Econometrica, Vol. 67, No. 6 (November, 1999), 1295–1339

WHY YOUTHS DROP OUT OF HIGH SCHOOL: THE IMPACT OF PREFERENCES, OPPORTUNITIES, AND ABILITIES

BY ZVI ECKSTEIN AND KENNETH I. WOLPIN[1]

In this paper, we develop and structurally estimate a sequential model of high school attendance and work decisions. The model's estimates imply that youths who drop out of high school have different traits than those who graduate—they have lower school ability and/or motivation, they have lower expectations about the rewards from graduation, they have a comparative advantage at jobs that are done by nongraduates, and they place a higher value on leisure and have a lower consumption value of school attendance. We also found that working while in school reduces school performance. However, policy experiments based on the model's estimates indicate that even the most restrictive prohibition on working while attending high school would have only a limited impact on the high school graduation rates of white males.

KEYWORDS: High school attendance, employment, sequential model, structural estimation, ability heterogeneity, preference heterogeneity.

I. INTRODUCTION

REDUCING THE RATE AT WHICH YOUTHS drop out of high school is considered a useful social goal. It is well documented that dropouts have lower earnings and are more likely to engage in antisocial behaviors.[2] However, the factors that cause youths to drop out of high school would seem to be diverse. According to data from the 1979 youth cohort of the National Longitudinal Surveys of Labor Market Experience (NLSY79), when a sample of white male youths who had not yet graduated from high school and were not currently attending were asked to choose from a list of reasons for their nonattendance, approximately 30 percent chose the response category "didn't like school" as the main reason, an additional 14 percent cited "offered a good job, chose to work" and 9 percent "lack of ability, poor grades."[3] Although the validity of these responses may be subject to question, it would not seem unreasonable to infer that preferences, for school

[1]We are grateful for support from the National Science Foundation under Grant No. SBR-9309636. Comments by Tom MaCurdy, John Rust, the editor and an anonymous referee greatly improved the paper.

[2]The private and social value of reducing the dropout rate depends on whether those youths who are dropouts would face improved labor market opportunities and would alter their behavior if they successfully completed the requirements for high school graduation. We show below that dropouts differ from graduates in important traits that affect their propensity to graduate and that at least some of those traits would have to be altered in order to increase the rate at which they graduate. Thus, the "effect" of graduating on opportunities and behaviors depends on which of those traits are altered as well as the direct effect of graduating *per se*.

[3]In addition, 20 percent chose an unspecified "other" category and 13 percent were under a school suspension or had been expelled. The data are from the 1979–1982 survey rounds of the 1979 youth cohort National Longitudinal Surveys of Labor Market Experience (NLSY79).

and for other substitute or complementary activities, as well as external and internal constraints are likely important elements in the decision.[4]

In this paper, we formulate and estimate an explicit sequential decision model of high school attendance and work that accommodates the variety of responses observed in the NLSY79 sample. We assume that youths choose among work-school combinations in order to maximize their expected lifetime utility at each decision period. The availability of high school transcripts in the NLSY79 enables us to model in some detail important institutional features of high school grade progression. To briefly outline the model, youths who attend school accumulate credits (courses) towards graduation and receive course-specific performance grades ("A"–"F"). Performance grades are probabilistic, depending on the individual's history of prior performance, the level of participation in the labor market (hours worked), and their known (to them) ability and motivation. Thus, graduation, which requires the accumulation of a fixed number of credits, is a probabilistic outcome that can be influenced by work decisions.[5] The labor market (randomly) offers up wages for part-time and full-time employment that depend also on some inherent skill "endowment" (potentially correlated with the ability and motivation associated with school performance) as well as labor market experience. Working, in addition to potentially reducing school performance, directly reduces leisure time, which is itself valued and which may differ among youths.[6] The value of attending high school consists of both its current consumption value, which is random, and the youth's perceived utility payoff (including the potential for increased earnings) to graduation, each of which may differ among the population.[7]

[4]For example, responses may be biased due to ex post rationalization. In addition, the large "other" category may subsume singularly important factors such as a low expected return to a high school diploma.

[5]The performance grade function should be thought of as a production function for new knowledge, with the grade earned in a course as the measure of new knowledge acquired. A complete specification of that production function would include the amount and quality of instruction time, the amount of time spent studying and doing homework, and the usage of complementary inputs such as computers and library resources. Unfortunately, the only measure of time allocation available longitudinally in the NLSY79 is market hours worked (study time is available in one survey round) and there are no measures of goods inputs. Strictly speaking, it is assumed that, for those attending school, hours spent working subtracts in a fixed proportion from the amount (and quality) of time that would be applied to learning.

[6]We do not have information on the types and intensities of nonmarket/leisure activities engaged in by youths, e.g., illicit income-earning (criminal) activities, the use of drugs (which might also be an input into the grade function if it reduces learning efficiency). Moreover, explicitly modeling the choice of those activities, although appropriate, would greatly expand the scope and complexity of the framework. One can view the value of those activities as subsumed in the (composite) value of leisure and of policies that are intended to limit the participation in those kinds of nonmarket activities, e.g., deterring crime or drug use through more severe punishment, as equivalent to a reduction in the value of leisure.

[7]School-based extra-curricular activities, such as participation in sports or clubs possibly can be thought of as contained in both the value of leisure and the consumption value of attending school.

Given their diversity of experiences and backgrounds, it is unlikely that youths begin high school with the same set of preferences (for leisure and for school), skills, abilities and motivation (with respect to school and the labor market) or expectations about the value of a high school diploma. Although preferences may change, skills may be augmented and expectations altered, the importance of these initial (upon entry into high school) traits may be large and persistent. For example, youths with persistently low motivation may indeed be able to earn a "B" in one course for idiosyncratic reasons (the teacher was an especially good motivator), and this additional knowledge may improve their chance of earning a higher grade in future courses, but the effect on future grades may be small relative to the persistent effect of low motivation.[8]

Moreover, youths who lack motivation to perform well in school may be more likely to choose to work while attending school given their greater likelihood of failure. Without accounting for such differences among youths, it may appear that working is more detrimental to school performance than it is. On the other hand, youths who have low motivation in the school domain may be less likely to work if they also have a relatively high value of leisure (disutility of work), in which case it may appear that working enhances school performance. In either event, in order to draw appropriate conclusions, it is clearly important to account for the existence of persistent heterogeneity in multiple traits that may themselves be related.

The model is estimated using data from the NLSY79. The estimation method combines the solution of the dynamic optimization problem with the maximization of a likelihood function that accounts jointly for annually observed work-schooling choices, wages, credits earned, and grades.[9] As argued above, identifying persistent initial traits is important both as an end in itself and as a means of guarding against unwarranted inferences. However, we do not know how to measure directly the kinds of traits, i.e., preferences, abilities, and expectations, that may be critical in determining whether or not a youth drops out of high school prior to graduation.

The methodology we adopt treats initial traits as unmeasured.[10] To accommodate this unmeasured heterogeneity at the time of initial high school entry, we assume that within the NLSY79 cohort there are a fixed number of discrete

[8]To the extent that these initial differences are important in determining performance in high school and they are not immutable, e.g., they are in part the outcome of parental investments, implementing early interventions that succeed in altering those traits might be efficacious.

[9]Methods of solving and estimating models with a discrete-choice dynamic programming structure are now well known and have been applied to a variety of issues. Examples are Eckstein and Wolpin (1989a), Keane and Wolpin (1994, 1997), Miller (1984), Pakes (1986), Rust (1987), and Wolpin (1984, 1987). Eckstein and Wolpin (1989b) and Rust (1992) provide useful surveys.

[10]The NLSY79 does have information on respondents' family background characteristics (parents' schooling, religion, country of origin, household structure during childhood). Our method does permit us to relate these characteristics to our estimates of initial traits.

types of youths who differ in the parameters that describe their preferences for school and leisure, their school ability and motivation and their expectation as to the value of a diploma. Because we cannot know a youth's type, the likelihood function is a mixture over types weighted by their sample probabilities (Heckman and Singer (1984)), where solving the dynamic optimization problem for each configuration of type-specific parameters provides the type-conditioned likelihood functions (Eckstein and Wolpin (1990), Keane and Wolpin (1997)).

To illustrate the importance of explicitly modeling the decision process, consider the following data relating work while attending school to subsequent graduation. During the week prior to each of the 1979–1982 NLSY79 survey rounds, those white male youths who were attending grade nine and who ultimately graduated from high school worked an average of 2.2 hours, but those who did not graduate averaged 4.0 hours.[11] Contrary to the inverse relationship between graduation rates and work in grade nine, graduation rates and work are unrelated for those attending grade 10, and positively related for grades 11 and 12.[12] Although the relationship is thus ambiguous, its interpretation would be unclear even if the same relationship had held at all grade levels. Almost everyone does attend grade nine, so that sample selection is probably not an important problem; however, with foresight, those who perceive their chances of graduating as small might be more likely to work. By grades 11 and 12, the selection issue is more problematic and the positive relationship between working and graduation might reflect inherent differences in ability and motivation rather than the acquisition of affective skills gained from employment that are also useful in school. The model we estimate accounts for this selection.

Given our parameter estimates, we are able to answer the question of whether working while attending high school affects performance in high school and, in addition, to determine the extent to which further restrictions on employment would affect the dropout rate. Determining the impact of work on performance in high school has been the subject of considerable study, although mostly not by economists, and is still much debated.[13] Indeed, an underlying premise of the Fair Labor Standards Act (FLSA), the main federal legislation regulating the use of child labor, is that working while attending school ad-

[11] The t value for the test of equality is 2.17. There are 422 observations.

[12] The differences and t values are 0.35 (.46) for grade 10, 1.96 (2.05) for grade 11, and 1.92 (1.36) for grade 12. Sample sizes are 719, 986, and 1,067.

