi_{10} + \pi_{11}^F(F_{1i} - \overline{F}_1) + \pi_{12}^F(F_{2i} - \overline{F}_2) \\ & + \pi_{13}^F(F_{3i} - \overline{F}_3) + \pi_{11}^Q(Q_{1i} - \overline{Q}_1) + \pi_{12}^Q(Q_{2i} - \overline{Q}_2) \\ & + \pi_{11}^Z(Z_i - \overline{Z}) + g(X_{1i}, \Gamma_1) + \xi_i, \end{aligned} \tag{7}$$

$$\begin{aligned} \log y_i = {} & \pi_{20} + \pi_{21}S_i + \pi_{22}S_i^2 + \pi_{23}\xi_i + \pi_{24}S_i\xi_i \\ & + \pi_{21}^F S_i F_{1i} + \pi_{22}^F S_i F_{2i} + \pi_{23}^F S_i F_{3i} + \pi_{21}^Q S_i Q_{1i} \\ & + \pi_{22}^Q S_i Q_{2i} + g(X_{2i}, \Gamma_2) + \varepsilon_i. \end{aligned} \tag{8}$$

Both X_{1i} and X_{2i} contain three cohort dummies, a race indicator, three dummies for the three region of birth, three indicators for the region of residence in 1973, and an indicator for residence in a metropolitan area (SMSA) in 1973. In addition to these regressors, X_{2i} includes a quartic in labor market experience and the main effects of the family background and school quality variables. The parameters are obtained by using optimal minimum-distance (OMD) estimation (Chamberlain 1984). In the present context, the OMD procedure seeks estimates of the 13 parameters of the marginal cost and marginal benefit schedules that are as close as possible to the predictions of the model, based on the 14 relevant regression coefficients from (7) and (8). The details of the estimation procedure are presented in the Appendix.

5. RESULTS

5.1 Educational Attainment and Gender Composition Among Siblings

Before proceeding with the estimation of the parameters of the model, I further examine the relationship between educational attainment and gender composition of the siblings. Table 4 presents regressions of completed education on the same specifications as Table 2, adding controls for the presence of any sisters, any brothers, and the fraction of females among siblings. This simple analysis might shed some light on the mechanisms through which family composition affects educational attainment. In each specification, the effects of family size are controlled for by a linear main effect or by including a series of unrestricted dummies. The entries in Table 4 again provide clear evidence that holding family size and background constant, men who grew up with at least one sister have significantly lower levels of education. The point estimates indicate that the presence of at least one sister reduces educational attainment by .17 to .21 year on average. Moreover, in all specifications, the "any sisters" variable is a stronger predictor of educational attainment than parental education, and it is also has a larger effect than number of siblings in the linear specifications. This finding is contrary to that of Butcher and Case (1994), who concluded that sibship gender composition had no effect on the educational attainment of men. They note that "For men in older age cohorts, completed education

Table 4. Gender composition, family size, and educational attainment

	(1) Linear effects	(2) Unrestricted effects	(3) Linear effects	(4) Unrestricted effects	(5) Linear effects	(6) Unrestricted effects	(7) Linear effects	(8) Unrestricted effects
Any sisters	−.2066	−.1679	—	—	—	—	−.1958	−.1149
	[.0534]	[.0617]					[.0830]	[.0904]
Any brothers	—	—	−.0633	.0523	—	—	—	—
			[.0536]	[.0631]				
Percent female	—	—	—	—	−.2821	−.2490	−.0243	−.1156
					[.0920]	[.0982]	[.1428]	[.1437]
Father's education	.1578	.1553	.1581	.1554	.1581	.1554	.1578	.1554
	[.0071]	[.0071]	[.0079]	[.0071]	[.0071]	[.0071]	[.0071]	[.0071]
Mother's education	.1856	.1844	.1859	.1845	.1855	.1843	.1586	.1844
	[.0078]	[.0078]	[.0078]	[.0078]	[.0078]	[.0078]	[.0078]	[.0078]
Number of siblings	−.1571	Yes	−.1657	Yes	−.1624	Yes	−.1571	Yes
	[.0078]		[.0079]		[.0075]		[.0078]	
R^2	.32	.33	.32	.32	.32	.33	.32	.32

NOTE: Dependent variable: years of education. Sample size is 17,300. Standard errors in brackets. Number of siblings is controlled for by linear main effects (columns 1, 3, 5, 7) or by unrestricted effects (19 dummies, columns 2, 4, 6, 8). All models include a race dummy, three cohort dummies, three region of birth dummies, three indicators for region of residence in 1973, SMSA status in 1973, and imputation dummies. Other family background controls are indicators for living with both parents at age 16 and for living on a farm at age 16.

