Center for Demography and Ecology

University of Wisconsin-Madison

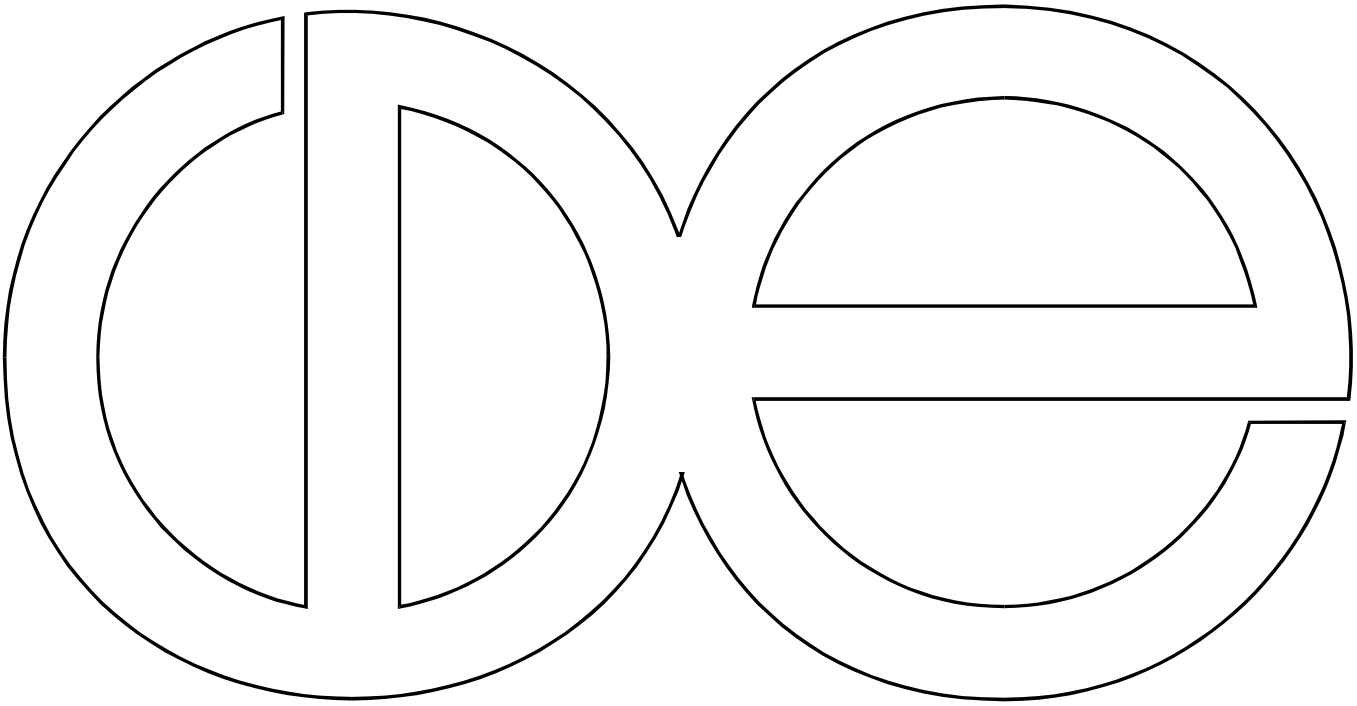
The Role of Expectations in Adolescent Schooling Choices: Do Youths Respond to Economic Incentives?

Kathryn Wilson

Barbara Wolfe

Robert Haveman

CDE Working Paper No. 2004-21



The Role of Expectations in Adolescent Schooling Choices:

Do Youths Respond to Economic Incentives?

Kathryn Wilson

Barbara Wolfe

Robert Haveman*

JEL codes: I20, J24

The Role of Expectations in Adolescent Schooling Choices: Do Youths Respond to Economic Incentives?

Abstract

We address the role of youths' own choice-conditioned expectations in understanding their schooling choices by constructing a choice (or "switching") model. We emphasize the effect of individual student perceptions regarding the "returns" associated with graduating from high school versus dropping out, while controlling for an extensive set of family and community factors. We find that youths' expected income returns to graduating from high school are influential in their schooling choices, even when an extensive set of background, economic, family, and neighborhood variables, designed to capture the effects of parental and governmental decisions, is introduced into the analysis.

Kathryn Wilson
Associate Professor of Economics
Kent State University

Barbara Wolfe
Professor of Economics, Population Health Sciences, and Public Affairs
University of Wisconsin–Madison

Robert Haveman
John Bascom Emeritus Professor of Economics and Public Affairs
University of Wisconsin–Madison

The Role of Expectations in Adolescent Schooling Choices:

Do Youths Respond to Economic Incentives?

Over the postwar period, the level of education attained by individuals in nations of the Organisation for Economic Co-operation and Development (OECD) has increased steadily. This is especially true for the United States.[1] In 1950, for example, the average American aged 25 or older had 9.3 years of education; by 2000, this had increased to 12.7 years. This increase reflects sizable increases in the rate of high school completion. In the cohort aged 16–24 in 1960, more than 27 percent were counted as high school dropouts; by 2001, the proportion of high school dropouts in this age cohort had fallen to 10.7 percent. Although the proportion of whites in any given age cohort who graduated high school has exceeded that for African-Americans over this period, this gap narrowed significantly during the 1970s and 1980s, and by 2000 had nearly closed.[2]

Observed trends in and patterns of educational attainment, human capital accumulation, and labor market success are determined by several factors. In addition to changes in a wide variety of background factors (such as ability, motivation, race, and gender), these involve decisions made by governments (as reflected in school resources and macroeconomic performance) and by parents (e.g., parental education, work effort, and child care time). Also relevant are the choices made by adolescents themselves in response to the net economic returns of the options open to them; in this paper we explore the schooling choices of youths in response to their perception of these returns.

Economic and sociological research on the determinants of the schooling outcomes of youths is extensive and distinguished.[3] These studies focus primarily on the family,

neighborhood, and peer group determinants of educational attainments, and are largely reduced-form in nature. The role of youth preferences, and the choices made in response to the economic incentives and constraints that young people face, have not been systematically explored in these studies. Hence, their results may attribute to background and family characteristics effects properly attributed to differences in youths' own choices in response to the incentives that they face.[4]

In this study, we address the role of youths' own choice-conditioned expectations in understanding their schooling choices by constructing a choice (or "switching") model. We emphasize the effect of individual student perceptions regarding the "returns" associated with graduating from high school versus dropping out, while controlling for an extensive set of family and community factors. Youths are presumed to form their expectations by referring to the outcomes of a slightly older cohort of individuals who have characteristics similar to their own. Because these expected income returns are choice-specific, we require estimates of each youth's choice-conditioned expected returns, taken to be unobserved by the analyst but known by the youth.[5]

We focus on the high school graduation outcome for three reasons. First, individuals must complete high school in order to consider postsecondary schooling options. Hence, the decision to complete high school carries with it an option value benefit.[6] Second, high school completion—the first schooling decision over which adolescents have some control—plays a crucial role in understanding the declining labor market position of low-skilled workers.[7] Finally, our analysis of the high school graduation choice is nested within an extensive economic literature that addresses the determinants of this outcome.

By focusing on the role of youth's own expectations, we have neglected other aspects of a complete analysis of the process of educational attainment. For example, we presume that our extensive set of family and neighborhood factors are exogenous to the outcome and independent of unobserved factors. Although we do not explicitly include school characteristics in our estimation, our detailed neighborhood and family variables are likely to accurately proxy for neighborhood school quality.[8]

In the next section, we present a utility-maximization model of an adolescent's choice regarding whether or not to drop out of high school. After describing the data, we explain our procedures for estimating the conditional expectations that we attribute to older teenagers, and present the results from estimation of our model. In this section, we also discuss the identification of the model. Our results indicate that youths do respond to economic incentives when making schooling choices, even when controlling for an extensive set of background, family, and neighborhood factors. Because this conclusion rests on a particular set of identifying restrictions in the form of determinants of expected earnings that are not directly related to the education choice, we test the robustness of our results to alternative assumptions, definitions, and identifying restrictions. We then use the estimated structural parameters to simulate the effects of changes in future choice-specific income expectations (and other variables) on the probability of high school completion.

I. A MODEL OF THE HIGH SCHOOL GRADUATION CHOICE

In our model of the high school graduation decision, youths respond as rational utility maximizers to expectations of utility (income) returns associated with the options available to

them. The utility function for a 16-year-old youth who values both consumption (C) and schooling (S) is:

$$U = U(C, S). \quad (1)$$

If we assume that the utility from schooling depends on a variety of background, family, and community variables, assign a linear functional form to utility from schooling, and allow different utility weights to be attached to schooling and consumption, the well-being of youth *i* is seen to be dependent on S and C:

$$U_{si} = B_s X_i + Z C_{si} + \mathbf{g}_{si}\,, \quad (2)$$

where X_i = a vector of nonconsumption (background, family, and community) variables that directly affect utility from schooling,

B_s = weights attached to nonconsumption utility gains of schooling,[9]

C_{si} = lifetime discounted stream of consumption conditional on graduation or not,

Z = weight of consumption in utility, and

$\mathbf{g}_{si}$ = random utility term conditional on graduation or not.

This function allows the utility effect of graduation to differ among youths depending on their family and community characteristics. These characteristics (e.g., parental education) may affect the perceptions and aspirations of young people and hence their assessment of the utility effects of high school graduation (including the option to select post–high school schooling options) apart from the income consequences of this choice. Since the consumption term, C_{si}, is separable and the weight of consumption on utility (Z) is not conditional on schooling level, utility from income is invariant to the graduation decision.

The youth maximizes utility subject to the following budget constraints:

$$Y_{si} = \$_s Q_i + >_{si} \quad (3a)$$

$$C_{si} \# Y_{si} \quad (3b)$$

where: Y_{si} = lifetime discounted income stream conditional on graduation or not,

$\$_s$ = schooling-specific parameters describing the influence of background, family, and community factors on income[10],

Q_i = variables which affect income, and

$>_{si}$ = random component of income conditional on graduation or not.

We presume that youths do not know their future income prospects with certainty, and form expectations regarding their choice-conditioned economic position by observing the incomes of a slightly older cohort of young adults with similar characteristics. Some of the youths in this older cohort have chosen to graduate from high school (g), whereas others have chosen to drop out (d).

$$E[Y_i|Q_i,g] = \$_g Q_j + >_{g\,j} \quad \text{for } Q_i = Q_j \quad (4a)$$

$$E[Y_i|Q_i,d] = \$_d Q_k + >_{d\,k} \quad \text{for } Q_i = Q_k \quad (4b)$$

If *i* (whose choice we analyze) chooses to graduate from high school, his/her expected income will equal that of *j*, a member of the reference group who graduated from high school, and who has a set of characteristics that influence income and that are like those of *i*. Similarly, if *i* chooses to drop out, his/her expected income will equal that of *k*, a member of the reference group who dropped out of high school with a set of characteristics that influence income similar to those of *i*.

The individual will choose to graduate from high school if the expected utility from graduating is greater than the expected utility from not graduating.

$$U_{gi} = B_g X_i + ZC_{gi} + \varepsilon_{gi} > B_d X_i + ZC_{di} + \varepsilon_{di} = U_{di} \quad (5)$$

Rearranging terms and incorporating the budget constraint, the probability that an individual will choose to graduate from high school is:

$$Pr_g = Pr[\varepsilon < BX + Z\{E[Y_g] - E[Y_d]\}] \quad (6)$$

where $\varepsilon = \varepsilon_d - \varepsilon_g$ and $B = B_g - B_d$.

Hence, we focus on the effect on the schooling choice of the expected income possibilities available to youths under both schooling options.[11] These expectations reflect the economic incentives that will influence youths' decisions. They are, in turn, influenced by the characteristics and prior choices of the youth, of their families, and of the larger society. These characteristics and prior choices can also directly affect youths' decisions, apart from any influence they have on the income expectations. Youths' choices of whether or not to graduate from high school are often made simultaneously with fertility (e.g., whether to have a birth out of wedlock) or marital decisions, but our model does not reflect such complex joint and dynamic interaction among decisions.