[13] See Greenberger and Steinberg (1986) for a survey of the literature. See also D'Amico (1984), Marsh (1991), and the more recent paper by Mortimer et. al. (1996). There has also been related research, also primarily by noneconomists and thus not explicitly decision-theoretic, on the relationship between working and college performance. See Hood and Maplethorpe (1980) for a somewhat dated summary. In the economics literature, Ehrenberg and Sherman (1987) address this issue by estimating what can be thought of as a statistical representation of an approximation to a sequential optimization problem under uncertainty.

versely affects school performance.[14] Working obviously reduces the amount of time available for other activities, including time devoted to studying.[15] Further, working may be arduous and have negative spillover effects on classroom attentiveness. Our results indicate that working while attending high school does reduce academic performance. However, the quantitative effects are small. Estimates of the behavioral model imply that implementing a policy that forced youths to remain in high school for five years or until they graduate, whichever comes first, without working would increase the number of high school graduates by slightly more than 2 percentage points (from 82 to 84.1 percent).

Estimation of the model allows us to address two central questions about dropout behavior, namely (i) who drops out of high school, i.e, how do dropouts differ from graduates in their unmeasured persistent initial traits and how are those traits related to observable family background characteristics, (ii) why do youths drop out, i.e., which initial traits, if any, are important in terms of explaining the propensity to drop out. Our findings indicate that dropping out of high school is confined to youths with specific traits: lower school ability and/or motivation, a lower expected value of a high school diploma, higher skills in the kinds of jobs that do not "require" a high school diploma, a higher value placed on leisure, and a lower consumption value of attending school. In addition, the reasons for dropping out are complex. For example, even though youths who drop out have lower ability/motivation, given their other traits most would still drop out if their level of ability/motivation was as high as the modal high school graduate.

Our results also imply that allowing for a small number of types in terms of initial traits provides an approximate sufficient statistic for the family background characteristics that are available in the data, e.g., parental schooling levels, family income, household structure, for explaining dropout behavior. Perhaps somewhat surprisingly, the discrete types were also sufficient statistics in terms of explaining completed schooling levels and criminal behavior.

[14] The FLSA was first passed by the U.S. Congress in 1938 and has since been strengthened by a series of amendments. The most severe restrictions apply to those minors under the age of 16, who are permitted to work only in nonmining, nonhazardous, and nonmanufacturing jobs and then only under conditions that do not interfere with their schooling or health. Minors under the age of 18 are prohibited from working in nonagricultural jobs that have been declared as especially hazardous. The FLSA also regulates the hours that minors can work. During times when schools are in session, 14- and 15-year-old children may be employed for no more than 18 hours weekly and for not more than three hours in any one day. The time of employment during the day is also restricted to between 7 a.m. and 7 p.m. However, full-time work, up to a maximum of 8 hours per day and 40 hours per week, is permitted during periods when schools are not in session. There are no hours limitations in the FLSA, regardless of time of year, for 16- and 17-year-old minors. State child labor laws tend to be even more restrictive, setting shorter daily and weekly hours limitations for periods when schools are in session.

[15] In the sociological literature, this substitution of time between school and work activities has been referred to as a zero-sum model (Coleman (1984), Marsh (1991)).

The structure of the paper is as follows. In the next section, we present the basic behavioral model. We briefly discuss the numerical solution method and the estimation method. Section III describes the data and presents descriptive statistics. In Section IV we present the specific parameterizations used in implementing the model and discuss the identification properties of the model. The following section provides the estimates of the model and discusses their implications. Section VI concludes.

II. MODEL

The Basic Structure

Choices about high school attendance and employment are made within institutional settings. The structural estimation approach requires that we explicitly specify the school environment in terms of grade level progression and high school graduation requirements and the labor market environment in terms of hours and wage opportunities.

The School Environment

Consider a youth who enters high school (grade level nine).[16] Each year a high school attendee must take (exactly) five course credits and must accumulate a total of 20 credits to obtain a (regular) high school diploma.[17] Completion of high school, therefore, takes a minimum of four years.[18] A letter grade is received for each credit and is converted to a 0–4.0 numerical scale, where "A" = 4.0, "B" = 3.0, "C" = 2.0, "D" = 1.0, "F" = 0. A passing grade in a course (A–D) represents incremental knowledge and progression towards the degree, i.e., a credit is earned. We denote the cumulative grade point average (GPA) achieved up to year t as G_t and the total number of credits earned up to year t as C_t. Completed grade levels (e_t) correspond to five-credit increments, i.e., $e_t = 8$ if $C_t < 5$, $e_t = 9$ if $5 \leq C_t < 10$, $e_t = 10$ if $10 \leq C_t < 15$, $e_t = 11$ if $15 \leq C_t < 20$, and $e_t = 12$ (high school graduation) if $C_t \geq 20$. In addition, we assume that

[16] Although the model is presented as the decision problem of a single youth, an important feature of our empirical implementation is that youths may be heterogeneous with respect to these factors (see below).

[17] We ignore the possibility of obtaining alternative certification such as the General Educational Development (GED) credential. The evidence suggests that a GED has a low pecuniary return (Cameron and Heckman (1993)).

[18] This structure of educational attainment during high school is only somewhat stylized. Although the actual attainment process varies considerably both within and among states, according to data from the National Center for Education Statistics (NCES) the average number of credits completed in four years of high school by public high school graduates was 21.5 in 1982. In 1990, the modal number of credits required for graduation was 20 (13 states) with eight additional states requiring 21 credits. The model is empirically implemented within the assumed structure by forcing compatibility of the data with the model as described below.

high school must be completed within five (calendar) years of first entry.[19] Thus, failing a total of six or more credits at any stage or having dropped out of school for more than one year precludes ever graduating from high school with a regular diploma.

The number of credits earned (c_t) in the school year is simply the number of courses (out of the five taken) in which the youth receives a passing grade. If course grades are statistically independent (conditional on a set of determinants including ability/motivation), then the probability of earning c credits is simply the cth power of one minus the probability of receiving an "F" in a single course. We assume that there is an underlying latent variable determining a course grade that reflects the incremental knowledge acquired (produced) in the course. Generically, the production function for incremental knowledge during any school year would depend on a student's start-of-year knowledge, on the student's ability and motivation, on the student's effort, and on school inputs, e.g., the quality of the teacher.[20] Operationally, incremental knowledge is modeled as depending on the number of cumulative credits so far earned (C_t) and the cumulative GPA so far achieved (G_t) (based on passing grades only, as failing grades are taken to imply that no incremental knowledge was acquired), school inputs are treated as unobservable iid time-varying factors, e.g., the quality of the instruction, and the youth's level of motivation/ability is treated as a permanent unobservable.[21] The exact form of the credit accumulation and grade functions are provided later.

As noted, student effort (study time) is assumed to vary (inversely) with hours worked in the labor market (h_t). We interpret the impact of hours worked on grades as determining the extent to which working while attending high school is detrimental to school performance. The major problem with this interpretation, i.e., with using hours worked to proxy for student effort, is that study time can be thought of as a decision distinct from participation in the labor market. In that case, the production function is misspecified.

Market Opportunities

Because the post-graduation decision problem of eventual high school graduates is not modeled, labor market opportunities (hours and wage offers) are specified only during periods in which an individual is not a high school graduate (periods of high school attendance and after the five year eligibility period). Hourly wage offers for part- and full-time work for non-high-school graduates are assumed to depend only on prior work experience as measured by cumula-

[19] This assumption, while not universally true, is only rarely violated in our sample (see below).

[20] It is not possible to allow these production functions to differ by subject matter both because of data limitations and tractability.

[21] Clearly, permanent differences in school quality will be confounded with ability/motivation as are differences in students' knowledge upon entry into high school, which are fixed initial conditions.

tive hours worked and on an idiosyncratic shock.[22] Specifically, the hourly wage offer is $w_t^j = \exp(w^j(H_t))\exp(\epsilon_t^j)$ for $j = p, f$, where H_t is accumulated hours worked up to t.

Choice Set

At the beginning of each school year an individual who is still eligible for high school ($t \leq 5$ and $e_t < 12$) chooses whether to attend school and/or whether to work in the labor market h_t hours per week (during the school year). The number of hours worked per week can take on any of six discrete values: $\{0, 10, 20, 30, 40, 50\}$. The first two positive hours correspond to part-time jobs, $h_t^p = \{10, 20\}$ and the last three to full-time jobs $h_t^f = \{30, 40, 50\}$. In each period, an individual receives both a part- and a full-time job offer. However, there is only a partial choice of hours. With probability $\pi_t^h(j,k)$ the individual receives a part-time job offer of $h_t^p = j$ hours and a full-time job offer of $H_t^f = k$ hours. Because an offer of each type of job is received in each period, $\Sigma_j \Sigma_k \pi_t^h(j,k) = 1$.

The choice set for someone who is eligible to attend high school, therefore, consists of six mutually exclusive and exhaustive dichotomous alternatives: $d_t^k = \{d_t^{sn}, d_t^{sp}, d_t^{sf}, d_t^{nn}, d_t^{np}, d_t^{nf}\}$, i.e., attend school and not work ($d_t^{sn} = 1$), attend school and work part-time ($d_t^{sp} = 1$), attend school and work full-time ($d_t^{sf} = 1$), not attend school and not work ($d_t^{nn} = 1$), not attend school and work part-time ($d_t^{np} = 1$), and not attend school and work full-time ($d_t^{nf} = 1$). On the other hand, a high school dropout who is ineligible to return to school ($t > 5$) chooses among only the latter three alternatives: $d_t^k = \{d_t^{nn}, d_t^{np}, d_t^{nf}\}$. As noted, we do not explicitly model the post-graduation decisions of those who graduate within the five year period, although we do take the value attached to graduation into account in the optimization problem and in the estimation.