appears to be negatively related to the presence of any sisters if one does not control adequately for the number of siblings in the family." Interestingly, even with flexible controls for the number of siblings, the OCG data show a strong negative relationship between educational attainment and the presence of any sisters. However, it is important to keep in mind that even with unrestricted dummies in family size, the model considered here is meant to be parsimonious. With a richer and larger dataset, one would condition on the full pattern of gender composition and consider other predictors of educational attainment, such as the gender of the next youngest sibling and birth order (see, e.g., Black, Devereux, and Salvanes 2005). Unfortunately, this information is not available in the OCG.

Table 4 also shows that the effects of sibling gender composition are not "exchangeable" in gender. As columns 3 and 4 indicate, conditional on family size and background, there are no differences in the educational attainment of men who grew up with at least one brother and those who did not: The estimated "any brothers" effect is small and insignificant. Moreover, the estimated difference is positive or negative, depending on the specification of the family size effects. Columns 5 and 6 show that the fraction of females in the sibship is also an important determinant of the educational attainment of men. In columns 7 and 8, both the "any sisters" and the percent female variables are included to determine which one has the strongest effect on educational attainment. It is apparent that the effect of sibling's gender composition on the educational attainment of males is mainly working through the presence of sisters in the sibship. In both specifications, the percent female variable is insignificant and is smaller in magnitude. The indicator for at least one sister has essentially the same magnitude as in column 1 and is relatively precisely estimated in the linear effects specification, whereas it is less precisely estimated in the unrestricted effects specification. Overall, the estimates in Table 4 demonstrate that holding family size and background constant, men who grew up with at least one sister have significantly lower educational attainment than those who did not.

These patterns are consistent with a model where parents care about the lifetime wealth and labor market earnings of their children and where the return to educational investments is lower for women (Berhman, Pollak, and Taubman 1982). In that case, the presence of sisters in the sibship will be negatively correlated with the educational attainment of males because more family resources will have to be allocated to females in order to equalize labor market earnings. Given that the gender composition of the sibship is random (conditional on other measures of family background), this suggests that indicators of the gender composition can potentially be used as variables affecting schooling only through their effects on the cost of schooling. One drawback of such an instrument is that fertility is a choice variable and that some theoretical models of fertility predict that family size and child ability are negatively related. Because the probability of having at least one sister increases with family size, it is possible that the "any sisters" variable might still be correlated with the unobserved ability factors, even after controlling for family size. Rosenzweig and Wolpin (2000) presented a discussion along those lines. Section 5.5 reports some tests of the validity of the "any sisters" variable as an instrument for years of education.

5.2 Two-Stage Least Squares Estimates of the Return to Education

As a prelude to the empirical implementation of the model developed in Sections 3 and 4, Table 5 presents a series of reduced-form regressions for schooling and log earnings, as well as the ordinary least squares (OLS) and two-stage least squares (2SLS) estimates of the return to education, using the presence of any sisters as an instrumental variable (IV) for years of completed education. All specifications are based on the same set of controls as in Tables 2–4, plus a quartic in labor market experience. The second and third columns of Table 5 indicate that holding family size and background constant,

Table 5. Reduced-form, OLS, and two-stages least square estimates of the return to schooling

	(1) OLS	(2) Reduced-form: log earnings	(3) Reduced-form: years of education	(4) 2SLS
Any sisters	—	−.0311	−.1609	—
		[.0084]	[.0492]	
Years of education	.0533	—	—	.1932
	[.0012]			[.0656]
Father's education	.0035	.0100	.1220	−.0136
	[.0011]	[.0011]	[.0065]	[.0081]
Mother's education	.0055	.0133	.1463	−.0150
	[.0012]	[.0012]	[.0073]	[.0097]
Number of siblings	−.0017	−.0073	−.1324	.0183
	[.0011]	[.0012]	[.0072]	[.0094]
Pupil–teacher ratio	−.0002	−.0019	−.0308	.0040
	[.0009]	[.0010]	[.0056]	[.0023]
Relative teacher salary	.1037	.1341	.5721	.0236
	[.0137]	[.0143]	[.0843]	[.0417]