The decision rule that maps expected returns into choices can be learned through empirical study only if data on both expectations and choices are available; however, available data sets reveal only the choices. In this circumstance, maintained assumptions about the expectations formation process are required to estimate decision rules conditional on those assumptions. If this process is, in fact, unknown, the estimation of expected income streams may be impossible. See Manski (1993a, 1993b).

Our approach to modeling the determinants of youths' schooling choices involves specifying the process by which individuals form expectations regarding the outcomes

associated with the choices available to them. In particular, the expectations of the youths whose choices we seek to understand are taken to be the observed returns associated with the choices experienced by similar individuals in a relevant reference group.[12] This procedure assumes that 16-year-old youths take those observed variables related to the schooling choice available to the researcher as being the determinants of the decisions of individuals in the reference group. In the absence of evidence regarding the influence of additional, unobserved factors on youth choice-conditioned expectations and schooling choices, we prefer this assumption to assuming that individuals in our primary sample also understand the unobserved selection process into educational level affecting members of the reference group. However, we do test the sensitivity of our results to those from a model that assumes that youths do perceive the effects of such unobserved factors in forming expectations; see below.

Our model is also related to that of Willis and Rosen (1979), who were among the first to directly introduce explicitly modeled perceptions of choice-related outcomes into an analysis of the youth schooling decision. In their model, family characteristics affect the schooling choice through their effect on the discount rate used by youths in forming present values of streams of future outcomes. Consistent with the education production literature, we allow family and neighborhood characteristics to affect both the expected returns to schooling and the nonpecuniary utility from additional schooling. Although Willis and Rosen assume rational expectations, presuming that youths do not systematically form incorrect predictions of their own labor market outcomes, we assume that youths look to the experiences of a reference group in forming their expectations. However, we also present results based on the rational expectations framework.

II. DATA ON THE PRIMARY AND SECONDARY YOUTH COHORT SAMPLES

Our estimates are based on two large longitudinal data sets constructed from a national stratified sample of families, the Michigan Panel Study of Income Dynamics (PSID).[13] The first data set—our primary sample composed of younger individuals whose schooling choices we model—includes 1,942 youths who were aged 0–6 years in the beginning year of the survey; they were followed until 1993, at which time they were young adults, ranging in age from 25 to 31 years.[14] A secondary sample consisting of a somewhat older cohort includes 1,326 youths who were aged 8–12 years in 1968, and who were 33 to 37 years old in 1993; this sample is used to estimate the choice-specific income expectation variables employed in our choice model.

For individuals in both data sets, we have extensive information on family status, income and source of income, parental education, neighborhood characteristics, and background characteristics such as race, religion, and location. In order to make comparisons of individuals with different birth years, we indexed the time-varying data elements in each data set by age. Individuals are considered to have graduated from high school if at any time during the sample period they are reported as graduates.[15] All monetary values are expressed in 1997 dollars using the Consumer Price Index (CPI) for all items.[16]

Finally, we added neighborhood information to our primary data set, constructed by matching small-area data from the 1970 and 1980 censuses to the location of the children.[17] The merged neighborhood data include information on the proportion of youths in the neighborhood that dropped out of high school and the proportion of high-income families in the neighborhood in which the family of each child in our primary sample lived for each year from 1968 to 1985.

III. INCOME EXPECTATIONS WITH SCHOOLING CHOICES

Empirical specification of our model relies on estimation of the choice-conditioned income expectations that we attribute to youth decision makers. For each youth in our sample, we calculate an annual expected income value for each relevant year, and then take the present discounted value of each choice-conditioned income stream as of age 16; this pair of estimates provides the core values for the expected personal income terms.

We rely on our secondary, older cohort to estimate the two choice-specific expected personal income[18] variables for each of the 1,942 youths in the primary sample. These expectation variables are obtained from estimated parameters of a series of personal income Tobit equations fit over observations in the secondary sample, together with the relevant observed characteristics of the youths in our primary sample (equation 4a). A Tobit maximum likelihood estimation procedure is used because there are a sizable number of observations with no reported income, especially at younger ages. We estimate 14 Tobit equations (one for each year from ages 19 to 32) for each of the two schooling groups—the group of high school graduates and the group that did not complete high school—with personal income as the dependent variable, a total of 28 equations.[19] The results of the first-stage estimates are generally as expected and are available from the authors.[20]

We use the relevant individual characteristics of each youth in our primary sample, together with the coefficient estimates from the two sets of 14 age-specific regressions fit over the reference sample, to predict income values (for each age from 19 to 32) for each primary sample observation.[21] Two 14-year series of predicted income expectations are obtained for each youth; one series represents their expected income trajectory over ages 19 to 32 conditional on

graduating from high school, and another 14-year series represents their expected income trajectory if they do not graduate.[22]

The weighted mean values of these predicted personal income expectations (and the standard deviations) are shown in Table 1 for each of the 14 years for each of the assumed schooling outcomes. These mean predicted values are shown for the entire primary sample, and for each of the two schooling groups in that sample.

The schooling-conditioned expected income patterns are revealing. During the first few years, predicted income assuming graduation is similar to that assuming failure to graduate. This reflects a combination of the higher earnings of high school graduates who directly enter the labor market; and the low (if any) earnings of those who continue their schooling. The income trajectory if youths do not graduate shows almost no real growth after age 22. In contrast, mean expected income for those who graduate increases over the 14 years. Beginning at age 22, the relative predicted income trajectories for the high school graduates reveal substantial growth. (Again, note that the predicted incomes with and without a high school degree are for the same individuals.)[23]

The bottom panels of Table 1 compare predicted incomes for the nongraduates (with their characteristics, and assuming both graduation and failure to graduate) with the conditional predicted incomes of the graduates. The predicted incomes without graduation of those who actually do not graduate are similar to the predicted incomes of the graduates if they had not graduated. However, after age 23 the predicted incomes of the actual graduates diverge rapidly from those of the nongraduates, even if we assume the latter had graduated. Perhaps most interestingly, youths who graduate have lower expected income if we assume that they do not

graduate than do those youths who in fact did not graduate; this suggests the existence of rational sorting. This finding is consistent with the theory of comparative advantage and with the empirical findings of Willis and Rosen (1979) for college attendance.[24]

We discount each of the choice-conditioned expected income streams through ages 19 to 32 for each youth in the primary sample up to age 16 (the most likely age for deciding whether or not to finish high school), using a discount rate of 3 percent. This procedure implicitly assumes that youths in our primary sample form their expectations of future schooling-conditioned incomes at age 16 by observing the realized incomes of persons like themselves who are in their late teens through early 30s. By including incomes over this age we capture the opportunity costs in the form of forgone earned income associated with education beyond high school.[25]

The discounted present values of these mean income streams are shown at the bottom of each panel in Table 1. The expected present value of income for the average youth in the sample, if he/she were to graduate high school, is $177,550; the average expected present value absent graduation is $97,212, for a difference of $80,338. The average gain from graduating high school is greater for those who, in fact, do graduate ($84,945) than for those who fail to graduate ($53,727), again suggesting rational sorting of individuals over the available schooling options.

IV. ESTIMATION OF THE HIGH SCHOOL GRADUATION CHOICE MODEL

In a second stage, we estimate the youth's high school graduation decision, focusing on the role played by the income expectations variables. For each individual, the difference in the present value of the two income predictions—the present value of income conditional upon

graduation minus the value conditional on not graduating; $E[Y_g] - E[Y_d]$—is taken to reflect the expected net opportunity gain associated with the decision to obtain a high school degree, and it is included in our model of the decision of whether or not to graduate high school.

$$(\text{Prob. HS grad} = 1) = BX_i + Z\,(E[Y_g] - E[Y_d]) - \varepsilon \qquad (7)$$

We estimate this model using a probit specification, similar to switching models in Manski (1987) and Lee (1979). The dependent variable in the model is equal to 1 if the youth graduated from high school, and 0 otherwise. The vector of coefficients B indicate how background, family, and community factors affect the utility of education, and hence the education choice, apart from their effect on expected income. The coefficient Z indicates the effect of the present value of the income difference on the high school graduation decision. If this coefficient is significant, we conclude that youths do respond to economic incentives in making their education choice.

When the expected income difference variable is taken to be the only factor influencing the high school graduation decision, the coefficient is positive and statistically significant (t-statistic = 7.8). Although this strong effect suggests that youths respond to economic opportunities in making schooling choices, it does not control for the direct effect of background, family, and neighborhood characteristics on the schooling choice.

A. Our Preferred Model

Columns 1–3 of Table 2 present our preferred estimate of the determinants of the high school graduation decision, equation (7). This two-stage framework is consistent with the utility maximization model of Section I, in which the error term in equation (2) reflects heterogeneity in

tastes and other unobserved factors. This model assumes that 16-year-old youths perceive the incomes of those high school graduates and nongraduates whose observed characteristics are similar to their own, and form their conditional expectations on the basis of these outcomes.[26] Our specification includes an extensive set of family and neighborhood variables, in addition to the expected income difference variable.

The coefficient on the net income gain of graduating from high school is positive and statistically significant, which supports the hypothesis that youths' expectations regarding the economic consequences of schooling attainment do influence the choices that they make; increasing the expected gain to graduating from high school seems likely to increase the prevalence of this outcome.[27]

Our findings on the responsiveness of individuals to their expectations of the consequences of choices are consistent with the findings of others that explore the role of income in determining choices to attend college or select particular majors. For example, Freeman (1971) found that college students respond to income incentives when selecting a major. Holding constant nonpecuniary characteristics, choice among degrees responds significantly to marginal income incentives (coefficient = .7). Our findings are also consistent with those of Manski and Wise (1983), who found that higher local wages reduced college attendance (coefficient = -.136), and Willis and Rosen (1979), who estimated that expected lifetime earnings gains associated with incremental schooling influence the decision to attend college; their estimated elasticity of enrollment to earnings, about 2, is marginally significant.[28] Our results are also consistent with a recent study (Jacob 2002) that finds that higher relative rates of return to higher education for women explain a large proportion of the gender difference in college attendance. Relative to these

prior studies, ours is unique in that the use of PSID data allow us to estimate the response of choice to income expectations based on individual characteristics.

Our estimates of the coefficients on other variables in the model are also consistent with many social science models of children's educational attainments. The race and gender variables suggest that African-American females are more likely than others to graduate once a variety of other factors are taken into account.[29] Young adults who have parents with a high school diploma are themselves more likely to choose to graduate, although having a parent who graduated from college does not further increase the probability of graduating from high school.[30] Those who spend more years during ages 6–15 living with a family that receives welfare (Aid to Families with Dependent Children, AFDC) benefits, or who experience more geographic moves, or whose parent is disabled are less likely to remain in school until graduation. Growing up in a poor family seems to play a very small role, whereas being the oldest child has a larger impact.[31] The neighborhood where a youth grows up also seems to play some role: those living in areas with a greater proportion of high school dropouts are themselves less likely to graduate.[32]

We accurately predict the outcome for 81 percent of the observations. The proportion is very high for those who graduated (97 percent); we correctly predict the nongraduation outcome in 15 percent of the cases.

B. Model with Sample Selection

Our preferred specification assumes that 16-year-old youths perceive the incomes of those individuals with observed characteristics similar to their own who did and did not graduate from high school, and that in forming their expectations these youths assume that members of the

reference group have been randomly assigned to the high school graduate or dropout groups. An alternative view is that assignment of the reference sample to the two educational outcomes is the result of a selection process that reflects all relevant determinants of schooling choice, and that individuals in the primary sample know this selection process and account for it in forming their prediction of expected personal income conditional on the schooling choice.

As a sensitivity test, we estimated an alternative specification that attempts to statistically control for this selection.[33] A two-stage Heckman-type selectivity correction model is fitted over the secondary, older cohort to estimate the two choice-specific expected personal income variables for each person in the primary sample.[34] First, we estimate a reduced-form probit equation with high school graduation as the dependent variable.[35] The appropriate lambda selectivity correction variable derived from this estimation is then included in each of the 14 age-specific personal income equations for each education group [equations (4a) and (4b)] in order to control for selection into one of the education groups.[36]

As in the preferred model, the difference in the present value of the two 14-year conditional income streams is the predicted term of interest.[37] Columns 4–6 of Table 2 indicate that the results from this three-stage model with sample selection are very similar to the preferred model. The income term is again positive and significant with a marginal effect that is slightly smaller than in the preferred model. The results for the other independent variables are also very similar to those in columns 1–3. We conclude that irrespective of the assumption made regarding the process of forming expectations, the educational choices of youths are guided by their perceptions of the incomes associated with the different options.