Preferences

Contemporaneous utility is assumed to be linear and additive in consumption, school attendance, and nonwork time, where consumption is the sum of the youth's earnings and family transfers. The utility specification rules out income effects on behavior. We further assume that the value of family transfers (in-kind as well as monetary) are not contingent on the youth's choices or school performance.[23]

Attending school involves mental effort as well as a loss of leisure time. On the other hand, learning may be valued *per se* and social activities, both intra-

[22] An additional hour of work experience is assumed to affect earnings equally regardless of whether that extra hour was the result of working in a part-time or a full-time job. In addition, it is assumed that there is no wage return to completing additional years of high school (without graduating).

[23] These assumptions are made, in part, because we have no data on intra-family transfers, but also in order to avoid having to model game-theoretic interactions between youths and their parents. Saving for college as a motivation for working while in high school is also ruled out.

and extra-curricular, may provide positive "consumption" value. The (net) utility obtained from attending school during the year depends on the psychic valuation attached to these various factors. However, the valuation of some of them may be affected by the extent of market work, e.g., working may constrain a student's ability to engage in extra-curricular activities, and by a student's level of maturity. We thus allow for a psychic value of school attendance, b_t^s, that explicitly depends on the level of work activity and that varies with age both systematically and stochastically, i.e., $b_t^{sj} = \bar{b}^{sj}(t) + \epsilon_t^{sj}$ for $j = \{n, p, f\}$. Nonwork time is valued at b_t^n dollars per hour and has a random and systematic age-varying component, i.e., $b_t^n = \bar{b}^n(t) + \epsilon_t^n$. There are L total hours per week that can be divided between work and nonwork.

To be concrete, current period utility at time $t(U_t)$ given the choice of alternative (k) is given by

$$
\begin{aligned}
&U_t^{sn} = b_t^n \cdot L + b_t^{sn},\\
&U_t^{sp} = b_t^n \cdot (L - h_t^p) + b_t^{sn} + b_t^{sp} + w_t^p \cdot h_t^p,\\
&U_t^{sf} = b_t^n \cdot (L - h_t^f) + b_t^{sn} + b_t^{sf} + w_t^f \cdot h_t^f,\\
&U_t^{nn} = b_t^n \cdot L,\\
&U_t^{np} = b_t^n \cdot (L - h_t^p) + w_t^p \cdot h_t^p,\\
&U_t^{nf} = b_t^n \cdot (L - h_t^f) + w_t^f \cdot h_t^f.
\end{aligned} \tag{1}
$$

The random elements subsumed in (1) $\{e_t^j, \epsilon_t^{sj}\}$, $j = \{n, p, f\}$ are assumed to be jointly serially independent and joint normal, i.e., $N(0, \Omega)$. The t-period shocks are revealed at t, but are unknown before t. Utility is normalized to consumption units, i.e., dollars. Notice that even if Ω is diagonal, the composite errors associated with the mutually exclusive choices will in general not be uncorrelated (e.g., every alternative contains ϵ_t^n and the consumption value of schooling is potentially correlated across the school-work alternatives) and with a logarithmic wage function the composite errors are also not normal.

The Optimization Problem

The individual is assumed to maximize the expected present discounted value of utility over an infinite horizon. Defining $V_t(S_t)$, the value function, to be this maximal expected present value at t, given the state space $S(t)$ at t, and given the discount factor β,

$$V_t(S_t) = \max_{\{d_t\}} E\left[\sum_{\tau=t}^{\infty} \beta^{\tau-t} \sum_k U_t^k d_t^k | S_t\right]. \tag{2}$$

The state space consists of all aspects of the history known to the individual that affect current alternative-specific utilities or the probability distribution of future utilities. As the model is specified, S_t includes G_t, C_t, H_t, and the ϵ_t's. The maximization of (2) is achieved by choice of the optimal sequence of feasible control variables $\{d_t^k\}$ given current realizations of the stochastic components of the utilities.

The value function can be written as the maximum over alternative-specific value functions. It is useful to define two value functions as follows, one for the post-high-school eligibility period ($t \geq 6$) and one for the eligibility period ($t \leq 5$):

$$V^1(S_t) = \max\left[V^{nn}(S_t), V^{np}(S_t), V^{nf}(S_t)\right], \quad t \geq 6,$$

$$V_t^2(S_t) = \max\left[V_t^{sn}(S_t), V_t^{sp}(S_t), V_t^{sf}(S_t), V_t^{nn}(S_t), V_t^{np}(S_t), V_t^{nf}(S_t)\right], \quad t \leq 5. \tag{3}$$

These alternative-specific value functions satisfy the Bellman equation (Bellman (1957)). For $t \geq 6$, i.e., for periods in which an individual (who did not graduate from high school) is no longer eligible to attend high school, these value functions are given by

$$V^k(S_t) = U_t^k + \beta E(V^1(S_{t+1}) | d_t^k = 1, S_t), \tag{4}$$

where $k = \{nn, np, nf\}$. Notice that the value function is stationary.[24] Expectations in (4) are taken over the stochastic components of the alternative-specific utilities and over the joint part- and full-time offered hours distribution.

To characterize the value functions during the high school eligibility period, define $\pi^s(c_t, g_t | S_t)$ to be the probability of earning c_t credits with grade point average g_t while attending school between time t and $t+1$ conditional on the state space at time t.[25] Although g_t is calculated only on passing grades (recall that failing grades are assumed not to contribute to incremental knowledge), we adopt the innocuous convention that $g_t = 0$ if $c_t = 0$. In addition, let V^D be the value of high school graduation at the time of graduation. Then for $t \leq 5$, the alternative-specific value functions are:

$$\begin{aligned} V_t^k(S_t) = U_t^k + \beta \Bigg[\sum_{c=0}^{5} \sum_{g=0}^{4} \pi^s(c_t, g_t | S_t, d_t^k = 1) \\ \times \big\{ I(t < 5, e_{t+1} < 12) E\left[V_{t+1}^2(S_{t+1}) | d_t^k = 1, S_t\right] \\ + I(t = 5, e_{t+1} < 12) E\left[V^1(S_{t+1}) | d_t^k = 1, S_t\right] \\ + I(e_{t+1} = 12) E\left[V^D | d_t^k = 1, S_t\right] \big\} \Bigg], \end{aligned} \tag{5}$$

where $I(\cdot)$ is an indicator function equal to one if the term inside the parentheses is true and zero otherwise. Over this period, the individual chooses among the six alternatives, $k = \{sn, sp, sf, nn, np, nf\}$. These value functions are nonsta-

[24] The alternative-specific utilities are specified below to ensure existence and uniqueness.

[25] If the individual chooses not to attend school at time t, then zero credits are earned between t and $t+1$ and the cumulative grade point average is unchanged.

tionary, i.e., they explicitly depend on t, because of the finite horizon over which a high school diploma must be earned.[26]

The value of graduating can be thought of as the solution to the optimization problem that the individual would face if he were to graduate from high school. As such, it implicitly embeds the decision about attending college. However, by treating it as estimable we avoid having to model that decision as well as the rest of the individual's lifetime decisions. As seen in (5), the expected value of graduation depends on the state space at t and thus evolves as prior decisions are made. As with the other aspects of the model, we provide the exact specification below.

Solution Method

The model does not admit to analytical solution, but can be numerically solved in a straightforward manner. The numerical complexity arises because the calculation of the value functions requires high-dimensional integrations. We follow the procedure in Keane and Wolpin (1994), using Monte Carlo integrations to evaluate the integrals that appear in (4) and (5).[27] In solving the stationary component of the model, it is assumed that wage offers are constant after $H = 50{,}000$, i.e., after accumulating the equivalent of 25 full-time (2,000 hours) years of working. This assumption provides a "terminal" stationary value for $\text{Emax}[V^{nn}(S), V^{np}(S), V^{nf}(S)]$ at $H = 50{,}000$. Stationary value functions are then solved recursively for $H < 50{,}000$ and pertain to all $t > 5$. The $E_t \max[V^{sn}(S), V^{sp}(S), V^{sf}(S), V^{nn}(S), V^{np}(S), V^{nf}(S)]$ functions for $t \leq 5$ are solved recursively from $t = 5$ as in any finite horizon model.

Estimation Method

Having solved for the Emax functions, equations (4) and (5) look like the indirect utilities associated with any panel data multinomial choice problem. There are three complications: (i) the errors as we have modeled the problem are not all additive and the composite errors of the mutually exclusive choices are correlated, (ii) in addition to the school-work choice, we observe the wage rate for those that work in a particular period (the accepted wage), the hours that they work, and their grades if they attend school, and (iii) the wage rate is

[26] Notice that the infinite horizon value function for nongraduates after period five is the terminal value function for the finite horizon decision problem. This treatment is similar to that found in Gilleskie (1998).

[27] Unlike the model in Keane and Wolpin, it is computationally manageable to evaluate the Emax functions at all elements of the state space rather than developing an approximating function. We used 250 draws to simulate the Emax function at each element of the state space.

clearly measured with error.[28] It is convenient to assume that the measurement error is multiplicative, i.e., $\ln w_t^{jo} = w^j(H_t) + \epsilon_t^j + \eta_t^j$, with $\eta_t^j \sim N(0,(\sigma_\eta^j)^2$ for $j = p, f$, where w_t^o signifies the observed wage.

At any time t, denote the vector of outcomes as $O_t = \{d_t^k, w_t^o, h_t, c_t, g_t\}$. The likelihood function for a sample of N individuals each observed from period $t = 1, \ldots, t_n$ is given by

$$\prod_{n=1}^{N} \Pr\left(O_{1n}, O_{2n}, \ldots, O_{t_n n} | S_0\right). \tag{6}$$

In general, the computation of (6) would require the calculation of multiple integrals of dimension at least equal to the number of periods times the number of alternatives.[29] Given the assumption of joint serial independence among the vector of shocks, the likelihood function can be written as the product of within-period outcome probabilities each of which is an integral of dimension equal to the number of alternatives.