NOTE: Sample size is 17,300. Standard errors in brackets, p values in parentheses. All models include a race dummy, three cohort dummies, three region of birth dummies, three indicators for region of residence in 1973, SMSA status in 1973, and imputation dummies. Other family background controls are indicators for living with both parents at age 16 and living on a farm at age 16.

men who grew up with at least one sister have lower education (about .16 fewer years) and lower earnings (about 3% less). The use of the "any sisters" indicator as an instrumental variable yields a 2SLS estimate of the return to schooling of about .193 (standard error = .066), which is 3 times as large as the corresponding OLS estimate reported in column 1. This result is consistent with the recent literature (see, e.g., Card 2000). In all studies, the IV estimates are larger than the corresponding OLS estimates, but in most cases the hypothesis that this difference is due to sampling error cannot be rejected. In this sample of men from the OCG, however, a Hausman test rejects the hypothesis that the difference between the OLS and 2SLS point estimates is due to sampling error (p value = .005).

5.3 Estimates of the Family Background Gradients

An alternative to 2SLS is a linear control function approach, as specified by (8) and (9). One advantage of the control function approach over 2SLS is that it permits the identification of the average causal effect, as well as the identification of the correlation between the endogenous variable (schooling) and the unobserved determinants of earnings a_i and b_i that are correlated with schooling. One drawback of this approach is that it requires stronger assumptions on the nature of the relationship between unobserved ability factors and the observable variables. With a larger dataset and more powerful instruments, one could relax such linearity assumptions and use the nonparametric control function methods proposed in Florens et al. (2004).

Table 6 reports the coefficients from the schooling and log-earnings control function regressions, following the specification of (7) and (8). Although interesting in its own right, this table is used to show the regression estimates from which the structural parameters of the model will be estimated using OMD. Consequently, I only briefly discuss the entries in Table 6. In column 1, coefficients from the reduced-form regression of schooling are reported. The estimates are essentially the same as the reduced-form coefficients displayed in the third column of Table 5. Column 2 reports the coefficients from the log-earnings regression specified in (8). In this specification, the residuals from the reduced-form regression of schooling are included in the earnings equation, as are their interactions with schooling. These controls will purge the other regression coefficients of any ability or endogeneity biases. The estimated average return to schooling is .161 (with standard error .063), which is smaller than the 2SLS estimate.

Table 7 presents OMD estimates of the parameters of the marginal benefit equation (in column 1) and marginal cost equation (in column 2) derived from the regression coefficients reported in Table 6. For each measure of family background, the "total effect" of that variable on the marginal return to schooling is reported in column 3. The total effect of a family background measure F_j on the marginal return to schooling corresponds to $\partial MB_i/\partial F_{ij} = b_j^F(1-\omega) + r_j^F\omega$ for $j = 1, 2, 3$, where $\omega = k_1/(k_1+k_2)$. The estimated intercepts of each schedule (b_0 and r_0) are displayed in row 1. The family background gradients for the marginal benefit and marginal cost equations (3.2) and (3.3) are shown in rows 2–4. The slopes of each schedule (i.e., k_1 and k_2) are reported in row 5. Finally, rows 6 and 7 show

Table 6. Estimates of the schooling reduced-form and "augmented" log-earnings regression

	Years of education	Log earnings
Any sisters	−.1597 [.0491]	—
Years of education	—	.1606 [.0625]
Years of education squared	—	−.0003 [.0008]
Reduced-form residuals	—	−.1212 [.0602]
Reduced-form residuals × education	—	.0016 [.0009]
Father's education[a]	.1220 [.0065]	.0010 [.0004]
Mother's education[a]	.1462 [.0073]	−.0005 [.0003]
Number of siblings[a]	−.1325 [.0072]	−.0010 [.0004]
Pupil–teacher ratio[a]	−.0303 [.0056]	−.0003 [.0002]
Relative teacher salary[a]	.5702 [.0842]	.0095 [.0033]

NOTE: Sample size is 17,300. Standard errors in brackets, p values in parentheses. See the note to Table 5 for a list of variables included in the regressions. The regressions correspond to (7) and (8) in the text.

[a]Interacted with education in the log-earnings model.