C. A Note on Model Identification

As in prior studies of the effect of expectations on youths' schooling decisions, our model needs to be identified if the results are to be reliable. We adopt a standard exclusions restriction approach to model identification; variables expected to affect income expectations but not the schooling choice (other than through the income terms) are included in the income estimations.[38] In our preferred model, variables which are traditionally related to earnings provide this identification for each age over which income is estimated.[39]

In order for the exclusion restrictions to provide valid identification, the variables must be correlated with future income in vector (Q) of income equations (4a) and (4b), but must not be correlated with the error term of the high school graduation choice equation (7). We include the total income of the youth's family when he or she is growing up and the percentage of the youth's neighborhood that is high income in (4a) and (4b), reflecting the tie between parent's and neighbors' economic success and their access to labor and transfer income opportunities.[40] We find little theoretical reason to expect them to be related to the high school graduation outcome (other than through changing income), even though they may influence post-secondary schooling choices, conditional on graduating from high school. Other aspects of family well-being and neighborhood circumstances, including the income-to-needs ratio of the family and the percentage of youths in the neighborhood who fail to complete high school, are included as direct determinants of the schooling choice through their inclusion in equation (7).

In addition to these two variables, we include the number of years the child lived in each of three regions of the country in equations (4a) and (4b), but not equation (7). These location variables, intended to capture the variation in regional prices and earnings, give added

identification to the model provided there are not strong and systematic regional variations in the probability of high school graduation.[41]

The instruments that we have chosen as exclusion restrictions are statistically significant in predicting income. Panel A of Table 3 presents age-specific values for a log-likelihood test of joint significance for the five identifying variables in the 14 equations estimated over ages 19–32 for the high school graduates. The identifying variables are statistically significant at the 10 (5) percent level, in 14 (8) of 28 estimates.[42] Although the measure of statistical significance is greater in the equations estimated for the high-school-graduate sample than for the smaller dropout sample, the identifying variables explain a larger portion of the variation in income for the high school dropouts.

Instruments must not be weakly correlated with the endogenous variable, even though they are statistically significant; both economic and statistical significance are important.[43] In OLS estimation of our income equations for high school graduates at each age, the R-squared (an indicator of correlation) increases by an average of 7.5 percent when the variables used for identification are added to the regressions, and by an average of 11.6 percent in the income equations for dropouts; see panel A of Table 3.

The existence of multiple exclusion restrictions enables a test of overidentifying restrictions which can confirm that, conditional on a single instrument being valid, other exclusion restrictions also influence the schooling choice only through their effect on income. Hence they should not be included in the estimation of the high school graduation choice probit [equation (7)]. Such tests involve regressing the residuals from equation (7) on the instruments, and testing the statistical significance of the regression. The test statistic is the uncentered R-

squared times the number of observations, which is distributed chi-squared with degrees of freedom equal to the number of overidentifying restrictions[44] If the instruments are not correlated with the error term of the graduation equation, then a regression of the residuals on the instrument should have a small R-squared statistic. The 5.04 value of this test using our instruments is well below the 7.78 critical value for statistical significance at the 10 percent level and thus the estimation passes the test of overidentification. In addition, we correlated each instrument with the error term, and obtained values that range from .026 to .089. Given that the overidentification tests are passed at standard econometric levels and that the correlation between the error term and the instruments is so low, the coefficient estimate on the income term is unlikely to be capturing the direct effect of the instruments on education rather than the effect of income in education.

Because different judgments regarding the choice of exclusion restrictions are possible, we have examined the robustness of our choice to alternative plausible instruments; the results are in panel B of Table 3. For each specification, a variable that is in the income equation but not the high school graduation equation is now included in both equations. Although exclusion restrictions are necessary for identification of the model, the model is robust to inclusion of any of the excluded variables in the graduation probit.[45] In addition, the tests of overidentification are passed for each of these alternative specifications of identification. Although the precision of the estimated coefficients varies somewhat with the choice of identifying restrictions, the simulated change in the probability of graduating high school remains within the range of .0119 to .0175—or percentage changes of from 1.4 to 2.1—under all of the alternatives.

Estimation of the model with selection requires additional identification. The initial probit equation (estimated to obtain the sample selection lambdas) requires variables which affect the

probability that individuals in the secondary sample will graduate from high school, but which do not have an effect on the expected incomes of these youths, except through the schooling outcome. The first such identifying variable (included in the first-stage probit but not the income equations) is the proportion of years the family moved when the youth was aged between 12 and 15. The second variable is related to family income. A variable measuring the average family income-to-needs ratio when the youth was aged between 12 and 15 is more closely related to the youth's schooling decision, whereas total family income (not adjusting for family size) reflects family labor and transfer market connections and is more proximately related to personal income when the individual is aged 19–29 (see the discussion in Section I.). Both variables are statistically significant in the selection equation; t-statistics are 2.24 for the family move variable, and 3.94 for the income-to-needs ratio. The test of overidentifying restrictions is also passed at conventional significance levels in this model with selection.

V. ROBUSTNESS TESTS

In estimating our preferred model, a variety of assumptions have been made and conventions adopted. Because judgments on these matters can differ, we have attempted to determine if the estimates from our preferred model (columns 1–3, Table 2) are robust to alternative reasonable assumptions. In Table 4, we present results when we systematically change the specification of the base model in a variety of dimensions, including:

- Alternative aggregation of education level
- Alternative process of forming expectations
- Alternative measures of income

- Allowing the effect of income on utility to vary by schooling level
- Alternative discount rates
- Group-specific discount rates
- Alternative time horizons
- Effect of uncertainty of income
- Alternative definition of high school graduation

We present the coefficients and t-statistics on the income variable(s), and a simulation of the expected response to a 10 percent increase in the expected income variable(s). The response simulations allow us to compare the effects of specified income changes using coefficients estimated across the alternative specifications.[46] Although there is some variation in magnitude, the basic results hold across all of the robustness specifications.

A. Alternative Aggregation of Education Level

The returns to graduation from high school estimated in the model include the returns of those who graduate but do not attend college and those who go on to college. We choose this specification, noting that a portion of the return to graduating from high school is the option of attending college, and many graduates do go on to attend college. Limiting the sample to those who do not attend college understates the returns to graduating from high school for those who do attend college and places no value on the option of attending college for those who may or may not. However, an alternative judgment may also have merit, in that those who are on the margin in deciding whether or not to graduate from high school may be least likely to attend college; thus the predicted earnings difference will overstate the expected returns to graduating for these

marginal students. Moreover, those who attend college typically have low earnings while they are in college, and thus the estimated income stream may understate the true returns for those who not only graduate from high school but also go to college.[47]

To address these issues, we present a specification that limits the reference sample to those who do not attend college. The predicted income streams are quite different from the preferred specification; the expected income from graduating is rather flat from ages 19 to 32, though at a higher amount than the expected income from dropping out. However, the net present value of income if the youth graduates is very similar in this alternative specification to that in the preferred specification. The elimination of college attendees increases age 19–22 earnings for graduates by dropping the lower earnings of college attendees during this age and lowers age 23–32 earnings for graduates by dropping the higher earnings of college attendees, so that the two factors essentially offset each other. Given the similarity of expected income differences, the similarity in the coefficient on income is not surprising.

B. Alternative Process of Forming Expectations

There is little agreement as to whether individuals have myopic or rational expectations when forming estimates of the implications of alternative choices (see above). As an alternative to our assumption of myopic expectations—that youths in the primary sample form their expectations by examining the incomes of similar individuals in the reference sample—we assume that individuals have rational expectations, in that their expected incomes do not vary systematically from their own actual incomes. A model which estimates both the income equations and the high school graduation equation fitted over the same sample[48] (rather than an

older cohort reference sample for the income equations) yields the results shown in panel B of Table 4. The coefficient on the income gain term is significant at the 10 percent level, and the marginal effect is identical to that in the preferred model.[49] The similarity of these results suggests that the information that youths obtain by looking to the outcomes of similar individuals in a reference group is not dissimilar from their own rational expectations, and adds to our confidence in the baseline results, which better fit both the data and our own views regarding the process of forming expectations.[50]

C. Alternative Measures of Income

In panel C of Table 4, we use two measures of income as alternative to the personal income measure used in the baseline models–labor market earnings and a family income-to-needs ratio.[51] In the third line of panel C, rather than estimating income at each age and taking the net present value, we calculate the net present value of income for each individual in the reference sample and estimate just two income equations: one for the expected net present value of income for graduates and one for dropouts.[52] The coefficients from these two equations are then used to obtain the expected net present value under each option for the primary sample.

D. Allowing the Effect of Income on Utility to Vary by Schooling Level

In the theoretical model, the utility associated with income is not dependent on the level of education. As an alternative, we relax this assumption, and allow Z to vary by education level, by including expected income with graduation and income without graduation separately in the regression. Panel D of Table 4 indicates that the results are relatively symmetrical for the two

income terms—income with graduation receives the same weight in utility as income without graduation. Although the positive marginal effect of increasing income if the youth is a high school graduate is somewhat larger than the negative marginal effect of decreasing income if the youth does not graduate from high school, the hypothesis that the two coefficient estimates are the same cannot be rejected. The two income variables are jointly significant when entered separately in the model.

E. Alternative Discount Rates

Panel E of Table 4 presents the results from using 1, 5, 10, and 25 percent discount rates as alternatives to the 3 percent rate used in our base specification.

F. Group-Specific Discount Rates

It has been argued that discount rates vary by socioeconomic group, rather than being invariant across individuals as the baseline estimates assume. For example, individuals from low-income families are often viewed as placing a higher weight on time, and Samwick (1997) provides some evidence of this pattern. In panel F of Table 4, we test the sensitivity of our results to this possibility in two ways. We first assign a discount rate of 10 percent to those with average family income less than twice the poverty line during their childhood years, and then to those who lived in a family with income below the poverty line at any time during their childhood years. (In both cases, a 3 percent discount rate is assigned to the rest of the sample.)

G. Alternative Time Horizons

In panel G of Table 4, we test the sensitivity of our estimated results by extending the time horizon beyond the age 32 boundary used in the baseline estimate. We first estimate a model in which we assume that predicted incomes at age 32 will continue to age 59. Next we limit the predicted income stream to age 27 by neglecting the last five years of income information.

H. Effect of Uncertainty in Expected Income

Use of the difference in expected income streams ignores differences in the uncertainty of the two streams.[53] In panel H of Table 4, we test the sensitivity of our baseline estimates to differences in uncertainty in year-to-year income expectations between the high school graduation and dropout options. In this analysis, we measure individual-specific income uncertainty[54] in both the graduation and dropout options over the 14-year age span (from ages 19 to 32), and introduce uncertainty measures for the graduation and dropout options[55] into the baseline model. Neither of the coefficients on the uncertainty terms is statistically significant.[56] Whereas the uncertainty measure in the graduation option enters with a negative sign, consistent with risk aversion, the sign on the uncertainty term for the dropout option is positive.[57] These results suggest that year-to-year income uncertainty does not play an important role in youths' schooling choices.[58]

I. Alternative Definition of High School Graduation

In our baseline model, we define individuals who receive their Graduation Equivalent Degree (GED)[59] to be high school graduates, consistent with the observed pattern of receipt of

GEDs.[60] Although there is still debate about how GED recipients should be classified,[61] the GED recipients in our reference sample more closely resemble graduates than nongraduates.[62]

In sum, these robustness tests reinforce the conclusion suggested by our baseline results, that youths respond to perceived economic gains in making schooling decisions. Our estimates of responses to income differences in schooling choices are largely invariant to alternative estimates based on substantially different judgments regarding the process of forming expectations, measures of outcome and expected economic returns, econometric specification and identification, time horizons, discount rates, and uncertainty.