To illustrate the calculation of the likelihood, it is sufficient to consider a specific outcome at some period. Suppose that the following is observed at period t: the individual chooses to attend school and work part-time ($d_t^{sp} = 1$), reports receiving a part-time wage rate of w_t^{po}, works h_t^p hours, and earns c_t credits during the school year with a grade point average g_t.

Further, assume that the individual entered the period having previously worked H hours and having earned a total of C credits with a cumulative GPA of G. The probability of that outcome is

$$\begin{aligned} (7) \quad & \Pr(d_t^{sp} = 1, w_t^{po}, h_t^p, c_t, g_t | G_t, C_t, H_t) \\ & = \pi(c_t, g_t | h_t^p, G_t, C_t) \cdot \pi(h_t^p) \\ & \cdot \int_{w_t^p} \Pr(d_t^{sp} = 1 | w_t^p, h_t^p, G_t, C_t, H_t) \cdot \Pr(w_t^p, w_t^{po} | G_t, C_t, H_t). \end{aligned}$$

We calculate the second line of (7), the joint probability of choosing $d_t^{sp} = 1$ and of observing a reported wage w_t^{po}, by a smoothed simulator.[30] The integration

[28] We observe a part-time wage for those that work part-time in a particular period and a full-time wage for those that work full-time in a particular period. Although there is likely to be measurement error in other variables, the measurement error in wages is obvious from the data whereas that is not the case for the other variables. Keane and Wolpin (1998) have recently developed an estimation method that allows for measurement error in all discrete choices.

[29] The measurement error in wages creates an additional integral in periods that include work.

[30] For each of K draws of the error vector, $\epsilon_t^n, \epsilon_t^f, \eta_t^p, \epsilon_t^{sn}, \epsilon_t^{sp}, \epsilon_t^{sf}$, noting that $\epsilon_t^p = \ln w_t^{jo} - \ln w^j(K_t, H_t) - \eta_t^p$, the kernel of the integral is

$$\exp\left[\frac{V_{tk}^{sf} - \max(V_{tk}^i)}{\tau}\right] \Big/ \sum_i \exp\left[\frac{V_{tk}^i - \max(V_{tk}^i)}{\tau}\right]$$

times the joint density of the observed and true wage. The first term in the kernel is the smoothed simulator of the probability that $d_t^{sp} = 1$, with τ the smoothing parameter. The smoothing parameter was set at 500, which is not unreasonable given that differences in value functions in the above kernel are as large as 10^5. The integral is then the average of the kernel over the K draws.

over the true wage in (7) is necessary because the choice depends on the true wage and we do not observe it. Other probability statements are calculated similarly.

Notice that the entire set of model parameters enter the likelihood through the choice probabilities that are computed from the solution of the dynamic programming problem. Subsets of parameters enter through other structural relationships as well, e.g., wage offer functions. The estimation procedure, i.e., the maximization of the likelihood function, iterates between the solution of the dynamic program and the calculation of the likelihood.

Preference and Endowment Heterogeneity

The model pertains to the decision problem of a single individual and the estimation method concomitantly applies to a homogeneous sample at the date of entry into high school. However, it is unlikely that at high school entry either ability/motivation, market skills, preferences for leisure or school, or expected valuations of high school graduation are the same for everyone. Moreover, while conceptually we may think of obtaining survey measurements of at least some of these factors, as of now adequate measures are not available.[31]

There are a number of ways to account for heterogeneity in these factors. One possibility is to treat all such heterogeneity as unobserved (by us). A standard (nonparametric) approach, and the one we adopt, is to allow for a finite mixture of types, say M types, each comprising a fixed proportion π_m ($m = 1, \dots, M$) of the population (e.g., see Heckman and Singer (1984), Keane and Wolpin (1997)). Thus, any given type would be described by a vector of parameters, given to them at the time of high school entry, corresponding to their preferences for leisure and school attendance, their school ability/motivation, their market skills, and the valuations they attach to graduation. Thus, for example, the deterministic part of the value of leisure for an individual of type m would be $\bar{b}_m^n$. The likelihood function in this case would be a finite mixture (or weighted average) of the type-specific likelihoods, namely

$$(8) \qquad \prod_{n=1}^{N} \sum_{m=1}^{M} \Pr\left(O_{1n}^m, O_{2n}^m, \dots, O_{t_n n}^m | \text{type} = m\right) \cdot \pi_m .$$

Thus, to calculate (8), the dynamic program must be solved for each type.

[31] Independent of the existence of heterogeneity, it is useful to think of the structure that we impose on the model in terms of functional form assumptions, distributional assumptions, and exclusionary restrictions, in addition to the assumptions of optimization and rational expectations, as being the necessary substitutes for not having measures of preferences, ability/motivation, and market skills. Although the NLSY respondents were administered a test often used as a measure of ability, the Armed Forces Qualifying Test (AFQT), it was given after differences in schooling had already emerged. There is evidence that the scores reflect these pre-existing differences in school attainment (Keane and Wolpin (forthcoming), Neal and Johnson (1997), Rosenzweig and Wolpin (1994)).

An alternative to the finite-mixture approach is to assume that the parameters representing initial preferences, ability/motivation, market skills, and the expected value of graduation are related to measured family background characteristics, e.g., parental schooling levels, family structure, etc., and that these measured characteristics control in part for population heterogeneity. It would be natural, in this case, to model each preference and endowment parameter as a parametric function of observable background characteristics, say a linear in parameters formulation, with, as in the pure finite-mixtures case, M type-specific constants.

Now, in the finite-mixtures case, by definition, all heterogeneity could be accounted for if there are as many types as there are people. However, to the extent that groups of individuals are identical, or nearly so, the number of types necessary to account for heterogeneity would be less than the number of people and the set of distinct family background types would map perfectly into the M heterogeneity types. In the nonparametric approach, estimates of the model's parameters would be inconsistent only if the number of types was misestimated. The gain to the nonparametric approach is considerable parsimony (there are M parameters for each component of preference and endowment heterogeneity vs. at a minimum M plus the number of family background characteristics for each). The loss is that we cannot estimate the impact of family background characteristics on preferences and endowments, which would entail, in addition, accounting in the estimation for the correlation between family background characteristics and unobserved heterogeneity. However, what is most important is that we capture the extent of heterogeneity fully with the nonparametric specification. As we discuss later, it is possible to evaluate whether the finite-mixture specification at least captures all of the heterogeneity associated with observable measured family background characteristics.

III. DATA

The NLSY79 consists of 12,686 individuals, approximately half of them male, who were 14 to 21 years of age as of January 1, 1979. There is a core nationally representative random sample and oversamples of blacks, Hispanics, and members of the military. This analysis is based on the white males in the core random sample who were less than age 15 as of October 1, 1977 and who had ever attended high school (at least grade level nine). There are 702 white males who meet these criteria (out of a total of 2439).[32] Interviews in the NLSY have been conducted annually since 1979, and we follow each individual from the first year they enter the ninth grade through either their year of high school graduation or, if they do not graduate, through their last interview (1991 at the latest).

The NLSY79 collects schooling and employment data as an event history retrospectively back to the preceding interview. Employment data include the

[32] The age restriction reduced the sample to 712 and the schooling restriction eliminated 10.

beginning and ending dates (to the calendar week) of all jobs (employers), all gaps in employment within the same job, usual hours worked on each job, and the usual rate-of-pay on each job. In the first interview, employment data were collected back to January 1, 1978. Given the age restriction imposed on our analysis sample, we effectively have a complete employment profile for each sample youth.

Schooling data include the highest grade attended and completed at each interview date, monthly attendance in each calendar month beginning in January 1980, school-leaving dates, and the dates of diplomas and degrees. In addition, the NLSY79 obtained and coded high school transcripts for much of the sample. Of the 702 youths in our subsample, 564 had usable transcript data.[33] The transcript data report all of the courses taken in high school, standardized (Carnegie) credits for each course, a grade level at which the course was taken (9 through 12), a course number and a grade for each course based on a 0 to 4.0 scale.[34]

For those who graduate in four years, which is the vast majority of youths, we essentially ignore the actual pattern of credit accumulation. Recall that in the model a student takes exactly five courses per year and needs to obtain a passing grade in 20 courses over at most five years in order to earn a diploma. Therefore for this group, we fixed the number of credits earned at five for each of the four years. Notice that in following this rule, those who fell behind but were able to make up failed credits and still graduate in four years are counted as having a normal grade progression.

There are a number of difficulties in using the transcript data for those who did not graduate within the four years, i.e., those who graduated in five years and those who never graduated. Most problematic is that calendar dates in which courses were taken are not reported. Thus, some transcripts report a large number of courses taken at a single grade level that obviously span more than one school year. Courses are supposed to be listed chronologically, but we don't know how many courses were taken in each school year. In addition, grade level designations are often meaningless. Some transcripts will show a youth advancing through several grade levels without ever having received a passing grade in a single course. Thus, we hand-edited each transcript of those who did not graduate within four years of entry. We attempted to adhere to some basic rules in establishing credits earned. However, it was impossible to codify the rules in a way that would avoid having to make judgements on a transcript-by-transcript basis.

Loosely speaking, our procedure was as follows. We began with grade level 9 and usually simply counted the number of credits earned over all courses with

[33]There were 588 youths with at least some transcript data. In 24 cases we could not sensibly match that data with the other available information. Those with transcript data are more likely to be high school graduates (80 percent as opposed to 65 percent) and are less likely to have attrited (4.6 percent as opposed to 7.9 percent in the 1983 interview).