Table 7. OMD estimates of the effects of family background on the return to schooling

	(1) Marginal benefit schedule	(2) Marginal cost schedule	(3) Total effect
1. Intercept	.1580 [.0437]	.0192 [.0433]	—
2. Father's education	.0011 [.0002]	−.0003 [.0003]	.0011 [.0002]
3. Mother's education	−.0005 [.0002]	−.0023 [.0003]	−.0006 [.0003]
4. Number of siblings	−.0010 [.0003]	.0005 [.0003]	−.0009 [.0003]
5. Slopes (k_1 and k_2)	.0005 [.0010]	.0112 [.0012]	—
Other parameters			
6. Absolute ability bias (λ_S)	−.1207 [.0424]	—	—
7. Comparative advantage bias term (ψ_S)	.0015 [.0005]		
8. Pupil–teacher ratio	−.0003 [.0001]		
9. Relative teacher salary	.0071 [.0012]	—	—
10. N × Objective [df] (p value)	1.01 [1] (.002)	—	—

NOTE: Standard errors in brackets, p values in parentheses. Based on the estimated coefficients and covariance matrix from Table 6. See the Appendix for details on the OMD procedure.

the absolute ability bias term (λ_S) and the comparative advantage bias term (Ψ_S), whereas row 10 reports the goodness-of-fit statistic associated with the model.

Based on this specification, the estimated average causal effect of education is .158 (with standard error .044), which is about 20% smaller than the 2SLS estimate from these data. This is not surprising because the framework underlying Table 7 decomposes the causal effect of education into an idiosyncratic component *and* a component due to variation in family background. The average discount rate is .0192 (with standard error .043). Men with better educated fathers have a higher return to schooling, because of the positive effect of father's education on the heterogeneous component of the marginal benefit, b_i. Conversely, men with better educated mothers have a lower return to schooling on average, resulting from the large reduction in the marginal cost of schooling associated with mother's education. One possible explanation for this finding is that better educated fathers increase children education through transmission of innate ability, whereas better educated mothers raise children education by enhancing the "family learning environment." This finding can also be stated in terms of the characteristics of the production function (1): Own education and father's education are q complements, whereas own education and mother's education are q substitutes (see Sato and Koizumi 1973). Interestingly, the differential impact of father's and mother's education on educational attainment and labor market earnings is also noticeable in Tables 2 and 3. In both cases, mother's education has a larger reduced-form effect than father's education on educational attainment and labor market earnings. Finally, men raised in larger families have lower marginal returns to schooling, a result entirely attributable to lower b_i in larger families, conditional on parental education.

An important implication of these results is that measures of family size and parental education are *not* valid instrumental variables for years of education because they affect the marginal benefit of an additional year of education. This point is important because measures of parental education are commonly used as instruments for schooling in the literature. Another interesting result in Table 7 is that the number of siblings is the measure of family background with the largest effect on the return to schooling. This follows because the positive effect of father's education on the return is essentially offset by the negative effect of mother's education. The estimated slope of the marginal benefit schedule k_1 is essentially 0, which suggests that the marginal return to schooling is roughly constant. As expected, the slope of the marginal cost is positive and much steeper. This follows from the fact that the marginal benefit schedule is almost horizontal, so the marginal cost of schooling must be increasing to ensure interior solutions. In these data, the absolute ability bias is negative and relatively important in magnitude at $-.12$ (with standard error .042), whereas the comparative advantage bias term (the projection coefficient of b_i on S_i) is .0015. These results are consistent with a model of nonhierarchical sorting: Individuals with higher levels of absolute ability acquire less schooling, whereas individuals with higher benefits of schooling acquire more schooling. In other words, there is positive selection on the slope of the earnings equation and negative selection on the intercept. This result has also been found in other applications by Willis and Rosen (1979), Garen (1984), and Carneiro, Heckman, and Vytlacil (2005). Finally, the goodness-of-fit statistic for this model is 10.01, which is slightly higher than the $\chi^2_{(1)}$ critical value, suggesting that this model and its embodied exclusion restrictions provide a too simplistic statistical representation of the data.

5.4 Interpretation of the Results

Taken as a whole, the results in this article provide new evidence on observable and unobservable sources of variation in the return to education. Contrary to the results of Ashenfelter and Rouse (1998) and Altonji and Dunn (1996), the results in this article indicate that family background variables play an important role in generating variation in the return to schooling across individuals. Allowing the returns to education to vary with family background variables reduces the estimated average causal effect of education by 20%.

Moreover, the results provide clear evidence on the relative importance of the different sources of heterogeneity in explaining schooling outcomes. Because $\Psi_S = kf$ in this model, the entries in Table 7 imply that the fraction of the total variance in schooling (9.86) attributable to variation in ability as opposed to variation in the cost (or tastes) for schooling is 13%. In other words, for these data, most of the difference in educational attainment across individuals can be attributed to differences in the cost of an additional year of education.