VI. SIMULATION OF THE EFFECTS OF SELECTED VARIABLES

Structural, as compared to reduced-form, estimation of the determinants of schooling choice implies that exogenous background, family, and neighborhood factors affect youths' choices both indirectly (through their effect on youths' perceived returns to education) and directly. Structural estimates, therefore, enable the researcher to quantitatively identify the process through which changes in exogenous factors affect educational outcomes, and this is an important gain. Here, we use the estimated coefficients in both the income equations and the final-stage probit of our baseline models to estimate the size of these indirect and direct impacts of the exogenous factors on youths' schooling choices.

In Table 5, we present simulation results for a variety of family characteristics (choices) with statistically significant coefficients in the final-stage probit equation of our preferred model. We present both the direct effect of simulated changes in these variables (based on the coefficient estimates in Table 2, columns 1–3), and the indirect effect (computed through measuring the

impact of changes in the variables on expected incomes, and in turn the effect of these income changes on the probability of graduating from high school, using the coefficient estimate on the expected income difference variable).

The first bank of results shows the simulated effect of increasing the expected income from high school graduation by both 10 percent and 25 percent, while holding constant the values of the other variables (including expected income if the youth is a dropout). We simulate that a 10 percent increase in future expected income conditional on being a high school graduate reduces the probability of dropping out by 1.6 percentage points, or by about 10 percent, and a 25 percent increase would decrease the probability by 3.7 percentage points, or by about 24 percent. These are substantial changes.

In the final row of the first bank of panel 1, we show the simulated effect on the probability of dropping out of actual changes in earnings returns to education from 1972 to 1990 (these changes are race- and gender-specific).[63] This exercise suggests that the actual change in the return to high school graduation has reduced the probability of dropping out of high school by nearly 20 percent over this period. According to the U.S. Department of Education, the percentage of high school dropouts among persons 16 to 24 years old fell from 15.0 percent in 1970 to 11.0 percent in 1992, a decrease of 26.7 percent. This suggests that changes in the expected returns to education explain a great deal of the actual change in high school completion rates.

The second panel of Table 5 shows the effect on the dropout probability of simulated changes in background factors significantly related to the educational choice. In the first simulation, we assume that both parents are at least high school graduates.[64] In this case, there is a

direct and an indirect effect; the direct effect is measured by the coefficient on the parental education variable in the final probit, and the indirect effect reflects the impact of parents' schooling on the conditional income expectations. Increases in parental education have a large effect on the probability that the children will drop out of high school. It appears that this is primarily through their direct impact on their children's choices, rather than by increasing the children's expected income opportunities; the direct effect is much larger than the indirect effect.

We also simulate the effect on the high school graduation choice of changes in youths' family structure during ages 6–15 (no time in a single-parent family), parental welfare receipt (no time on welfare), family mobility (no moves), and neighborhood quality (a neighborhood in which 10 percent or fewer young adults were high school dropouts). In all cases the simulated effects are quite small. For example, if we assume that all youths live with both parents while growing up, our model simulates a reduction in the probability of dropping out of 3.7 percent; the direct effect is negative while the indirect effect is positive (though very small). Likewise, the simulated effect of an absence of welfare receipt is also small; here again the negative direct effect dominates the positive indirect effect.

In general, we find that simulated changes in these family and neighborhood variables do affect the schooling choice, but that the direct effect of these changes exceeds the indirect effect operating through changes in the expected returns to schooling (the income variables).

VII. CONCLUSION

We have modeled the high school completion choices of older adolescents as a rational response to expectations regarding the utility (income) gain associated with graduating from high

school relative to dropping out. We obtained estimates of choice-specific incomes through an explicit expectations formation process in which youths refer to the choice-conditioned outcomes of an older cohort in framing their own estimates of the economic impacts of the choices with which they are confronted. We find that youths' expected income returns to graduating from high school are influential in their schooling choice, even when an extensive set of background, economic, family, and neighborhood variables, designed to capture the effects of parental and governmental decisions, are introduced into the analysis.[65] Consistent with economic models, youths appear more likely to choose to graduate from high school as expected returns from additional schooling increase. Our results are consistent with those of prior research that has investigated the determinants of college attendance and the selection of fields of study.

These results suggest that policies designed to increase the income returns to high school graduation—either by increasing the income associated with additional schooling or decreasing the costs of schooling—may reduce the dropout rate. However, such measures may also tend to leave a smaller group of nongraduates relatively (but not absolutely) more disadvantaged. These results also suggest that measures designed to increase the economic well-being or the returns to working of low-skilled workers (through, say, increases in the minimum wage or subsidies to low-wage employment) could have the adverse side effect of increasing the number of youths who would choose to drop out rather than complete high school.

References

Altonji, Joseph, and Thomas Dunn. “The Effects of School and Family Characteristics on the Returns to Education.” National Bureau of Economic Research Working Paper No. 5072, 1995.

Arellano, Manuel, and Costas Meghir. “Female Labour Supply and On-the-Job Search: An Empirical Model Estimated Using Complementary Data Sets.” *Review of Economic Studies,* 59, 1992, 537–57.

Axinn, William, Greg J. Duncan, and Arland Thornton. “The Effects of Parents’ Income, Wealth and Attitudes on Children’s Completed Schooling and Self-Esteem,” in *Consequences of Growing Up Poor*, edited by Greg J. Duncan and Jeanne Brooks-Gunn. New York: Russell Sage Foundation, 1997, 518–40.

Becketti, Sean, William Gould, Lee Lillard, and Finis Welch. 1988. “The PSID after Fourteen Years: An Evaluation.” *Journal of Labor Economics*, 6(4), 1988, 472–92.

Bound, J., D. Jaeger, and R. Baker. “Problems with Instrumental Variable Estimation when the Correlation Between the Instruments and the Endogenous Explanatory Variable is Weak.” *Journal of the American Statistical Association*, 90(430), 1995, 443–450.

Cameron, Stephen V., and James J. Heckman. “The Nonequivalence of High School Equivalents.” *Journal of Labor Economics*, 11(1), 1993, 1–47.

Corcoran, Mary, and Terry Adams. “Race, sex and the intergenerational transmission of poverty,” in *Consequences of Growing Up Poor*, edited by Greg J. Duncan and Jeanne Brooks-Gunn. New York: Russell Sage Foundation, 1997, 461–517.

Datcher, L.. “Effects of community and family backgrounds on achievement.” *The Review of Economics and Statistics*, 64, 1982, 32–41.

Das, M., and B. Donkers. “How Certain Are Dutch Households about Future Income? An Empirical Analysis.” *Review of Income and Wealth*, 45, 1999, 325–38.

Dominitz, Jeff, and Charles F. Manski. “Eliciting Student Expectations of the Returns to Schooling.” *Journal of Human Resources*, 31(1), 1996, 1–26.

Duncan, Greg J. “Families and Neighbors as Sources of Disadvantage in the Schooling Decisions of White and Black Adolescents.” *American Journal of Education*, 103(1), 1994, 20–53.

Duncan, Greg J., and Saul Hoffman. “Welfare Benefits, Economic Opportunities, and Out-of-Wedlock Births among Black Teenage Girls.” *Demography*, 27(4), 1990, 519–35.

Evans, Williams, Wallace Oates, and Robert Schwab. “Measuring Peer Group Effects: A Study of Teenage Behavior.” *Journal of Political Economy*, 100(5), 1992, 966–91.

Farber, Henry S. “Job Creation in the United States: Good Jobs or Bad?” Industrial Relations Section Working Paper No. 385, Princeton University, 1997.

Fitzgerald, John, Peter Gottschalk, and Robert Moffitt. “An Analysis of Sample Attrition in Panel Data: The Michigan Panel Study of Income Dynamics.” Mimeo, December 1996.

Freeman, Richard. *The Market for College-Trained Manpower*. Cambridge, MA: Harvard University Press, 1971.

————. “Supply and Salary Adjustments to the Changing Science Manpower Market: Physics, 1948–1973.” *American Economic Review*, 65(1), 1975, 27–39.

————. “Demand for Education,” in *Handbook of Labor Economics, Vol. 1*, edited by O. Ashenfelter and R. Layard. Amsterdam: North-Holland Press, 1986.

Goldin, Claudia. “America’s Graduation from High School: The Evolution and Spread of Secondary Schooling in the Twentieth Century.” *Journal of Economic History*, 58(2), 1998, 345–74.

Greene, William H. *Econometric Analysis*, Third Edition. New Jersey: Prentice Hall, 1997.

Hanushek, Eric. “The Economics of Schooling: Production and Efficiency in Public Schools.” *Journal of Economic Literature*, 24, 1986, 1141–77.

————. “Remedial Math: Rather than Spend More on Schools, We Should Spend More Wisely.” *The New Democrat*, 7(6), 1995, 25–27.

————. “Conclusions and Controversies about the Effectiveness of School Resources.” *Federal Reserve Board of New York Economic Policy Review* 4(1, March), 1998, 11–28.

Hanushek, Eric, with Charles S. Benson, Richard B. Freeman, Dean T. Jamison, Henry M. Levin, Rebecca A. Maynard, Richard J. Murnane, Steven G. Rivkin, Richard H. Sabot, Lewis C. Solomon, Anita A. Summers, Finis Welch, and Barbara L. Wolfe. *Making Schools Work: Improving Performance and Controlling Costs*. Washington, DC: Brookings Institution, 1994.

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

————. “The Determinants of Children’s Attainments: A Review of Methods and Findings.” *Journal of Economic Literature*, 33(4), 1995, 1829–78.

Haveman, Robert, Gary Sandefur, Barbara Wolfe, and Andrea Voyer. “Trends in Children’s Attainments and Their Determinants as Family Income Inequality Has Increased,” in

Social Inequality, edited by Kathryn Neckerman. New York: Russell Sage Foundation, 2004.

Heckman, James A. "Sample Selection Bias as a Specification Error." *Econometrica* 47(1), 1979, 153–61.

Heckman, James, Jingjing Hsee, and Yona Rubinstein. "The GED Is a Mixed Signal." Mimeo, University of Chicago, 1999.

Heckman, James, Anne Layne-Farrar, and Petra Todd. "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*, edited by Gary Burtless. Washington, DC: Brookings Institution, 1996.

Hill, Martha "The Panel Study of Income Dynamics: A User's Guide." *Guides to Major Social Science Data Bases*, Vol. 2. Newbury Park, CA: Sage Publications, Inc., 1992.

Jacob, Brian. "Where the Boys Aren't: Non-Cognitive Skills, Returns to School and the Gender Gap in Higher Education." National Bureau of Economic Research Working Paper No. W8964, 2002.

Johnston, Jack, and John DiNardo. *Econometric Methods*, 4th Edition. New York: McGraw-Hill, 1997.

Lee, Lung-Fei. "Identification Estimation in Binary Choice Models with Limited (Censored) Dependent Variables." *Econometrica*, 47(4), 1979, 977–86.

Lillard, Lee, and C. W. A. Panis. "Panel Attrition from the PSID: Household Income, Marital Status, and Mortality." Mimeo, Rand Corporation, 1994.

Manski, Charles. “Academic Ability, Earnings and the Decision to Become a Teacher: Evidence from the National Longitudinal Study of the High School Class of 1972,” in *Public Sector Payrolls*, edited by David Wise. Chicago: University of Chicago Press, 1987.

————————. “Adolescent Econometricians: How Do Youths Infer the Returns to Schooling?” in *Studies of Supply and Demand in Higher Education*, edited by Charles T. Clotfelter and Michael Rothschild. Chicago: University of Chicago Press, 1993a.

————————. “Identification of Endogenous Social Effects: The Reflection Problem.” *Review of Economic Studies*, 60(3), 1993b, 531–42.

Manski, Charles, and David Wise. *College Choice in America*. Cambridge, MA: Harvard University Press, 1983.

Manski, Charles F., Gary Sandefur, Sara McLanahan, and Daniel Powers. “Alternative Estimates of the Effects of Family Structure During Adolescence on High School Graduation.” *Journal of the American Statistical Association*, 87, 1992, 25–37.

Manski, Charles F., and John D. Straub. “Worker Perceptions of Job Insecurity in the Mid-1990s: Evidence from the Survey of Income Expectations.” *Journal of Human Resources*, 35(3), 2000, 447–79.