[34]For more information see the NLS Handbook.

unique course numbers for which passing grades were received. However, we normally did not allow a student to accumulate more than one course credit in a particular subject in a single school year with additional such credits allocated to the next year. (Recall that we are dealing with students who did not graduate from high school.) If the number of credits earned over these courses totaled five or more, we assigned five credits to grade level 9 preliminarily. For those allotted five credits for grade level 9, we then looked at their grades in the major subjects, i.e., English, Math, Social Studies, and Science. For each of those courses with a unique course number in which a failing grade was reported, we subtracted its associated credits (a maximum of one credit was deducted for each subject). If failing grades were received in all of the four major subjects, zero credits were assigned for that year. A course for which a failing grade was reported and that was repeated at the same grade level was treated as course work taken in the following school year, as were failed courses with the same course numbers that were attached to the next higher grade level. We then repeated this procedure for the repeated grade level 9 courses and the grade level 10 courses, assigning credits earned to year two of high school. We repeated this procedure until all of the courses on the transcript had been exhausted.

This procedure determined the sequence of credits earned. The grade level assigned to a particular year was based on accumulated credits as per the model. Thus, a youth who earned three credits in year one of high school was assigned a completed grade level of eight upon entering year two of high school. If the youth earned two or more credits in the second year, then grade level would increase by one upon entering the third year of high school; if not, the grade level would remain at eight.

The calendar-time placement of the sequence of credits earned, and thus grade levels completed, was obtained from the main survey school enrollment data. From this match, it is apparent that the grade-level progression calculated from the transcript data as above does not match self-reported grade levels in the main survey. If self-reported grade levels reflect actual school assignments, it is evident that our accounting of credits earned does not correspond to the standards for progression practiced by high schools. Self-reported highest grade completed is considerably higher on average than is the same measure based on our credit accounting. For those who do not graduate from high school, the average level of schooling completed prior to their last year of attendance as self-reported is 10.0, but it is only 9.4 based on the accumulated credits as we have calculated from the transcript data.

GPA is also calculated from the transcript data and is based on only the five major subjects (the four above plus foreign language) that are taken in the school year. To conform to the model of knowledge acquisition, GPA is based only on passing grades. We discretized GPA into seven categories: 1.0, 1.5, 2.0, 2.5, 3.0, 3.5, 4.0. Those GPA values between 1.0 and 1.25 were assigned 1.0, those between 1.25 and 1.75 were assigned a 1.5, etc.

The number of hours worked is calculated differently for those in school during the year than for those not in school. For those who did not attend school at any time during the school year, say from time t to $t+1$, we calculate hours worked per week based on employment data between October 1 of year t and September 30 of year $t+1$.[35] For those who were attending school during the year, we base the calculation of hours worked per week on data covering the period between October 1 and June 30. Although we could have calculated summer work hours separately, the model does not accommodate separate summer work decisions and we wanted, most importantly, to account accurately for school-year weekly work hours. The discretized values of hours worked per week are obtained from the reported continuous hours as follows: $h=0$ if $h<5$, $h=10$ if $5 \leq h < 15$, $h=20$ if $15 \leq h < 25$, $h=30$ if $25 \leq h < 35$, $h=40$ if $35 \leq h < 45$, and $h=50$ if $h \geq 45$.[36] Hourly wage rates were calculated by cumulating weekly earnings over the relevant period (12 months for nonschool attendees and nine months for school attenders) and dividing by hours worked. Annual earnings are calculated by multiplying the average wage times average hours times 50 (weeks).

Descriptive Statistics

Each observation begins at first entry into the ninth grade and ends either at high school graduation (four or five years later) or the last time the youth was surveyed up to the 1991 interview. The longest period over which we observe an individual who dropped out of high school is thirteen years. Table I shows the choice distribution and the number of observations in each period. By definition, everyone in the sample is attending school in period one.[37] In grade nine 82.6 percent of the youths in our sample attend school without working (hours worked per week less that five on average), 15.3 percent attend school and work part-time (5 to 24 hours per week on average) and only 2.1 percent work full-time while attending school (25 or more hours per week on average). Working while simultaneously attending school rises rapidly with years since high school entry (and thus with age), with part-time work while attending school reaching 40 percent of all youths and full-time work while attending almost 10 percent in the third year (grade 11 with normal progress). In that year, 6 percent of youths were not attending school, with about two-thirds of them working. By the fourth year, only 30 percent of the youths are in school and not working, with almost one-fifth working full-time while attending. In addition,

[35] We actually cumulated over the first week of each month and multiplied by 13/3 to obtain monthly hours worked.

[36] Because the employment history only goes back to January 1, 1978, in computing hours worked for the period October 1, 1977–September 30, 1978, we assigned average hours worked based on the January 1, 1978–September 30, 1978 period. For those not attending school for the entire year, we used hours worked only over the period of school attendance, again because our primary concern was to accurately measure work hours while in school.

[37] A very small proportion of youths never attended high school.

TABLE I

CHOICE DISTRIBUTION BY PERIOD (PCT.)[a]

Period	*sn*	*sp*	*sf*	*nn*	*np*	*nf*	No. Obs.
1	82.6	15.3	2.1	—	—	—	564
2	65.1	29.8	3.4	.5	.5	.5	553
3	44.5	40.0	9.5	1.7	1.9	2.4	535
4	30.0	40.3	18.2	2.9	3.9	4.7	516
5	7.6	6.5	4.4	13.0	27.2	41.3	92
6	—	—	—	13.4	22.0	64.3	82
7	—	—	—	8.7	17.3	74.1	81
8	—	—	—	10.0	21.3	68.8	80
9	—	—	—	12.8	18.0	69.9	78
10	—	—	—	10.5	10.5	79.0	76
11	—	—	—	16.0	13.3	70.7	75
12	—	—	—	14.5	8.1	77.4	62
13	—	—	—	20.8	12.5	66.7	24

[a]*sn*: in school, not working; *sp* in school, working part-time; *sf*: in school, working full-time; *nn*: not in school; *np*: not in school, working part-time; *nf*: not in school, working full-time.

11.5 percent are not attending school, and of these about 40 percent are working full-time and 34 percent part-time. Of those youths who attend school for at least four years, 78 percent worked at least one of those years and over one-quarter worked at least three years.

In year four, 421 of the 516 youths who were in the sample for at least four years (81.6 percent) graduated from high school. Of the 92 nongraduates who were observed in year five, 17 (18.5 percent) attended school in year five and of those, six (35 percent) graduated in year five. Thus, about 83 percent of the sample who are observed until graduation or, otherwise, for at least five years, graduated from high school, with 98.6 percent of them graduating in four years.[38] Returning to high school after dropping out is a rare occurrence. In our sample, only three youths dropped out and returned (one graduated). However, this is an understatement of the phenomenon both because high school dropouts are under-represented in the transcript sample and because transcripts are themselves more likely to be incomplete in the case of returnees. In Keane and Wolpin (1997) it was found that 6.5 percent of a similar sample of youths from the NLSY79 returned to high school and completed an additional grade (see also Light (1995)).

If the only motive for high school attendance is to earn a diploma and if the five-year limit on attendance assumed in the model is strictly binding, then any youth who falls more than one grade level behind in terms of credit acquisition would drop out. To the extent that there is some consumption value to attending high school or some labor market payoff to having completed credits without

[38]As noted, our transcript sample undercounts high school dropouts and so the graduation rate is overstated. Even ignoring that, the 48 attriters between periods one and four are probably mostly dropouts; if they all were dropouts, the overall graduation rate would be 77 percent for our sample.

graduating, individuals may remain in school even after having fallen too far behind to graduate within the five-year period. Table II shows the relationship between attendance rates and grade level progression. First, for a given grade level completed, the proportion attending school monotonically declines with years since high school entry. However, a significant number of youths with no possibility of graduation within the five-year horizon still choose to attend school (e.g., 39 (14) percent of those who are still attending grade level eight in period three (four) do not drop out), which indicates an additional payoff to attendance beyond that to graduation (or the enforcement of mandatory school leaving ages).[39] Second, holding constant the number of grade levels that a youth has fallen behind (e.g., grade level eight in period two, grade level nine in period three, etc.), the proportion attending school declines over time (moving diagonally from the upper left), which is consistent with the existence of a finite horizon. Third, attendance among youths who have not fallen behind in grade level is almost universal.

IV. FURTHER PARAMETERIZATIONS

To estimate the model, it is necessary to adopt explicit forms for the following functions: grades, wage offers, hours offers, and the values of leisure, school attendance, and graduation. The functional forms that were adopted do not represent a priori specification, but rather are the result of an iterative specification search based on assessing the fit of the model to summary statistics in the data. That is, parameters were added as estimation proceeded in order to improve the fit of the model to elemental aspects of the data, specifically to the

TABLE II

ATTENDANCE RATE (PCT.) BY GRADE COMPLETED AND PERIOD

	Grade Level Completed			
Period	8	9	10	11
1	100 (564)[a]	—	—	—
2	84.3 (51)	99.8 (502)	—	—
3	39.3 (28)	80.0 (45)	98.7 (462)	
4	13.6 (22)	33.3 (27)	53.7 (41)	99.3 (426)
5	0.0 (19)	12.5 (24)	15.6 (32)	52.9 (17)

[a]Number of observations with completed grade in parentheses.

[39] It is also possible that such youths have a possibility of graduating in ways that our model does not accommodate, e.g., taking additional credits either during the regular school year or in the summer.

data shown in Tables I and III.[40] We were especially concerned about distinguishing between unobserved heterogeneity and other explanations for the age and transition patterns in the data. We, therefore, liberally added age variables and lagged choice variables in order that the degree of heterogeneity not be overstated.[41] After the specifications were set, we iterated until the likelihood converged.