Finally, there is no evidence that the benefits of an additional year of education are declining with the level of education. The entries in Table 7 suggest that the slope of the marginal benefit schedule is essentially 0. When combined with the negative absolute ability bias reported in Table 7, this suggests a novel interpretation of the recent findings from studies of the causal effect of education based on instrumental variables. Similar to the results in Table 5, these studies (see Card 2000 for a survey) have documented IV estimates of the return to schooling systematically exceeding the corresponding OLS estimates. The leading explanations for this pattern are as follows: (1) small (absolute) ability bias combined with a downward bias in the OLS estimate due to measurement error in reported schooling (Griliches 1979; Angrist and Krueger 1991); and (2) heterogeneity in the returns to education (along with declining marginal benefit to educational investments) combined with instrumental variables that affect the schooling outcomes of individuals who would have relatively low schooling in the absence of the supply-side innovation (see, e.g., Angrist and Imbens 1995). The results in Table 7 are *not* consistent with these two explanations. First, attenuation bias alone cannot explain the large gap between the OLS and 2SLS estimates in Table 5: Using reinterview data, Bielby, Hauser, and Featherman (1977) reported that the reliability of reported schooling is about 94% in the OCG. Second, the marginal returns to education are essentially constant across education levels (i.e., $k_1 \approx 0$). Therefore, as implied by the results of Tables 5 and 7, one novel explanation for the larger IV estimates is a *negative* absolute ability bias in the OLS estimates. Under the assumptions of Section 3, it can be shown that the OLS estimate reported in Table 5 converges to $b_0 + \lambda_S + \Psi_S\overline{S}$; that is, the OLS estimate is confounded by an ability bias (λ_S) and a comparative advantage bias ($\Psi_S\overline{S}$). As long as $|\lambda_S| > |\Psi_S\overline{S}|$, a negative value for λ_S implies that the simple OLS estimate is biased downward. With the relatively high reliability of reported schooling in these data (94%), and essentially no concavity in the "structural" earnings function, this is the only explanation why the 2SLS estimates are about 3 times as large as the OLS estimates of the return to education.

5.5 Sensitivity Analysis

Finally, I briefly describe the results of a sensitivity analysis. The full set of results is available in the result appendix (Deschenes 2006). Three issues are investigated: (1) the validity of the exclusion restrictions, (2) measurement error in own and parental schooling, and (3) robustness of the OMD to finite-sample bias. I discuss the results of the sensitivity analysis in that order.

As discussed in the previous section, sibling gender composition (conditional on family size) can be rightfully excluded from the marginal benefit equation (3.2) if it affects educational attainment but has no independent effect on earnings. If gender composition is still related to unobserved determinants of earnings ability after controlling for family background and size, then it does not satisfy the exclusion restriction. Although this assumption is not directly testable (because a_i, b_i, and r_i are all unobservable), various pieces of evidence can be examined to evaluate its validity.

First, I use data from the Project TALENT (1960) study of 400,000 high school students in grades 10–12 in 1960. These students responded to a lengthy questionnaire and cognitive tests during the two full days of class devoted to the study (see Kuhn and Weinberger 2005 for a more detailed description of these data). I estimate regression models relating the different measures of ability (such as math and verbal test scores) on measures of family background and sibling composition, grade dummies, and a host of demographic variables. The estimated effects of sibling gender composition on the various test scores are small in magnitude and insignificant, whereas parental education is a strong predictor of achievement. Second, with the OCG data I used the interaction between an index of "poor" family background and the presence of any sisters as an additional instrumental variable. This approach is motivated by the fact that the negative effect of the presence of any sisters on the educational attainment of men (conditional on family size) should be larger for poorer households if it only operates through the marginal cost of schooling. The reduced-form estimates confirm that the (negative) effect of the "any sisters" variable on educational attainment is larger for individuals with poorer family backgrounds. The 2SLS estimate of the return to schooling is slightly smaller but more precisely estimated than the 2SLS estimate reported in Table 5. The results also show that the presence of any sisters has a very small and insignificant effect on log earnings (−.003, with a standard error of .011). These two simple tests do not provide any evidence against the maintained assumption that holding family size and background constant, the gender composition among siblings is unrelated to the unobserved determinants of earnings.