Mattila, J. P. “Determinants of Male School Enrollments: A Time-Series Analysis.” *Review of Economics and Statistics*, 64(2), 1982, 242–51.

Murnane, Richard J., John B. Willett, and Katherine P. Boudett. “Do High School Dropouts Benefit from Obtaining a GED?” *Educational Evaluation and Policy Analysis*, 17(2), 1995, 133–47.

Murnane, Richard J., John B. Willett, and John H. Tyler. “Who Benefits from Obtaining a GED? Evidence from High School and Beyond.” National Bureau of Economic Research Working Paper No. W7172, 1999.

Peters, H. Elizabeth, and Natalie C. Mullis. “The Role of Family Income and Sources of Income in Adolescent Achievement,” in *Consequences of Growing Up Poor*, edited by Greg J. Duncan and Jeanne Brooks-Gunn. New York: Russell Sage Foundation, 1997, 340–81.

Ribar, David. “A Multinomial Logit Analysis of Teenage Fertility and High School Completion.” *Economics of Education Review*, 12(2), 1993, 153–64.

Rivkin, Steven G., Eric A. Hanushek, and John F. Kain. “Teachers, Schools, and Academic Achievement.” National Bureau of Economic Research Working Paper No. 6691, 2002.

Samwick, Andrew A. “Discount Rate Heterogeneity and Social Security Reform.” National Bureau of Economic Research Working Paper No. W6219, 1997.

Schweitzer, Mark E. “Workforce Composition and Earnings Inequality.” *Economic Review–Cleveland Federal Reserve Bank* 33(2), 1997, 13–24.

Teachman, Jay D., Kathleen M. Paasch, Randall D. Day, and Karen P. Carver. “Poverty during Adolescence and Subsequent Educational Attainment,” in *Consequences of Growing Up Poor*, edited by Greg J. Duncan and Jeanne Brooks-Gunn. New York: Russell Sage Foundation, 1997.

U.S. Bureau of the Census. *Statistical Abstract of the United States*. http://www.census.gov/prod/2004pubs/03statab/intlstat.pdf, 2003.

U.S. Department of Education. *Digest of Education Statistics, 2002*. http://nces.ed.gov/fastfacts/display.asp?id=16, 2003.

Willis, Robert, and Sherwin Rosen. “Education and Self-Selection.” *Journal of Political Economy* 87(5), 1979, S7–S35.

Wilson, Kathryn. “Using the PSID to Study the Effects of School Spending.” *Public Finance Review* 28(5), 2000, 428–51.

———————. “The Determinants of Educational Attainment: Modeling and Estimating the Human Capital Model and Education Production Functions.” *Southern Economic Journal* 67(3), 2001, 518–51.

Notes

[*]Kathryn Wilson is Associate Professor of Economics at Kent State University. Barbara Wolfe is Professor of Economics, Population Health Sciences, and Public Affairs at the University of Wisconsin–Madison. Robert Haveman is John Bascom Emeritus Professor of Economics and Public Affairs at the University of Wisconsin–Madison.

[1]In 2000, the U.S. proportion of the working-age population with less than a high school diploma was 13 percent, compared to Germany at 18 percent, France at 36 percent, Australia at 41 percent, and Canada and the United Kingdom at 18 and 17 percent, respectively. (U.S. Bureau of the Census, 2003, Table 1331.)

[2]The latest data, for March 2001, show the high school dropout rate for non-Hispanic whites and blacks aged 16–24 at 7.3 and 10.9 percent, respectively. Among Hispanics, 27 percent of those aged 16–24 years have dropped out. (U.S. Department of Education, 2003, Table 108.) Persons with a General Education Development (GED) are counted as completers.

[3]Haveman, Sandefur, Wolfe and Voyer (2004) review several of these contributions. The most recent include Axinn, Duncan, and Thornton (1997), Peters and Mullis (1997), and Teachman, Paasch, Day, and Carver (1997).

[4]Haveman and Wolfe (1995) note that: "[T]here are but few studies that attempt to account systematically for the interdependence among determinant variables so necessary for establishing true causal links. . . . Moreover, the studies . . . are primarily reduced-form estimates, with little attempt to characterize the choices made as responses to economic incentives; there are few structural models to be found." (pp. 1872–73).

[5]Willis and Rosen (1979) were among the earliest to have explicitly introduced choice-

conditioned expectations into the analysis of schooling choices. Another early attempt is a study by Duncan and Hoffman (1990), in which expected incomes (including welfare benefits) conditional on whether or not there is a nonmarital birth are introduced as choice determinants in a model of African-American teenage female fertility. Manski (1987) and Manski and Wise (1983) have presented structural models of the choices of young people regarding whether or not to become a teacher and whether or not to enroll in postsecondary school programs. None of these studies secure full choice-conditioned expectation estimates based on an explicit expectations formation process. Several recent surveys ask explicitly expectational questions in order to determine the correspondence of survey response probabilities to actual outcomes. See for example, Survey of Economic Expectations as described in Manski and Straub (2000), the USB-Panel Survey of the Netherlands (Das and Donkers, 1999), and a survey of youths in Madison, WI. (Dominitz and Manski, 1996).

[6]The decision of whether or not to complete high school is also less complicated by issues of choice-conditioned tuition costs or available student financial aid than are subsequent schooling decisions. High school attendance does carry an opportunity cost in terms of forgone earnings and there may be nonpecuniary costs such as exposure to violence which may vary by schools. Our model incorporates both these costs.

[7]For example, Farber (1997) documents the significantly lower hourly wages of high school dropouts relative to graduates from 1979–96, for both existing and new jobs.

[8]The role of school-based resources in influencing educational attainment is controversial. Hanushek (1998 and 1986) and Hanushek et al. (1994) conclude that the evidence on the independent effects of school characteristics is not substantial. More recent studies,

including Hanushek (1995), and Rivkin, Hanushek and Kain (2002) suggest a greater role for school inputs, especially teacher quality. Wilson (2000, 2001) explores the effects of school quality using school data appended to the Michigan Panel Study of Income Dynamics.

[9]Because schooling (S) is measured dichotomously, those elements of the utility function with the subscript *s* take on a different value depending on the graduation choice.

[10]Allowing **$** to vary with schooling attainment status allows background, family, and community to affect the returns to schooling.

[11]Equation (6) posits that the schooling choice is determined only by point estimates of the expected present value of incomes. Clearly, other moments of the expected subjective income distributions are also relevant, especially the variance. In Section V we present the results of a variant of our model that incorporates uncertainty in year-to-year income expectations between the high school graduation and dropout options.

[12]This is the same expectations framework as that used by Freeman (1971) and Manski (1993a, 1993b). Manski (1993a) relates this method to how econometricians forecast. “I instead assume that youth form their expectations in the manner of practicing econometricians: youths observe the incomes realized by members of the preceding generation who chose schooling, and they make inferences from these observations.” (p. 49)

[13]The PSID data provide longitudinal information on 5,000 families beginning in 1968. We use data covering 26 years of information, beginning with 1968.

[14]Only those children who remained in the survey until 1988 are included. Haveman and Wolfe (1994) studied the effect of attrition on this sample, and concluded that, with the exception of race, those who attrited do not appear to differ from the remaining sample. Previous studies of attrition in the PSID also find little reason for concern that attrition has reduced the

representativeness of the sample. See Becketti, Gould, Lillard, and Welch (1988), Lillard and Panis (1994), and Fitzgerald, Gottschalk, and Moffitt (1996). In a few cases, observations could not be used and are excluded from the analysis. These include persons with two or more contiguous years of missing data. Those observations with but one (contiguous) year of missing data were retained and the missing data were filled in by averaging the data for the two years contiguous to the year of missing data. For the first and last years of the sample, this averaging of the contiguous years is not possible. In this case, the contiguous year's value is assigned, adjusted if appropriate using other information that is reported.

[15]The PSID survey respondent may inaccurately report the education level of other individuals in the household. If the sample member is the "head" or "wife" in the household at any time during the sample period, we use the education level reported at that time. This measure is likely to be self-reported and accurate. For sample members who are not the "head" or "wife" we take the highest reported level of education. In Section V, we check the sensitivity of our estimates to this assumption by including only diploma graduates in the estimation.

[16]The U. S. Census Bureau describes the CPI as the best measure for adjusting payments to consumers when the intent is to allow them to purchase, at today's prices, the same market basket of consumer goods and services that they could purchase in an earlier reference period. "It is also the best measure to use to translate hourly and weekly earnings into inflation free dollars." See http://stats.bls.gov/cpifaq.htm, Question 1.

[17]Combined geographic codes added to the annual PSID data for 1968 to 1985 by the Michigan Survey Research Center were matched to 1970 and 1980 Census data. Using the Census data, we assigned neighborhood values to the neighborhood in which each family in the

PSID lived. In most cases, this link is based on a match of the location of our observations to the relevant Census tract or block numbering area (67.8 percent for 1970 and 71.5 percent for 1980). For years prior to 1970 we use 1970 data; for years after 1980 we use 1980 data; and for 1971–1979 we used a weighted combination of 1970 and 1980 data [weights are .9 (1970) and .1 (1980) for 1971; .8 (1970) and .2 (1980) for 1972 and so on].

[18]Personal income is defined as the sum of the person's own earnings, transfer benefits, and unearned income from all other sources. Omitting transfer income (including cash welfare benefits) would have left out an important component of the expected economic well-being concept specified in our model. We use personal rather than family income since the latter incorporates issues of family composition and allocation which are outside our model and, for the most part, our observation. In Section V, we test the robustness of our results to alternative income measures.

[19]The variables included in these equations are those judged likely to be related to the personal income dependent variable. These include personal characteristics (dummy variables for race, gender, race*gender, and firstborn), family characteristics (mother's education, father's education, parental education missing, log of family income, years in poverty, years the family received AFDC, number of siblings, years in a single-parent family, years mother worked, years head was disabled, and years in a Standard Metropolitan Statistical Area [SMSA]), neighborhood characteristics (percentages of youths that drop out of high school and families with high income), region dummy variables (West, Northeast, and South), and gender interaction variables (female*years mother worked, female*years in single-parent family, and female*years family received AFDC). (Inclusion of the three interaction variables was guided by a test of the

hypothesis that the intercept is the only difference between males and females in the income equations, using an alternative specification. In this specification, the coefficients on variables in which gender was interacted with each independent variable were not jointly significant. However, years mother worked, years in a single-parent family, and years receiving AFDC were each individually significant when interacted with gender.) The means and standard deviations of the variables are shown in the first four columns of Appendix Table 1. The 12–15 age range is used for variables describing circumstances during adolescence. This procedure differs from that of Willis and Rosen (1979), who calculate a growth rate based on income observed at two points in time only; our data allow us to observe income every year for 14 years and hence, we use the multiyear income stream.

[20]In the income regressions, the African American and female variables have negative and significant coefficients, while African American*female has positive and significant coefficients. The magnitude of these coefficients is large. Simulating the sample as African American reduces predicted income by 26 percent for high school graduates; and by 37 percent for nongraduates. Similarly, simulating the sample as female reduces predicted income by 23 percent for the sample of graduates and by 46 percent for nongraduates. Having a parent with higher education has a negative and significant effect on income for younger ages, but the coefficient becomes positive and significant in estimates for older ages. For the sample that did not graduate, having a father who graduated from high school has a negative coefficient for all ages. The coefficient on the family income/needs ratio is generally insignificant for nongraduates; for graduates, the income/needs ratio has a similar pattern to that described above for parental higher education. Other variables that have a statistically significant relationship

with income are percent of years spent in an SMSA, which has a positive sign, and the region dummy variables. Percentage of years living in the West or South (relative to the omitted dummy variable for North Central) is positively associated with income for high school graduates, whereas nongraduates who lived in the South have lower income.

[21]We have excluded predicted incomes for youths aged 17 and 18 because it is difficult to accurately identify earnings levels at these ages, when earnings account for the bulk of personal income.