Grade functions: Recall that the grade earned in a course is conceptualized as indicating the incremental knowledge acquired in the course. This latent variable is assumed to depend at period t on cumulative grade point average (G), cumulative credits earned as summarized by grade level completed (e), the number of years in high school (t), whether any credits have yet been earned ($I(C=0)$), on current hours worked (h) and on type.[42] Although the number of credits earned depends only on the number of courses in which a passing grade is received and graduation is contingent only on earning the requisite number of credits, because the latent variable is a function of G, it is necessary to model the entire grade distribution. Given that grades are qualitatively ordered, it is natural to model the grade function in an ordered response framework. We adopt an ordered logit for convenience. Specifically, the probability of receiving an "F" in a single course is given by

$$\pi_t^F = 1/(1+\exp(\theta^F)),$$

where

$$\theta_t^F = \sum_m \theta_{0m}^F I(\text{type}=m) + \theta_1 G_t + \theta_2 e_t + \theta_3 t + \theta_4 I(C_t = 0) + \theta_5 I(h_t \geq 500) + \theta_6 I(h_t \geq 1000) + \theta_7 I(h \geq 1500).$$

The probability of receiving a "D" in a course is given by

$$\pi_t^D = 1/(1+\exp(\theta^F + \theta^D)) - \pi_t^F,$$

where

$$\theta^D = \sum_m \theta_{0m}^D I(\text{type}=m).$$

[40]Although this method of iterating between model specification and data clearly contaminates statistical measures of model fit, it would seem that such a strategy is unavoidable given the complexity of the behavioral model. Given the limited data we have, and in particular the small proportion of high school dropouts, using only part of the data in estimation, e.g., only the first few years of high school, and assessing the model based on it's out-of-sample fit, is not feasible.

[41]We report below on the ability of the model to fit the age pattern of choices, as depicted in Table I, without resorting to the inclusion of age directly.

[42]Note that time since entry into high school is essentially the same as age; in our sample over three-fourths enter high school at age 14 and age at entry is a fixed characteristic captured by a person's type. It is necessary to include the indicator of zero credits because grade point average is undefined at $C=0$. We normalize $G=0$ at $C=0$. The indicator function also captures the fact that grading systems may differ in the ninth grade, which reflects the fact that there is often a school change between the ninth and tenth grades.

TABLE III

ACTUAL AND PREDICTED SUMMARY MEASURES OF SCHOOL ACHIEVEMENT AND EMPLOYMENT

	Actual	Predicted
Pct, high school graduates:	82.9	82.0
Pct. dist. grade completion levels less than 12:		
8	22.9	24.0
9	26.7	25.0
10	37.2	34.1
11	14.0	17.0
Pct. attending school by years since entry:		
1	100.0	100.0
2	98.4	98.0
3	94.0	95.5
4	88.6	91.1
5	18.5	22.8
Pct. dist. course credits earned per period:		
0	2.4	1.3
1	1.2	2.0
2	2.6	2.8
3	2.5	3.5
4	1.9	4.0
5	89.3	86.4
Mean GPA by grade level attending (passing grades):		
9	2.36	2.35
10	2.30	2.26
11	2.38	2.36
12	2.40	2.43
Pct. dist. wkly hours worked while attending school:		
0	58.8	54.8
10	19.6	23.4
20	13.0	15.6
30	6.5	4.4
40	1.6	1.2
50	0.5	0.6
Mean hourly wage ($):		
In school, work part-time	4.45	4.29
In school, work full-time	4.73	5.61
Not in sch., work part-time	5.50	6.09
Not in sch., work full-time	6.77	7.10

Similarly, the other grade probabilities follow the ordered logit formulation with $\theta^j = \Sigma_m \theta^j_{0m}$ for $j = C, B, A$.[43] The joint credit-grade distribution necessary for the solution of the dynamic programming problem and for the calculation of the likelihood is easily derived from the single-course grade probabilities.

Notice that if the parameters $\theta_5 = \theta_6 = \theta_7 = 0$ there is no effect of working while in school on school performance. The effect of working 10 hours per week ($h = 500$) on the log-odds of failing a course is given by $-\theta_5$, the effect of working 20 hours per week by $-(\theta_5 + \theta_6)$, and the effect of working 30 or more hours per week by $-(\theta_5 + \theta_6 + \theta_7)$.

Wage-offer functions: The part-time and full-time wage offer functions are assumed to be linear in work experience as measured by accumulated hours worked up to 50,000 hours. Wage offers may also differ by unobserved type reflecting differential market skills at the time of entry into high school. Specifically,

$$w_t^j = \sum_{m=1}^{M} \alpha^j_{0m} \cdot I(\text{type} = m) + \alpha^j_1 H_t \cdot I(H_t < 50{,}000) + \alpha^j_1 \cdot 50{,}000 \cdot I(H_t \geq 50{,}000)$$

for $j = p, f$. We assume, for parsimony, that measures of school performance in high school such as credits earned or GPA do not affect wages offered to nongraduates.

Hours-offer functions: The joint part-time and full-time hours offer function is assumed to be different during the high school eligibility years than during the post-eligibility years. We also allow age to affect hours offers during the high school years, reflecting possibly both legal constraints and employer preferences. Specifically, the hours functions are given by the multinomial logit form:

$$\pi_t^h(j,k) = \exp\left(\theta^h_{0jk} + \theta^h_{1jk} * t\right) \Big/ \sum_{j'} \sum_{k'} \exp\left(\theta^h_{0j'k'} + \theta^k_{1j'k'} * t\right) \quad \text{for} \quad t \leq 5,$$

$$\pi_t^h(j,k) = \exp\left(\theta^h_{0jk}\right) \Big/ \sum_{j'} \sum_{k'} \exp\left(\theta^h_{0j'k'}\right) \quad \text{for} \quad t \geq 6.$$

There is also assumed to be a job-finding cost if one was neither employed nor in school. A cost is borne upon choosing either work alternative while not attending school ($d_t^{np} = 1$ or $d_t^{nf} = 1$) if the nonwork (nonschool) alternative was chosen in the previous period ($d_{t-1}^{nn} = 1$). Utility in these alternatives, as given in (1), is reduced by $c^{np} \cdot I(d_{t-1}^{nn} = 1)$ if $d_t^{np} = 1$ and by $c^{nf} \cdot I(d_{t-1}^{nn} = 1)$ if $d_t^{nf} = 1$.

Value of leisure: The hourly value of leisure is assumed to differ by type. In addition, the value of leisure may differ in the high school eligibility years from

[43] By allowing for heterogeneity in all the θ^j's, the impact of working on the grade distribution can differ flexibly by type. An even more flexible form in which the determinants of θ^F also entered into the other θ^j's would add too many parameters to be computationally tractable.

the post-eligibility years, i.e.,

$$\bar{b}^n = \sum_{m=1}^{M} b_{0m}^n I(\text{type} = m) + b_1^n I(t \leq 5).$$

Net consumption value of school attendance: As seen in (1), the utility value of attending school is assumed to depend explicitly on work status:

$$\bar{b}_t^s = \bar{b}_t^{sn} \cdot I(d_t^{sn} = 1) + \left[\bar{b}_t^{sp} + \bar{b}_t^{sn}\right] \cdot I(d_t^{sp} = 1) + \left[\bar{b}_t^{sf} + \bar{b}_t^{sn}\right] \cdot I(d_t^{sf} = 1).$$

This psychic value may reflect a number of positive or negative factors including the value attached to the social aspects of attending school and the value attached to learning *per se* as well as to the effort it entails. We allow for unobserved type heterogeneity and for the value of attendance to change with the level of schooling and with time since first entry into high school. In addition, we include a cost to dropping out and returning to high school as well as a cost to attending high school in the fifth year of eligibility. Both of these reflect the fact that in either case the individual is no longer synchronized with his entry cohort. Working either part- or full-time may reduce the psychic value of attendance if it inhibits participation in social activities (e.g., extra-curricular programs) or if it implies increased effort in learning. Such an effect may depend on years since entry (age cum maturity), and on whether the individual has been able to adjust to the joint activity as measured by work participation in the previous period. Thus (letting $d_t^s = 1$ indicate school attendance in any work status),

$$\bar{b}_t^{sn} = \sum_{m=1}^{M} b_{0m}^{sn} I(\text{type} = m) + b_1^{sn} t + b_2^{sn} e + b_3^{sn} I(d_{t-1}^s = 1) + b_4^{sn} I(t = 5),$$

$$\bar{b}_t^{sp} = b_0^{sp} + b_1^{sp} t + b_2^{sp} I(t = 1) + b_3^{sp} I(d_{t-1}^{sp} = 1) + b_4^{sp} I(d_{t-1}^{sf} = 1),$$

$$\bar{b}_t^{sf} = b_0^{sf} + b_1^{sf} t + b_2^{sf} I(t = 1) + b_3^{sf} I(d_{t-1}^{sf} = 1) + b_4^{sf} I(d_{t-1}^{sp} = 1).$$

Value of high school graduation: The deterministic component of the present value of the utility of graduating from high school, which includes not only monetary but also psychic returns, is allowed to differ by type and to depend on achieved cumulative grade point average upon graduation and upon work experience gained while attending high school, namely

$$V^D = \sum_{m=1}^{M} \gamma_{0m} I(\text{type} = m) + \gamma_1 G^D \cdot I(e = 12) + \gamma_2 H^D \cdot I(e = 12).$$

In solving the optimization problem, the value attached to graduating is optimally updated as seen in (5). Because credits earned, cumulative grade point average, and hours worked evolve over time, the forecasts of cumulative GPA and hours worked at the time of graduation, G^D and H^D respectively, as well as of the likelihood of graduating, changes as decisions are made and stochastic shocks are realized.