It is well known that 2SLS estimates are not affected by classical measurement error, although they are possibly biased under other forms of measurement error (see, e.g., Kane, Rouse, and Staiger 1999). Less is known, however, on the effects of classical measurement error in nonlinear models, and few analytical formulas describing the bias are available. In the context of the model presented in Section 3, identification of the parameters of the marginal benefit schedule requires consistent estimates of the interaction terms between family background and schooling in the augmented log-earnings regression. Conversely, identification of the parameters of the marginal cost schedule, based on the schooling reduced-form estimates, is not affected by classical measurement error in schooling. Even if schooling and father's education are reported with classical measurement error, the nonlinearity due to the interaction terms introduces nonclassical measurement error in the regression. The asymptotic bias in the OLS slope estimates depends on multiple features of the (unobserved) joint distribution of the measurement error components. I assume a reliability of .90 for own schooling and .85 for father's education and use simulations to assess the magnitude of the biases (see Bielby et al. 1977 for analysis of the measurement error in schooling in the OCG). The results of a simple analysis suggest that the higher order terms are more sensitive than the main effects. In all specifications used, the simulation results suggest that the interaction between schooling and father's education is biased downward and that the relevant entries in Table 7 should be inflated by a factor of 1.25. Therefore, the results presented in this study should be interpreted as conservative estimates of the effects of family background on the return to schooling.

In certain applications, OMD can be biased downward (in absolute terms) if there is correlation between the sampling error in the vector of moments and the sampling error of the elements in the weighting matrix (Altonji and Segal 1996). Following Altonji and Segal (1996), who concluded that equally weighted minimum distance (EWMD) dominates OMD, I estimate three alternatives to OMD: EWMD, variance-weighted minimum distance (VWMD), and direct "one-step" nonlinear least squares (NLLS) estimates of the parameters. These alternative estimates of the parameters are similar to those reported in Table 7.

6. CONCLUSION

This article develops and implements a simple model of schooling and earnings where the return to schooling varies across individuals and where family background characteristics play a direct role in generating the heterogeneity. The model illustrates the influence of family background variables in the optimizing behavior of individuals and, thus, generalizes the standard causal model of schooling and earnings. It is shown that a correct assessment of the relationship between family background and return to schooling entails the identification of the effect of family environment on both the marginal benefit and the marginal cost of schooling. Only with such information can the impact of family background characteristics on subsequent labor market outcomes be determined.

The empirical analysis, based on a large sample from the 1973 Occupational Change in a Generation survey, documents several patterns concerning the relationship among family background, schooling, earnings, and return to schooling. Parental education raises both the educational attainment and the labor market earnings received by individuals. Men from larger families acquire less schooling and have lower earnings. Moreover, the negative effect of family size is shown to vary with the gender composition of the sibship. When holding family background and sibship size constant, men raised with more sisters have lower educational attainment and earnings. These inferences are robust to a wide variety of specifications.

The patterns are then interpreted in the context of the model. The identification of the parameters of the marginal benefit and cost functions requires two exclusion restrictions. This article considers measures of elementary and secondary school quality as variables affecting only the marginal benefit of schooling, conditional on family background. The randomness embodied in the gender composition among siblings, holding family size and background constant, is used as an exogenous source of variation affecting only the marginal cost of schooling. The results provide new evidence on the effects of family background on the return to schooling. Men raised in larger families have a lower return to schooling, with each additional sibling reducing the return to schooling by as much as 5% of the conventional Mincerian estimate. The *combined* effects of parental education are more modest. Men who grew up with better educated fathers have a higher return to education, whereas those who grew up with better educated mothers have a lower return to schooling. This finding suggests that own education and father's education are q complements in the production of earnings capacity, whereas own education and mother's education are q substitutes in the production of earnings capacity. Overall, accounting for family background differences reduces the estimate of the average causal effect of education by 20%. The disparity of these results clearly indicates that different aspects of familial environment have different effects on the marginal benefit and marginal cost of schooling. Family size and father's education entirely operate through the benefits of an additional year of education. The negative impact of maternal education on the return to schooling is solely attributable to lower costs per year of education and, hence, improved educational prospects. These insights should be a key component of any appraisal of policies targeted at children from disadvantaged families.

Finally, this article documents new facts about the sources of variation in schooling outcomes and in the components of the return to education. A new finding is that almost 90% of the total variance in schooling outcomes is attributable to differences in the costs (or tastes) of schooling, as opposed to differences in ability. There is little evidence of declining marginal benefits of an additional year of education (i.e., the human capital production function is linear in years of education). The results of the optimizing model of schooling choices indicate a negative absolute ability bias and a positive comparative advantage bias. These last two pieces of evidence support a new interpretation of why the IV estimates exceed the OLS estimates of the return to education: A negative absolute ability bias combined with constant marginal benefit to schooling is the only explanation consistent with the results presented in this article.