[22]Implicit in our use of this expected gain indicator as a proxy for the difference in youths' expected outcomes in the two states is the assumption that youths presume there is no difference in the income generation process between the younger and the older cohort (i.e., there is no cohort effect). The characteristics of the youths who do and who do not graduate high school overlap substantially. Our reduced-form model predicting this choice fails to explain a high proportion of the choices made, suggesting only limited self-selection in terms of the economic opportunities facing youths in their schooling choices; adolescents with low forgone income (associated with failing to graduate) and with high forgone income are observed to both graduate and drop out before graduating. This avoids a potential identification problem in the use of these income expectation variables to explain the choices observed.

[23]The income trajectory associated with the high school graduation option is influenced by the inclusion in the graduate category of some youths who will continue schooling beyond high school graduation. The projected income stream will be depressed over ages 19–22 because those attending college and not working are included. Conversely, the inclusion of these college-bound youths in the graduate category leads to a higher projected income stream after they have

completed their schooling. These two effects tend to offset each other in the calculation of the present value of the trajectories (see below). In section V. we present the results of estimating our model when the graduate category is designed to include only those youths who graduate high school but do not continue on to college.

[24]In discussing their results, Willis and Rosen explain, "plumbers may have very limited potential as highly schooled lawyers, but by the same token lawyers may have much lower potential as plumbers than those who actually end up choosing that kind of work. This contrasts with the one-factor ability-as-IQ specifications in the literature which assume that the best lawyers would also be the best plumbers and would imply strictly hierarchical sorting in the absence of financial constraints." (p. S11)

[25]The youths in our primary sample may have passed the age at which the high school choice is made by the time members of the reference group are in their older twenties. We are implicitly assuming that the estimated life-cycle earnings trajectory for these years is correctly perceived by the individuals whose decisions we model. Data limitations preclude extending the observation period beyond age 32. See Heckman, Layne-Farrar, and Todd (1996) on the use of earnings estimated during the 20s as proxies for the lifetime income experience of individuals pursuing careers with long training requirements.

[26]Unweighted data are used for estimation; 1,544, or 79.5 percent, of the youths in our sample graduated from high school. The unweighted means and standard deviations of these variables are shown in Appendix Table 2. Since the regression controls for race and family income, the variables that were the basis of oversampling in the original sampling scheme, the unweighted regression will give consistent coefficient estimates. (See Hill, 1992)

[27]Since the difference in income is an imputed value based on the coefficient estimates of

the 28 income regressions for the reference sample, the standard errors from the high school graduation regression may be biased downward. The residual in the high school graduation equation includes both the regular error term and the coefficient estimate on the income term times the difference between income and the fitted value of income. To examine the magnitude of the bias, we estimated the high school graduation equation using ordinary least squares and corrected the standard errors according to the method outlined in Appendix B4 of Arellano and Meghir (1992). The corrected standard errors incorporate all 28 of the covariance matrices from the 28 income regressions. The corrected standard errors were only changed at the third decimal place or higher (on average, the corrected standard errors are about 5 percent larger than their uncorrected counterparts; 5.5 percent larger for the predicted income term), and there was no change in the level of statistical significance for any variable. Therefore, we conclude that the degree of bias is very small, and present the regular probit standard errors for all regressions.

[28]Other studies also find similar relationships for related questions. Freeman (1975) used time series data for the United States to examine the relationship between market conditions and the supply of new entrants to physics. Mattila (1982) also used U.S. time series data for 1956 to 1979 in studying the response of school enrollment rates to the expected rate of return. These types of studies are summarized in Freeman (1986), who finds that the estimated elasticity of supply of more highly educated persons to changes in salaries ranges from 1 to 2, consistent with our finding of an elasticity of about 1 in the probability of dropping out to income if one graduates (Table 5).

[29]The female African American results are consistent with those found in Manski et. al (1992) and Jacob (2002). Our results are not sensitive if race is defined instead as “white” and

"nonwhite."

[30]The negative and significant coefficient which suggests that having a father with at least some college education reduces the probability of graduation is unexpected. In a reduced-form version (without expected education) the coefficient of the father's college education is positive but not statistically significant. In the income equations, the father's college education is negative for nongraduates and positive for graduates after age 23, suggesting a generally positive role via expected income differences. In the secondary sample, only 2 percent of nongraduates but 15 percent of graduates had a father with at least some college.

[31]We also estimated specifications that included an interaction of expected income with race, income, and poverty. The interaction terms were entered individually, and none were significant, suggesting that the effects of expected income do not vary across these measures.

[32]We are treating these variables as exogenous in this specification. If the family's decisions regarding moving or neighborhood quality or parental work depend on, say, unobserved family characteristics such as motivation, and these factors influence the child's school attainment, coefficient estimates and simulated effects based on them may be biased. See, for example, Evans, Oates, and Schwab (1992).

[33]Use of selection correction techniques is necessary to reliably estimate the determinants of the extent of an activity (e.g., labor supply) chosen by individuals only some of whom are observed to engage in the activity. See Heckman (1979) and Lee (1979). Here, however, the issue is whether members of a younger cohort of youth, in forming their perceptions of choice-conditioned incomes, model the complete selection process by which members of the reference group become graduates or dropouts.

[34]This specification assumes that unobserved variables included in $,_i$ and $>_i$ affect both

education choices and personal incomes. Tastes for a professional career (ambition) or aversion to a status that carries social disapproval (stigma), such as being a high school dropout, are examples. If this is the case, an endogeneity problem exists in obtaining reliable income predictions, because incomes observed for youths who did and did not graduate from high school depend on factors beyond those observed and measured. In effect, parameter estimates (and hence predicted values) from income equations fitted over the two groups will be biased because the selection process that assigns observations to the groups reflects the unobserved variables included in the error terms. The Heckman technique accounts for this process and yields unbiased parameter estimates in the income equations.

[35]The independent variables included in this selection equation are those expected to affect the graduation decision [equation (7)]. They include all variables in the full model of Table 2 (except, of course, the predicted income variable), typically measured over ages 12–15. The definitions, means, and standard deviations of these variables are shown in the last two columns of Appendix Table 1.

[36]We estimate the income equations as OLS rather than Tobit. The results of the initial probit estimate and the choice-specific income models are generally as expected and are available from the authors. We also attempted alternative specifications for this selection model. We attempted to estimate a Tobit model with selection using maximum likelihood (see Greene, 1997), which is our preferred specification; however, the model would not converge for every age and the estimated variance matrix was singular for some ages. We also estimated the model using Tobit for income and including a Heckman selection term, recognizing that this could lead to inconsistent results, and obtained estimates for the final choice equation that are virtually

identical to the results using OLS with selection. These estimates are available upon request from the authors.

[37]In using the coefficients to predict income at each age, conditional on the individuals in our primary sample choosing to graduate or not, we use a fitted value of the lambda term. This is appropriate provided youths know things about themselves that are unobservable to the analyst, and that these unobservable characteristics are captured in the lambda term. In the sensitivity analysis in Appendix Table 4, we also present results when no adjustment for lambda is made in the predicted income term. The predicted income stream in this model with selectivity is very similar to that shown in Table 1 for graduation, whereas the expected income stream without graduation is somewhat higher than the estimates from the preferred model without sample selection. The present value estimates are $175,900 and $107,360, respectively.

[38]If the determinants of income (Q) are the same as the nonincome determinants of education (X), there will be multicollinearity between X and **$**^Q in equation (6), making it impossible to separate the effect of economic incentives in the education decision from the nonincome effects and estimate equation (7). In order to recover the coefficient estimate of the effects of the returns to schooling, the predicted income term cannot be a linear function of the other independent variables included in the high school graduation estimation.

[39]Identification is also secured by timing. The period over which family variables are measured differs between the analyses of the income equations (where variables are averaged over ages 12–15) and the high school graduation equation (where they are averaged over ages 6–15). This difference contributes to model identification because the income terms are not a linear combination of the same variables that are in the graduation equation. Our approach

requires fewer restrictive assumptions than that of Willis and Rosen, who gain identification by assuming that family background has no effect on future income, but only education, and that ability affects future income but not schooling choices.

[40]This choice is consistent with findings that parental income affects both future income (see Datcher 1982 and Corcoran and Adams 1997) and the returns to schooling (see Altonji and Dunn 1995); it assumes that these variables reflect family and peer group linkages to the labor market and "transfer market," and hence influence the schooling choice through their effect on income expectations.

[41]Some regional patterns in educational attainment patterns do exist, but they are not sufficient to preclude use of regions as an identifying variable, given the other family and neighborhood controls included in the model. However, see Goldin (1998).

[42]The critical values for this test are 9.2 and 11.1 at the 10 and 5 percent levels of significance, respectively.

[43]Bound, Jaeger, and Baker (1995) show how biased estimates can result if the correlation between the instruments and the endogenous variable is weak, even thought the estimated relationship is statistically significant (because of, say, large sample size).

[44]See Johnston and DiNardo (1997), pp. 336–38. This is asymptotically equivalent to a Basmann test.

[45]As would be expected, when the high school graduation equation is estimated with all the identifying variables included, the income term is not statistically significant.

[46]Appendix Table 4 contains a parallel table for the model with sample selection. That table includes an additional robustness test in which no adjustment is made for the lambda coefficient (lambda is given a value of zero for each individual in the primary sample), rather

than using the fitted value of lambda in the predicted income terms. This is consistent with the assumption that individuals in the primary sample recognize that the education of those in the reference sample is a result of their selection process but they do not consider their own unobservable characteristics in forming their predictions.

[47]The robustness tests for extended time horizons presented in panel G of Table 4 also alleviate this problem.

[48]We cannot use the primary sample to estimate the model with this expectation process because we do not observe the actual income of these individuals at later ages. However, the procedure in estimating this model is identical to that in estimating the baseline model. We first estimate 28 income equations for the reference sample and use the coefficient estimates to form predicted incomes for the reference sample, and then estimate the high school graduation probit. Background variables are measured over ages 12–15 rather than 6–15 in the high school graduation equation because we observe these variables beginning at age 12 for those in the reference sample.

[49]The t-statistic in the rational expectations model with selection (Appendix Table 4) is (2.57).

[50]There are two cautions to interpreting the results of this sensitivity analysis. First, the tests of overidentifying restrictions are not passed at standard significance levels for this specification. This implies that the exclusion restrictions used to identify this model may include variables that should be included in the high school graduation equation for the reference sample. Because the reference sample is used in this model rather than the younger sample, it is possible that the valid identifying restrictions used in our preferred estimate have a different relationship to the education outcome in this model because changes may have taken place over

time, resulting in the failure of the overidentification tests to be passed in this case. Alternatively, the failure of the tests to be passed may be due to the inclusion of the same variables (other than the exclusion restrictions) in both the income and education models. (The variables are measured over ages 12–15 in both the income and the education equations in the rational expectations model, whereas in the preferred specification they are measured over ages 12–15 in the income equations but over ages 6–15 in the education equation.) Second, the correction of standard errors is more involved for this model because the predicted income term is estimated over the same sample as the high school graduation equation. The t-statistic presented is based on the uncorrected standard error and thus overstates the significance level.

[51]The income-to-needs ratio equals the income of the family unit in which the youth lives divided by the official U.S. poverty line appropriate to the family. We are reluctant to use the family income-to-needs ratio since it would reflect parental income if youths live at their parents' home, hence overstating their access to income. Alternatively, the personal income variable may understate income if youths cohabit and share income.

[52]The income equation is estimated by OLS rather than Tobit because there is no mass point at zero for the net present value of the income variable.

[53]Our model predicts that expected income returns influence the schooling choices of youths; risk aversion is accounted for only in the utility function via the declining marginal utility of income.

[54]Uncertainty is measured as the absolute percentage difference in income compared to the trend of income:

$$\text{Uncertainty}_i = 1/T * \sum_{t=1 \text{ to } T} |Y_{it} - \bar{Y}_{it}| \div Y_i$$

where

$Y\text{-}_{it}$ = trend income at each age = Average $Y_i + [(Y_{iT} - Y_{i1})/(T - 1)] * (t - T/2)$;

i = individual; t = year (where year 1 = age 19, . . ., year 14 = age 32);

T = total years observed (14);

Y_{it} = income of individual i at time t; and

$Y_i = 3_{t=1 \text{ to } T} Y_{it}/Z$ (average income of individual i at ages 19–32).