We have made a number of accommodations in the specification of the value of graduation in order to conserve on parameters and avoid identification problems (see below). In keeping with the prior specifications, we do not allow for type-specific interaction effects, e.g., in the wage offer function the effect of work experience on wage offers might vary by type or in the course grade function the effect of working on the probability of receiving an "F" might be less likely for high ability types. In the present case, the specification rules out type-specific GPA effects (the effect of GPA might depend on whether one is a type that is likely to attend college) and type-specific work experience effects (the effect of work experience while in high school might be more valuable for high school graduates who do not attend college). To conserve on parameters, we consistently allow only for first-order heterogeneity effects.[44]

In addition, as with the rest of the model, prior random shocks to wages or preferences only affect decisions through the state variables, C, G, and H. This assumption ignores, for example, the possibility that working while in high school might be motivated by saving for college, in which case the accumulated earnings while in high school would affect the probability of attending college and thus the value of graduating from high school. However, allowing accumulated earnings to be a state variable that affects the value of graduation would require modeling savings decisions, which is beyond the scope of this paper. Rather, we assume that whatever constraints there are on borrowing that affects post-graduation decisions are captured by the type-specific constants, the γ_{0m}'s.

Identification Issues

The model contains 92 parameters. The ordered logit grade function itself contains 23 parameters (25% of the total), 16 of which are used to distinguish among the $M = 4$ ability/motivation types.[45] Given that the hours effects in the grade function are at the heart of the issue, our motivation was to maintain as much flexibility in the form of the grade function as possible, conditional on computational tractability. Of the 69 remaining parameters, 20 are associated with types, i.e., each of the 4 types is described by a vector of 5 attributes in addition to ability/motivation (the utility value of leisure, the utility value of school attendance, skill endowments associated with part-time wage offers, skill endowments associated with full-time wage offers and the value of high school graduation). Individuals are composites of these type-specific attributes and it is their combination that determines the type-specific dropout propensities.

As was shown in Table I, working while attending high school increases monotonically with age. The model provides a number of explanations for this

[44] Actually, after completing the estimation, we re-estimated the model allowing for interactions of cumulative GPA and work experience with type (see below).

[45] Four types fit the data distinctly better than did fewer types. Below, we report on the results with two types. We did not attempt to add a fifth type because doing so would have added 10 more parameters. We also did not attempt to estimate the discount factor (set at .97) because it is unlikely to be well identified separately from the value of graduation parameters.

pattern. First, to the extent that wage offers (when schooling is less than 12 years) rise with work experience, the probability of working will rise with age as work experience is accumulated. Individuals who receive a wage offer in period one large enough to induce them to work while attending school will have a higher probability of drawing a wage offer large enough to induce them to work in period two and so on. Moreover, the accumulation of work experience increases the expected present value of lifetime utility associated with being a high school dropout, but may not increase by as much, if at all, the expected present value of being a graduate.

Second, there is a "pure" age effect that arises from the graduation credit requirement (earning 20 credits in five years). Essentially, this effect arises because the propensity to work at any given age depends on the probability of graduation as viewed at that age, which in turn depends on the number of cumulative credits failed up to that age. The most straightforward case is for those individuals for whom graduation is no longer feasible (having fallen behind more than five credits). For them, it is clear that the impact of working on the probability of failing will no longer act as a disincentive to working while attending school (assuming they do attend). Similarly, for those still on track to graduate, the disincentive effect of working will decline with age for the same number of credits failed, i.e., the probability of graduating is lower in period 3 for those who have accumulated 9 credits (one credit failed) than in period 4 for those who have accumulated 14 credits (also one credit behind). Offsetting this increased graduation probability is the fact that the average total number of credits failed is increasing. On net, this pure age effect would seem to be indeterminate for the group for whom graduation is still feasible. In simplified versions of the model which incorporate no other reasons for systematic age effects, we have been able to simulate an increasing propensity with age to work while attending school.

Third, heterogeneity in preferences and endowments implies that those who drop out of high school during the five periods of eligibility will be a select sample. If those who drop out work less while attending school, then it will appear as if working while attending school is increasing in age. Finally, the model allows for direct age (maturation) effects and for lagged state dependencies; there are 13 parameters that capture age effects and 11 parameters that capture lagged state dependencies (4 of which appear in the ordered logit grade function). Not only do they significantly improve the fit of the model (see below), but they guard against overstating the degree of (unobserved) heterogeneity.

Although the model is complex it is nevertheless possible to provide intuition concerning the identifiability of the parameters. Consider first the wage offer and course grade functions. Their identification obviously rests on the wage and course grade data. A useful way to view the estimation problem with respect to each of them is as one of sample selection. We only observe wages for accepted job offers and we only observe course grades for those who choose to attend school. The solution to the optimization problem provides the sample selection rules. The functional form, distributional and exclusionary assumptions embedded in the model serve the same purpose as would a sample selection correction in either a two-step or full information procedure.

Consider next the identification of the utility function parameters. If the model were static (individuals behaved myopically), then identification of the utility function parameters would come from observing the choices and accepted wages. Now, recall that the dynamic optimization problem looks like a static multinomial choice problem with the future component of the value function treated as a known quantity based on its recursive solution (given the discount factor). What is unclear, however, is how the parameters reflecting the consumption value of schooling that enter the static problem can be distinguished from the parameters of the value of graduation function because both are directly related to the value of attending school.

The basic intuition for the identification of these two effects is easiest to establish for a homogeneous population. Suppose, as is the case in our data, that some fraction of the sample continues to attend school even though their probability of graduation is zero. For them, the only reason to attend school allowed for in the model would be that they receive a positive consumption value, which is thus identified. However, in prior periods in which the probability of graduation is not zero, choices do depend on the value attached to graduation, which is thus also identified.

Identification of these parameters is problematic if no individuals continue in school when their probability of graduation is zero. Similarly, there would be a problem when there are heterogeneous types some of whom always have a positive probability of graduating. To consider these cases, note first that the weight attached to the value of graduation in the choice-specific value functions changes systematically over time as the probability of graduation changes with accumulated credits and grade point average (in the extreme case we considered above it fell to zero at some point).[46] Now, if neither the value of graduation nor the consumption value of schooling changes over time or if they change over time in systematic but different parametric fashions, having multiple periods identifies them separately. However, they are not identified nonparametrically, i.e., if they each take on a different value at each period. In the specification we adopt, the consumption value of schooling is modeled as a linear function of age (with dummies for periods one and five) and the expected value of graduation evolves with age only as cumulative GPA and work experience changes.[47]

[46] When the probability of graduation doesn't change over time, the consumption value of schooling and the expected value of graduation cannot be separately identified. Indeed, for one of the types we estimate that the probability of graduation is close to zero from the first period and the estimate of the value that type attaches to high school graduation has a huge standard error.

[47] Age effects were included in the consumption value of schooling because it was the most direct way to fit the age trend in the school-work choice data. It does have the problem that youths are assumed to have perfect foresight about the age trend.

V. ESTIMATION RESULTS

Although the model has a large number of parameters, it is also being fit to a lot of data: the sequence of school-work hour choices over five years and the sequence of hours worked choices for nongraduates for up to an additional decade, the sequence of achieved grade point averages and credits earned during high school, and the sequence of accepted hourly wage rates for over a decade. The full set of parameters of the model (together with standard errors) are provided in Appendix A. We first present evidence about model fit before discussing some of the individual parameter estimates and the implications of the estimates for dropout behavior and for the efficacy of policy interventions.

Model Fit

Consider first the fit of the model to basic data elements. Table III compares the actual and predicted values of a number of summary measures of school achievement and employment. In the first row, it is seen that 82.9 percent of our sample graduate from high school. The model prediction, based on a simulated sample of 5,000 youths, is 82.0. As seen, the model fits the data quite closely in most other school and work dimensions as well. There are, however, several exceptions. With respect to schooling, the model overstates the fraction of nongraduates who complete eleventh grade (17.0 vs. 14.0), overstates the percent who attend school particularly in the fifth year of high school eligibility for those who have not already graduated (22.8 vs. 18.5), overstates the percent of school attendees who fail at least one course in a year (13.6 vs. 10.7) and understates the percent who fail all five courses in a year (1.3 vs. 2.4), overstates the percent who work part-time while attending high school (23.4 vs. 19.6 percent working 10 hours per week and 15.6 vs. 13.0 percent working 20 hours per week) and concomitantly understates the percent working more than 20 hours per week (6.2 vs. 8.6 percent working 30, 40, or 50 hours per week) and the percent not working (54.8 vs. 58.8 percent). The largest deviations occur with respect to the hourly wage rate; the model overstates the mean hourly wage of those employed part-time while in school ($5.61 vs. $4.73) and those working part- and full-time while not attending school ($6.09 vs. $5.50 and $7.10 vs. $6.77 respectively). Overall, the differences do not appear especially large.

The model does less well in fitting the age pattern of choices (Table I), even though, as we have discussed, the model allows for a number of possible avenues for generating age effects (including a number of parameters directly related to age). Out of the five high school eligibility periods, a contingency table chi-square test rejects that the model fits the school-work choices in periods 1, 2, and 4. The model overstates the extent to which individuals work while attending school in periods 1 (25% vs. 17%) and 2 (43% vs. 32%) and understates it in period 4 (51% vs. 59%).[48] Moreover, a generalized chi-square test that accounts for the fact that there are estimated parameters (Andrews (1988)) rejects the fit for periods 3 and 5 as well even though in period 3 the model overstates working while attending school by only 3 percentage points and in period 5 by only 4 percentage points.