ACKNOWLEDGMENTS

I am grateful to Orley Ashenfelter and David Card for many helpful discussions. I also thank Ted Bergstrom, Ken Chay, David Lee, Peter Kuhn, the associate editor, two anonymous referees, and seminar participants at the University of California, Berkeley, and the University of California, Santa Barbara, for their comments.

APPENDIX: OPTIMAL MINIMUM–DISTANCE ESTIMATION

Let π denote the vector of relevant regression coefficients from (7) and (8):

$$\pi = [\pi_{10}, \pi_{11}^F, \pi_{12}^F, \pi_{13}^F, \pi_{11}^Q, \pi_{12}^Q, \\ \pi_{21}, \pi_{22}, \pi_{23}, \pi_{24}, \pi_{21}^F, \pi_{22}^F, \pi_{23}^F, \pi_{21}^Q, \pi_{22}^Q]'.$$

When multiple family background and school quality measures are included, the results of Section 4.1 generalize to

$$\begin{aligned}
\pi_{10} &= \frac{b_0 - r_0}{k_1 + k_2}, & \pi_{21}^F &= b_1^F,\\
\pi_{11}^F &= \frac{b_1^F - r_1^F}{k_1 + k_2}, & \pi_{22}^F &= b_2^F,\\
\pi_{12}^F &= \frac{b_2^F - r_2^F}{k_1 + k_2}, & \pi_{23}^F &= b_3^F,\\
\pi_{13}^F &= \frac{b_3^F - r_3^F}{k_1 + k_2}, & \pi_{21}^F &= b_1^Q,\\
\pi_{11}^Q &= \frac{b_1^Q}{k_1 + k_2}, & &\\
\pi_{21}^Q &= b_2^Q, & &\\
\pi_{12}^Q &= \frac{b_2^Q}{k_1 + k_2}, & &\\
\pi_{21} &= b_0 - \sum_{j=1}^{3} b_j^F \overline{F}_j - \sum_{k=1}^{2} b_k^Q \overline{Q}_k, & &\\
\pi_{22} &= -.5k_1, & &\\
\pi_{23} &= \lambda_S, & &\\
\pi_{24} &= \Psi_S. & &
\end{aligned}$$

Optimal minimum-distance estimates (OMDs) are obtained by minimizing the following quadratic form:

$$\widehat{\theta} = \min_{\theta} [\widehat{\pi} - f(\theta)]' \widehat{W} [\widehat{\pi} - f(\theta)],$$

where $\widehat{\pi}$ is the vector of estimated regression coefficients, $f(\theta)$ is the vector of restrictions imposed by the model on the regression coefficients, $\widehat{W}$ is an estimate of the inverse covariance matrix of $\widehat{\pi}$, and θ is the vector of parameters:

$$\theta = [b_0, b_{11}, b_{12}, b_{13}, b_{Q1}, b_{Q2}, k_1, r_0, r_{11}, r_{12}, r_{13}, k_2, \lambda_S, \Psi_S]'.$$

Chamberlain (1984) showed that, under mild regularity conditions, the optimal minimum-distance estimator is asymptotically efficient. Moreover, the value of the objective function at the optimum can be used to perform specification tests of the model.

[Received October 2004. Revised May 2006.]

REFERENCES

Altonji, J. G., and Dunn, T. A. (1996), "The Effect of Family Characteristics on the Return to Schooling," *Review of Economics and Statistics*, 78, 665–671.

Altonji, J. G., and Segal, L. M. (1996), "Small-Sample Bias in GMM Estimation of Covariance Structures," *Journal of Business & Economic Statistics*, 14, 353–366.

Angrist, J. D., and Imbens, G. W. (1995), "Two-Stage Least Squares Estimation of Average Causal Effects in Models With Variable Treatment Intensity," *Journal of the American Statistical Association*, 90, 431–442.

Angrist, J. D., and Krueger, A. B. (1991), "Does Compulsory School Attendance Affect Schooling and Earnings," *Quarterly Journal of Economics*, 106, 979–1014.

Ashenfelter, O., and Rouse, C. E. (1998), "Income, Schooling and Ability: Evidence From a New Sample of Twins," *Quarterly Journal of Economics*, 113, 253–284.