In essence, we calculate an annual income trend value (the predicted value of annual income from a linear trend fitted to the expected income terms), and take the average over ages 19–32 of the absolute value of the percentage by which expected income deviates from this trend.

[55]Each individual in the primary sample has two calculated uncertainty measures, reflecting the variation in expected income if the individual graduates or does not graduate. On average, for the primary sample the estimated uncertainty if the individual graduates is less than the estimated uncertainty if the individual drops out.

[56]T-statistic = 0.54 for uncertainty of income if a graduate, and 1.21 for uncertainty of income if not a graduate.

[57]We also investigated the difference in uncertainty between the graduation and dropout options by calculating our uncertainty measure over predicted incomes for ages 19–32 for both high school graduates and dropouts in the reference sample. (Measured uncertainty was .83 for graduates and .71 for dropouts.) We then regressed these uncertainty measures on the variables included in the income regressions plus a dummy variable for high school graduation. The coefficient on the graduation variable is not statistically significant, indicating that controlling for the determinants of income, income uncertainty does not differ for graduates and dropouts.

[58]However, these findings do not speak to another potential uncertainty associated with the schooling choice, i.e., uncertainty as to where in the income- level path individuals will

initially be located. We have assembled some evidence regarding this difficult issue. First, we estimate that the standard deviation of the net present value of the actual income (not controlling for other observable factors) of graduates in the reference sample is 134 percent of the income of dropouts, suggesting that graduates face the greater level of income uncertainty. Second, we calculated the square of the residuals from the income regressions for the reference sample; this value indicates the extent of income variation that is not explained by observable factors. This measure of residual variation in income for high school graduates is 112 percent of that of dropouts, suggesting greater income uncertainty in the income level for high school graduates. If youths are risk averse, the greater income uncertainty in the high school graduate option implies that the utility returns to graduating from high school would be biased upward in estimates that ignore this uncertainty. It follows that the coefficient on the expected income difference variable in our baseline estimates may have some downward bias, though we are unable to estimate its extent.

[59]The General Educational Development certificate (GED) is the school leaving credential most commonly available to school dropouts in the United States. Awarding of the GED typically occurs after the normal age for high school completion, and requires the taking and passing of high school equivalent tests. More specifically, this requires having state-specific passing scores on a seven and one-half hour, five-part written battery of tests assessing skills and knowledge in five areas: writing, social studies, science, reading, and mathematics.

[60]Nearly one-half of those who obtain a GED do so within one year of the time their cohort graduated from high school. See Murnane et al. (1999).

[61]See Cameron and Heckman (1993), Murnane et al. (1995), Murnane et al. (1999), and

Heckman et al. (1999).

[62]In the reference sample 54 individuals received a GED, and in the primary sample 74 individuals did so. The average net present value of the personal income stream (ages 19–32) of GED recipients in the reference sample is only 2 percent lower than that of those who graduated with a standard certificate. This pattern of similarity is consistent with younger recipients of GEDs in Heckman et al. (1999).

[63]We obtain race- and gender-specific earnings by education level in 1972 and 1990 from Schweitzer (1997). White men and white women had the greatest increase in the earnings return to a high school diploma; the increases for nonwhite men and women were small. Schweitzer (1997) understates the returns to obtaining a diploma because those with schooling beyond high school are not included in the comparison. Rather, the earnings of high school graduates (but not those attending college) are compared with the earnings of those who did not graduate from high school. In addition, the differentials only reflect earnings differences, whereas the model estimates reflect differences in all sources of personal income.

[64]In this simulation, we increase the schooling level of all parents who have less than a high school diploma to the high school diploma level; the schooling level of parents with a high school diploma or more is not altered.

[65]We caution that although our estimates include a rich specification of the parental choice/opportunity factors influencing children's decisions, we are unable to measure other factors that may significantly influence these decisions. The quality and the socioeconomic/racial composition of the children's schools, the attitudes, expectations, and child-rearing skills of the parents of the children, and the inherent mental powers and physical characteristics of the children themselves are also likely to play some role. These considerations (plus, for example,

other neighborhood characteristics) remain as unobserved variables, given the information in our data.

TABLE 1
Predicted Personal Incomes, if High School Graduate and if Nongraduate

	If Graduate		If Nongraduate	
	Mean	St. Dev.	Mean	St. Dev.
Entire Sample (n=1,942)				
Age 19	$9,112	$2,327	$7,902	$4,905
Age 20	10,695	2,793	9,069	5,639
Age 21	12,175	3,624	11,038	7,381
Age 22	14,303	3,599	11,440	6,944
Age 23	16,045	4,279	9,857	6,961
Age 24	16,710	4,384	9,725	5,525
Age 25	17,932	5,316	8,750	5,337
Age 26	18,553	6,114	8,021	6,734
Age 27	19,517	6,438	10,491	7,411
Age 28	20,235	7,442	9,582	8,223
Age 29	20,236	7,830	8,719	6,610
Age 30	20,793	8,417	7,966	6,054
Age 31	21,535	9,146	6,843	5,066
Age 32	22,149	9,498	7,317	5,508
Net Present Value	$177,550	$51,274	$97,212	$61,118
Actual High School Graduates (n=1,544)				
Age 19	$9,177	$2,284	$7,947	$5,001
Age 20	10,743	2,792	9,110	5,766
Age 21	12,241	3,630	11,132	7,565
Age 22	14,451	3,580	11,565	7,090
Age 23	16,295	4,275	9,944	7,114
Age 24	17,086	4,335	9,798	5,561
Age 25	18,370	5,282	8,739	5,332
Age 26	19,111	6,093	7,835	6,754
Age 27	20,079	6,418	10,467	7,445
Age 28	20,932	7,439	9,478	8,323
Age 29	20,968	7,838	8,673	6,684
Age 30	21,585	8,453	7,916	6,133
Age 31	22,404	9,212	6,686	5,072
Age 32	23,048	9,616	7,270	5,517
Net Present Value	$182,120	$50,966	$97,175	$61,940
Actual High School Nongraduates (n=398)				
Age 19	$8,738	$2,534	$7642	$4,303
Age 20	10,413	2,791	8,831	4,844
Age 21	11,793	3,567	10,492	6,194
Age 22	13,448	3,596	10,716	5,990
Age 23	14,599	4,007	9,353	5,989
Age 24	14,541	4,026	9,302	5,303
Age 25	15,405	4,781	8,811	5,371
Age 26	15,330	5,174	9,091	6,523
Age 27	16,272	5,535	10,625	7,220
Age 28	16,211	6,062	10,179	7,603
Age 29	15,999	6,303	8,984	6,169
Age 30	16,216	6,548	8,254	5,581
Age 31	16,512	6,867	7,751	4,941
Age 32	16,955	6,764	7,587	5,456
Net Present Value	$151,150	$44,707	$97,423	$56,220

TABLE 2
Determinants of High School Graduation
(N = 1,942)

Variable	Main Specification (2-stage) Coefficient Estimate (1)	T-Statistic (2)	Marginal Effect (3)	Model with Selection Coefficient Estimate (4)	T-Statistic (5)	Marginal Effect (6)
Constant	1.099***	5.28	0.272	1.087***	5.23	0.270
Expected Income if Graduate - Expected Income if Not Graduate[§]	0.516***	3.06	0.128	0.451***	2.88	0.112
Mother High School Graduate = 1	0.305***	3.46	0.076	0.306***	3.47	0.076
Mother Some College = 1	-0.051	-0.29	-0.013	-0.035	-0.20	-0.009
Father High School Graduate = 1	0.318***	2.94	0.079	0.258**	2.46	0.064
Father Some College = 1	-0.436**	-2.04	-0.108	-0.391*	-1.87	-0.097
Missing Father or Mother Education	-0.009	-0.08	-0.002	-0.020	-0.19	-0.005
Race (African American = 1)	-0.074	-0.60	-0.018	-0.069	-0.57	-0.017
Sex (Female = 1)	-0.294*	-1.73	-0.073	-0.260	-1.56	-0.065
Race × Sex	0.495***	3.00	0.122	0.455***	2.83	0.113
Proportion of Years Lived with a Single Parent, Ages 6–15	0.125	0.74	0.031	0.094	0.56	0.023
Proportion of Years Mother Worked, Ages 6–15	0.006	0.04	0.001	0.022	0.15	0.006
Proportion of Years Lived in Poverty, Ages 6–15	-0.141	-0.86	-0.035	-0.088	-0.53	-0.022
Proportion of Years Received AFDC, Ages 6–15	-0.520**	-2.48	-0.129	-0.458**	-2.23	-0.114
Proportion of Years Lived in SMSA, Ages 6–15	-0.130	-1.35	-0.032	-0.156	-1.61	-0.039
Proportion of Locational Moves, Ages 6–15	-1.311***	-6.52	-0.325	-1.326***	-6.59	-0.330
First Born = 1	0.465***	4.32	0.115	0.460***	4.29	0.114
Average Family Income-to-Needs Ratio, Ages 6–15	-0.003	-0.07	-0.001	0.042	1.08	0.011

(table continues)

TABLE 2, continued

Variable	Main Specification (2-stage) Coefficient Estimate	T-Statistic	Marginal Effect	Model with Selection Coefficient Estimate	T-Statistic	Marginal Effect
Proportion of High School Dropouts in Neighborhood, Ages 6–15	-0.011***	-2.64	-0.003	-0.011***	-2.68	-0.003
Proportion Years Head Disabled, Ages 6–15	-0.308**	-2.48	-0.076	-0.320***	-2.58	-0.080
Average Number of Siblings, Ages 6–15	-0.036	-1.43	-0.009	-0.038	-1.49	-0.009
Sex × Years Mother Worked, Ages 6–15	0.181	1.03	0.045	0.166	0.95	0.041
Sex × Years Received AFDC, Ages 6–15	0.588**	2.34	0.145	0.550**	2.21	0.137
Sex × Years in Single-Parent Family, Ages 6–15	-0.465**	-2.50	-0.115	-0.426**	-2.33	-0.106

§ This difference is the present discounted value of the expected income if a graduate minus the present discounted value of the expected income if not a graduate. See Table 1.

TABLE 3
Model Identification Assumptions:
Coefficients and T-Statistics of Present Values of Income Gain from Graduation, and Simulated Change in Probability of High School Graduation–Preferred Model (N=1,942)

A. Relevance of Identifiers

	Log-likelihood test statistic for joint significance of identifying variables[a]		Percent change in R-squared when identification variables are included[b]	
Income Regression	High School Graduation	High School Dropout	High School Graduation	High School Dropout
Age 19	14.0	12.4	15.5	23.2
Age 20	10.2	9.8	6.9	18.0
Age 21	13.6	9.2	9.1	11.5
Age 22	4.8	5.8	4.0	10.3
Age 23	4.8	6.0	4.1	11.0
Age 24	8.2	3.0	7.0	7.1
Age 25	16.2	5.2	12.0	13.5
Age 26	13.8	8.4	7.4	10.4
Age 27	12.8	12.4	8.5	15.0
Age 28	10.2	7.0	7.2	12.0
Age 29	9.8	4.8	6.0	11.2
Age 30	7.4	3.2	3.7	4.5
Age 31	10.4	4.8	6.8	11.3
Age 32	11.4	1.2	7.5	3.9

B. Alternative Exclusion Restrictions for Model Identification

Included in High School Graduation Equation	Estimated Income Difference Coefficient	T-Statistic	Simulated Change in Probability of Graduating[c]
Family income	0.4646	2.61	.0143
Neighborhood income	0.5752	2.99	.0175
West	0.4990	2.94	.0153
Northeast	0.4134	2.34	.0127
South	0.5050	2.97	.0155
All three regions	0.3863	2.09	.0119

[a]The critical value is 9.2 at the 10 percent level and 11.1 at the 5 percent level.
[b]The percentage change in R-squared is calculated for an OLS regression on income.
[c]The simulated change in the probability of graduating shows the change in the probability of graduating if expected income associated with graduating were to increase by 10 percent.