Becker, G. S. (1967), *Human Capital and the Personal Distribution of Income*, Ann Arbor: University of Michigan Press.

Berhman, J. R., Pollack, R. A., and Taubman, P. (1982), "Parental Preferences and Provision for Progeny," *Journal of Political Economy*, 90, 52–73.

Bielby, W. T., Hauser, R. M., and Featherman, D. L. (1977), "Response Errors of Black and Nonblack Males in Models of Intergenerational Transmission of Socioeconomic Status," *American Journal of Sociology*, 82, 1242–1288.

Black, S., Devereux, P., and Salvanes, K. (2005), "The More the Merrier? The Effects of Family Size and Birth Order on Children's Education," *Quarterly Journal of Economics*, 120, 669–699.

Butcher, K. F., and Case, A. (1994), "The Effect of Siblings' Sex Composition on Women's Education and Earnings," *Quarterly Journal of Economics*, 109, 531–563.

Card, D. (1999), "The Causal Effect of Education on Earnings," in *Handbook of Labor Economics*, Vol. 3, eds. O. Ashenfelter and D. Card, Amsterdam: North-Holland, pp. 1802–1859.

——— (2000), "Estimating the Return to Schooling: Progress on Some Persistent Econometric Problems," *Econometrica*, 69, 1127–1160.

Card, D., and Krueger, A. B. (1992), "Does School Quality Matter: Returns to Education and the Characteristics of Public Schools in the United States," *Journal of Political Economy*, 100, 1–40.

Carneiro, P., Heckman, J., and Vytlacil, E. (2005), "Understanding What Instrumental Variables Estimate: Estimating Marginal and Average Returns to Education," working paper, University College London.

Chamberlain, G. (1984), "Panel Data," in *The Handbook of Econometrics*, Vol. 2, eds. Z. Griliches and M. Intrilligator, Amsterdam: North-Holland, pp. 1247–1318.

Deschenes, O. (2006), "Result Appendix to Estimating the Effects of Family Background on the Returns to Education," working paper, University of California, Santa Barbara, Dept. of Economics.

Florens, J.-P., Heckman, J., Meghir, C., and Vytlacil, E. (2004), "Instrumental Variables, Local Instrumental Variables, and Control Functions," Working Paper CWP15/02, Institute for Fiscal Studies.

Garen, J. (1984), "The Returns to Schooling: A Selectivity Bias Approach With a Continuous Choice Variable," *Econometrica*, 52, 1199–1218.

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

Hanushek, E. A. (1986), "The Economics of Schooling: Production and Efficiency in Public Schools," *Journal of Economic Literature*, 24, 1141–1177.

Hausman, J. A. (1978), "Specification Tests in Econometrics," *Econometrica*, 46, 1377–1398.

Heckman, J. J., Layne-Farrar, A., and Todd, P. (1996), "Does Measured School Quality Really Matter? An Examination of the Earnings-Quality Relationship," in *Does Money Matter? The Effects of School Resources on Student Achievement and Adult Success*, Washington, DC: Brookings Institution, pp. 192–288.

Heckman, J. J., and Vytlacil, E. (1998), "Instrumental Variables Methods for the Correlated Random Coefficient Model: Estimating the Rate of Return to Schooling When the Return Is Correlated With Schooling," *Journal of Human Resources*, 23, 974–987.

Kane, T. J., Rouse, C. E., and Staiger, D. (1999), "Estimating Returns to Schooling When Schooling Is Misreported," Working Paper 7235, National Bureau of Economic Research.

Kuhn, P., and Weinberger, C. (2005), "Leadership Skills and Wages," *Journal of Labor Economics*, 23, 395–436.

Rosenzweig, M. R., and Wolpin, K. I. (2000), "Natural 'Natural Experiments' in Economics," *Journal of Economic Literature*, 38, 827–874.

Sato, R., and Koizumi, T. (1973), "On the Elasticities of Substitution and Complemetarity," *Oxford Economic Papers*, 25, 44–56.

Taubman, P. (ed.) (1977), *Kinometrics: The Determinants of Socio-Economic Success Within and Between Families*, Amsterdam: North-Holland.

Willis, R., and Rosen, S. (1979), "Education and Self-Selection," *Journal of Political Economy*, 79, S7–S36.

Woolridge, J. M. (2000), "Instrumental Variables Estimation of the Average Treatment Effect in the Correlated Random Coefficient Model," mimeo, Michigan State University, Dept. of Economics.