TABLE 4
Alternative Specifications of Model of High School Graduation Choice: Coefficients and T-Statistics of Present Values of Income Gain from Graduation, and Simulated Change in Probability–Preferred Model (N=1,942)

	Estimated Income Difference Coefficient	T-Statistic	Simulated Change in Probability of Graduating[a]
Preferred Specification (Table 2, columns 1–3)	0.5155	3.06	.0158
Model with Sample Selection (Table 2, columns 4–6)	0.4510	2.88	.0137
A. *Alternative Aggregation of Education Level*			
Limit reference sample to those who did not attend college	0.6024	3.78	.0164
B. *Alternative Process of Forming Expectations*			
Rational expectations estimated for reference sample (n=1,356)	0.5425	1.90	.0158
C. *Alternative Measures of Difference in Present Values of Income*			
Personal Earnings	0.4449	3.06	.0129
Family Income-to-Needs Ratio	0.0063	1.86	.0125
Net Present Value of Income (rather than income at each age)	0.4558	2.84	.0140
D. *Schooling-Specific Effect of Income*			
$Y_{HS} \div 1000$	0.5065	1.89	.0156
$Y_{NO} \div 1000$	-0.5190	2.78	-.0106
E. *Alternative Discount Rates*			
1 percent	0.4358	3.12	.0162
5 percent	0.6017	2.99	.0154
10 percent	0.8385	2.82	.0141
25 percent	1.3384	2.32	.0106
F. *Group-Specific Discount Rates*			
If income-to-needs ratio < 2, discount rate = 10%	0.4156	2.54	.0104
If ever in poverty ages 6–15, discount rate = 10%	0.3804	2.35	.0097
G. *Alternative Time Horizons*			
Expected income extended to age 59	0.1999	3.74	.0214
Expected income truncated at age 27	0.6373	2.63	.0125
H. *Effect of Uncertainty of Income*			
Add measures of expected income uncertainty if a graduate and if not a graduate	0.4665	2.69	.0143
I. *Alternative Definition of High School Graduation*			
Eliminate GED from sample (n=1,868)	0.5155	2.96	.0159

[a]The simulated change in the probability of graduating shows the change in the probability of graduating if expected income associated with graduating were to increase by 10 percent.

TABLE 5
Simulated Impact of Changes in (1) Expected Income if Graduating and (2) Parental and Family Characteristics on Probability of Dropping Out of High School
(Specifications from Table 2, column 1)

Increasing the Present Value of Expected Income if Graduate

	Probability Of Dropping Out	Percent Change
Base probability	.1553	
10% increase in expected income if graduating	.1395	-10.2%
25% increase in expected income if graduating	.1184	-23.7%
actual change in earnings differential 1970-1992*	.1244	-19.9%

Changes in Family and Background Characteristics

	Direct Effect	Indirect Effect	Total Effect	Percent Change
Both parents high school graduates	-.0585	.0041	-.0557	-35.9%
Years in single parent family = 0	-.0068	.0022	-.0057	-3.7%
Years receiving AFDC = 0	-.0051	.0021	-.0033	-2.1%
Years family moved location = 0	-.0052		-.0052	-3.3%
Neighborhood % Dropouts decreases 10%	-.0036	-.0008	-.0043	-2.8%

*The actual change in earnings differential is gender- and race-specific, and is from Schweitzer (1997).

APPENDIX TABLE 1
Summary Statistics of Variables Used in Income Equations and Selection Equation

Variable	Variables Used in Income Equations: High School Graduates Mean	High School Graduates Std. Dev	Nongraduates Mean	Nongraduates Std. Dev	Variables Used in Selection Probit Mean	Selection Probit Std. Dev
Average Number of Siblings, Ages 12–15	3.00	1.92	3.61	2.16	3.12	1.98
First Born = 1	0.16	0.37	0.09	0.28	0.15	0.36
Proportion of Years Lived with a Single Parent, Ages 12–15	0.27	0.42	0.38	0.45	0.29	0.43
Proportion of Years Mother Worked, Ages 12–15	0.55	0.42	0.44	0.40	0.53	0.42
Proportion of Years Received AFDC, Ages 12–15	0.14	0.30	0.26	0.36	0.16	0.31
Mother High School Graduate = 1	0.50	0.50	0.22	0.41	0.44	0.50
Mother Some College = 1	0.13	0.34	0.02	0.14	0.11	0.31
Father High School Graduate = 1	0.37	0.48	0.13	0.34	0.32	0.47
Father Some College = 1	0.15	0.36	0.02	0.12	0.12	0.33
Missing Father or Mother Education	0.26	0.44	0.31	0.46	0.27	0.44
Race (African American = 1)	0.49	0.50	0.59	0.49	0.51	0.50
Sex (Female = 1)	0.52	0.50	0.47	0.50	0.51	0.50
Race × Sex	0.27	0.44	0.29	0.45	0.27	0.45
Proportion of Years Lived in SMSA, Ages 12–15	0.75	0.41	0.72	0.43	0.75	0.41
Proportion Years Head Disabled, Ages 12–15	0.19	0.33	0.32	0.40	0.22	0.35
Proportion of High School Dropouts in Neighborhood, Ages 12–15	17.11	10.38	23.51	10.93	18.39	10.80
Sex × Years Mother Worked, Ages 12–15	0.28	0.41	0.20	0.34	0.26	0.40
Sex × Years in Single-Parent Family, Ages 12–15	0.14	0.34	0.17	0.35	0.15	0.34
Sex × Years Received AFDC, Ages 12–15	0.06	0.21	0.13	0.29	0.08	0.23
Proportion of Years Lived in Poverty, Ages 12–15	0.23	0.36	0.44	0.41		
LN Family Income, Ages 12–15	10.59	0.64	10.15	0.57		
Lived in West, Ages 12–15	0.15	0.35	0.13	0.33		
Lived in Northeast, Ages 12–15	0.17	0.37	0.07	0.25		
Lived in South, Ages 12–15	0.43	0.49	0.53	0.50		
Proportion of Families in Neighborhood with High Income, Ages 12–15	0.18	0.15	0.11	0.09		
Lambda from First Stage Used in Third Stage Model	0.29	0.21	-1.18	0.40		
Proportion of Locational Moves, Ages 12–15					0.14	0.20
Average Family Income-to-Needs Ratio, Ages 12–15					2.12	1.67

APPENDIX TABLE 2
Summary Statistics of Variables Used in High School Graduation Probit
(N=1942)

Variable	Mean	Std. Dev
Expected Income if Graduate - Expected Income if Not Graduate§	67,785	52,161
Mother High School Graduate = 1	0.49	0.50
Mother Some College = 1	0.12	0.33
Father High School Graduate = 1	0.41	0.49
Father Some College = 1	0.16	0.36
Missing Father or Mother Education	0.24	0.43
Race (African American = 1)	0.48	0.50
Sex (Female = 1)	0.50	0.50
Race × Sex	0.25	0.43
Proportion of Years Lived with a Single Parent, Ages 6–15	0.29	0.40
Proportion of Years Mother Worked, Ages 6–15	0.56	0.36
Proportion of Years Lived in Poverty, Ages 6–15	0.24	0.32
Proportion of Years Received AFDC, Ages 6–15	0.15	0.28
Proportion of Years Lived in SMSA, Ages 6–15	0.74	0.41
Proportion of Locational Moves, Ages 6–15	0.16	0.18
First Born = 1	0.20	0.40
Average Family Income-to-Needs Ratio, Ages 6–15	2.26	1.65
Proportion of High School Dropouts in Neighborhood, Ages 6–15	17.13	9.34
Proportion Years Head Disabled, Ages 6–15	0.18	0.29
Average Number of Siblings, Ages 6–15	2.65	1.62
Sex × Years Mother Worked, Ages 6–15	0.30	0.42
Sex × Years Received AFDC, Ages 6–15	0.07	0.23
Sex × Years in Single-Parent Family, Ages 6–15	0.16	0.35

§ The present discounted value of the expected income if a graduate minus the present discounted value of the expected income if not a graduate.

Appendix Table 3

Model Identification Assumptions:
Coefficients and T-Statistics of Present Values of Income Gain from Graduation, and Simulated Change in Probability of High School Graduation–Alternative Model (N=1,942)

A. Relevance of Identifiers

	Log-likelihood test statistic for joint significance of identifying variables[a]		Percent change in R-squared when identification variables are included[b]	
Income Regression	High School Graduation	High School Dropout	High School Graduation	High School Dropout
Age 19	14.5	17.5	17.7	21.5
Age 20	10.2	13.0	9.3	16.9
Age 21	12.5	6.2	10.4	7.9
Age 22	4.7	6.8	4.0	9.9
Age 23	5.2	10.5	3.8	11.1
Age 24	6.5	5.0	4.5	7.3
Age 25	14.0	9.9	7.6	13.3
Age 26	9.4	7.3	4.4	8.4
Age 27	13.2	12.8	7.7	14.5
Age 28	11.1	5.2	5.6	5.9
Age 29	8.8	4.7	4.1	6.8
Age 30	4.7	0.8	2.2	1.2
Age 31	10.6	4.2	5.6	8.3
Age 32	9.9	1.4	5.0	2.2

B. Alternative Exclusion Restrictions for Model Identification
(exclusion restriction variable included in final high school graduation equation)

Included in High School Graduation Equation	Estimated Income Difference Coefficient	T-Statistic	Simulated Change in Probability of Graduating[c]
Family income	0.3994	2.37	.0127
Neighborhood income	0.5374	2.88	.0169
West	0.4221	2.67	.0134
Northeast	0.3706	2.30	.0118
South	0.4414	2.76	.0140
All three regions	0.3253	1.90	.0104

[a]The critical value is 9.2 at the 10 percent level and 11.1 at the 5 percent level.
[b]The percentage change in R-squared is calculated for an OLS regression on income.
[c]The simulated change in the probability of graduating shows the change in the probability of graduating if expected income associated with graduating were to increase by 10 percent.

Appendix Table 4
Alternative Specifications of Model of High School Graduation Choice: Coefficients and T-Statistics of Present Values of Income Gain from Graduation and Simulated Change in Probability-Model with Sample Selection (N=1,942)

	Coefficient Estimate	T-Statistic	Simulated Change in Probability[a]
Baseline Specification with Selection	0.4510	2.88	.0137
Sample Selection			
No Correction for Lambda in Predicted Income	0.5293	2.80	.0176
A. *Alternative Aggregation of Education Level*			
Limit reference sample to those who did not attend college	0.4250	2.71	.0130
B. *Alternative Process of Forming Expectations*			
Rational expectations estimated for reference sample (n=1,356)	0.6387	2.57	.0172
C. *Alternative Measures of Difference in Present Values of Income*			
Personal Earnings	0.4714	2.99	.0136
Family Income/Needs Ratio	0.0052	1.76	.0086
Net Present Value of Income (rather than income at each age)	0.4273	2.72	.0131
D. *Schooling-Specific Effect of Income*			
$Y_{HS} \div 1000$	0.4378	1.79	.0133
$Y_{NO} \div 1000$	-0.4556	2.67	-.0099
E. *Alternative Discount Rates*			
1 percent	0.3798	2.91	.0140
5 percent	0.5287	2.84	.0134
10 percent	0.7483	2.74	.0124
25 percent	1.2730	2.44	.0099
F. *Group-Specific Discount Rates*			
If income ÷ needs < 2, discount rate = 10%	0.3609	2.37	.0090
If ever in poverty ages 6 to 15, discount rate = 10%	0.4507	2.33	.0114
G. *Alternative Time Horizons*			
Extended to age 59	0.1624	3.27	.0174
Shortened to age 27	0.5803	2.64	.0112
H. *Effect of Uncertainty of Income*			
Add measure of expected income uncertainty	0.3866	2.30	.0118
I. *Alternative Definition of High School Graduation*			
Eliminate GED from sample (n=1,868)	0.4689	2.91	.0143

[a]The simulated change in the probability of graduating shows the change in the probability of graduating if expected income associated with graduating were to increase by 10 percent.

Center for Demography and Ecology
University of Wisconsin
1180 Observatory Drive Rm. 4412
Madison, WI 53706-1393
U.S.A.
608/262-2182
FAX 608/262-8400
comments to: bwolfe@lafollette.wisc.edu
requests to: cdepubs@ssc.wisc.edu