Given the extensiveness of the data, there are a large number of ways of looking at model fit, at least as many ways as there n-way tables that can be constructed relating choices, wages, credits, and grades to each other and over time. As a way of testing the model's fit to higher order cross-classifications of the data, Table IV looks at how the model replicates two multi-variate relationships, the propensity to attend school and the propensity to fall behind in course credits (to fail at least one course in a period). The propensity to attend school (or obversely, not to dropout) at any period is simply estimated as a logistic function of the state variables, which is what a researcher might have done without specifying the structure we have imposed. The propensity of failing a course in any period is given by the same determinants as in the structural ordered logit grade function. Note, however, that neither of these relationships is consistently estimated given the assumptions of the behavioral model; they do not account for heterogeneity or sample selection. Our purpose is not to draw inferences about behavior or technology, but simply to assess whether some important aspects of the covariance structure generated by the model are similar to that in the data. The first column (labeled actual) shows the relationships obtained from the data and the second column (labeled predicted) those obtained from the simulated sample based on the estimated behavioral model. The numbers reported show the changes in the odds-ratios (the probability of the choice divided by one minus the probability of the choice) due to a unit change in the state variable; a coefficient over one indicates an increase in the odds-ratio and a coefficient less than one a reduction. A 95% confidence interval for each odds-ratio is shown in parentheses.[49]

[48]A likelihood ratio test strongly rejects a specification that excludes the 13 specific age parameters that appear in the model; the ln likelihood is −9,288 without age effects vs. −9,166 with the age effects. With no direct age effects, the proportion predicted to attend school without working falls from 54.3% to 44.5% between period 1 and 2 and then to 43.0% in period 3, a considerably smaller decline than in the data (82.6% to 65.1% to 44.5%, Table I) or in the model that allows for age effects (74.4% to 54.2% to 42.6%). Thus, although the model without age effects does generate some of the decline observed in the data, the restrictions imposed on the estimates by having to fit other aspects of the data limit its ability to do so. Direct age (maturation) effects are important in explaining the age pattern of choices.

[49]The reported confidence interval for the simulated data arises solely from the simulation error associated with a finite sample. The additional error that arises from the fact that the simulated data are based on estimated parameters is not included. It would be possible to bootstrap confidence intervals incorporating estimation error from the estimated variance-covariance matrix of the parameters. Given the rank of the variance-covariance matrix and the fact that the hypothesis that the actual data and the simulated data are the same is rejected at almost any level of significance, we did not pursue this computationally intensive procedure.

TABLE IV

ACTUAL AND PREDICTED SCHOOL ATTENDANCE AND COURSE CREDIT LOGISTIC REGRESSIONS

	Attended School ($I(d_t^s = 1)$)		Failed One or More Courses ($I(c < 5)$)	
	Actual	Predicted	Actual	Predicted
Grade comp. (e)	6.23[a] (4.00, 9.71)[b]	5.41 (4.74, 6.16)	.008 (.003, .061)	.007 (.005, .009)
Yrs. since entry (t)	.112 (.067, .189)	.102 (.087, .120)	51.1 (22.5, 117.8)	45.3 (34.9, 58.8)
Cum. GPA (G)	2.00 (1.22, 3.29)	1.46 (1.28, 1.67)	.236 (.157, .355)	.175 (.147, .208)
$I(C > 0)$	.945 (.076, 11.7)	1.53 (.437, 5.35)	.033 (.012, .093)	.020 (.013, .031)
Cum. Hrs. Wkd. (H)	.9997 (.9995, .9999)	1.0002 (1.0000, 1.0003)	—	—
In Schl. Prior Pd. ($I(d_{t-1}^s = 1)$)	24.75 (5.94, 103.2)	3.39 (2.50, 4.61)	—	—
Wkd. 10 Hrs/Week	—	—	1.32 (.799, 2.17)	.949 (.813, 1.11)
Wkd. 20 Hrs/Week	—	—	1.09 (.512, 2.30)	1.02 (.849, 1.23)
Wkd. 30 or More Hrs./Week	—	—	2.78 (1.11, 6.98)	1.25 (.923, 1.71)
Pseudo R^2	.639	.560	.432	0.510
H_0: pred. = actual	reject		reject	

[a] Coefficients are percent changes in the odds-ratio divided by 100 plus one for a unit change in the variable, i.e., $\partial(p/1-p) = \beta \cdot (p/1-p)$.

[b] 95% confidence interval in parentheses.

For both logistic regressions, the parameter estimates obtained from the simulated data are in most cases qualitatively the same as those based on the actual data. The main quantitative difference in the attendance relationship is that it exhibits considerably less state dependence in the simulated data than in the actual data; having attended school in the previous period is, in the actual data, associated with about a 25-fold increase in the odds-ratio of choosing the same alternative the next period, but only about a 3-fold increase in the simulated data. (Recall that the behavioral model directly includes lagged state dependence of this form.) In the course credit relationship, the main difference is that the "effect" of working 30 hours per week or more on the probability of failing a course is much greater in the actual data than in the simulated data. In almost all cases, the point estimates from the simulated data fall within a 95% confidence of the estimates from the actual data. However, chi-square tests of whether the estimates from the simulated data are jointly the same as those from the actual data reject that hypothesis at any reasonable level of significance.

Another way to assess the validity of the model would be to consider whether particular parameter estimates conform to priors. Such an evaluation would require that there exist objective priors about which there was some consensus. However, there is only one study of which we are aware that has estimated comparable parameters, an analysis of schooling and occupational choice decisions by Keane and Wolpin (1997) using a superset of the NLSY79 observations used here. Their model differs from the one here in that they model the schooling decision more crudely (for example, they do not consider the possibility of failure), but do not limit school attendance decisions only to the high school period, and they follow all individuals (not just high school dropouts) into the labor market. In addition, they model occupational choice in the labor market, but only consider the full-time work decision. As in this model, Keane and Wolpin (KW) allow for preference and endowment heterogeneity.

A number of parameters can be compared: the consumption value of schooling, the value of home production, and the present value of high school graduation. Because it is difficult to match unobservable types across the two studies, we compare estimates that are averaged over types. In addition, because the studies use different discount factors, which in the current model acts to scale parameters, we have re-estimated the model using the discount factor in KW (.936). The consumption value of schooling provides the cleanest comparison.[50] The estimates are reasonably close, $9,638 in KW vs. our estimate of $10,997. Comparing the value of home production is complicated by the fact that the estimate in KW is an annual amount while we estimate the hourly value of leisure. KW's estimate is $13,381 per year while we estimate the hourly value to be $3.27. The two are equated at 4092 annual hours of home time (11 hours per day), which is not too unreasonable an estimate of the total stock of time (net of sleep and personal hygiene).[51] Finally, KW estimate the present discounted value of lifetime utility (as of age 26) to be $445,277, while we estimate it to be $595,979.[52] The proximity of these estimates is perhaps the most surprising result of the comparison given that our estimate, unlike KW, does not make use of post-graduation data on earnings.

[50] The figures are obtained from Tables 8 and 9 in KW. Because KW's model does not allow for working while attending school, the comparison is based on the estimates of the consumption value of schooling while not working, that is the re-estimated b^{sn}'s.

[51] One difference in the estimates is that KW find that those types with high values of leisure are also those with high consumption values of schooling, while we find the opposite.

[52] In deriving the estimate from their tables, we used only the individuals who had initial schooling levels of 10 or more in order that the group correspond more closely to those who would ultimately receive high school diplomas. In keeping with that restriction, the estimate from our model is based only on the expected value of graduation for types 2 and 4. G was set to 3.0 and H to 2000 in our estimate, but their effects on the value of graduation are too small to make any real difference. An implicit assumption in KW is that the total amount of time available (L in our model) to be split between work and leisure is equal to the amount of hours spent in full-time work. Note that L scales the present value of utility by the value of leisure times L. In calculating the average value of graduation for our model, we set $L = 2500$ to correspond to the KW assumption.

Dropping Out

Who drops out of high school? Table V presents completed schooling levels by type. It is clear from this table that dropping out is confined to two types, type 1 and type 3. Type 1's comprise about 25 percent of all dropouts and type 3's the rest. Although all of the youths in these two groups are dropouts, the two types differ considerably in their completed schooling levels; type 3's are spread over the entire range of grade levels, while the vast majority of type 1's do not complete the ninth grade and almost none complete the tenth. Type 3's also have a higher GPA in the courses that they pass than type 1's, 1.75 vs. 1.53.

Interestingly, there is essentially no type that consists of both graduates and dropouts. Type 2's and type 4's all graduate, with the latter comprising 6.6 percent of all graduates. However, the two types differ considerably in their GPA's. Type 4's are essentially straight A students (GPA = 3.99) while the type 2's are C+ students on average (GPA = 2.39).[53]

That type 4's have the highest GPA is the result of their type-specific values of the θ_0's (see Table A.1), i.e, type 4's have the highest levels of whatever "permanent" traits are associated with school performance, whether ability, motivation or initial (at the time of high school entry) knowledge.[54] In Table VI, the ranking of types by all of the unobserved heterogeneity parameters is shown.[55] Type 4's rank first (or essentially tied for first) in all of the parameters that are schooling related, namely the consumption value of attending school,

TABLE V
PERCENT DISTRIBUTION OF GRADE COMPLETION LEVELS BY TYPE

Grade Level	Type 1	Type 2	Type 3	Type 4
8	82.5	0	4.1	0
9	16.7	0	27.6	0
10	0.9	0	45.1	0
11	0	0	22.6	0
12	0	100.0	0.6	100.0
Cumulative terminal GPA	1.53	2.39	1.75	3.99[a]
Percent of sample	4.6	76.5	13.5	5.4

[a] 3.6% of the sample reported a terminal GPA of 4.0.

[53] In the data, 6.0 percent of the youths who graduate have 4.0 averages (actually 3.75–4.0 due to the discretization).

[54] The difference in the coefficients that determines the probability of failure (types 1 and 3 vs. either type 2 or 4) are statistically significant (see Table A.1). However, the model does not well identify the cut-off values of the ordered logit that determine the probabilities of the specific passing grades. Recall that in the estimation, we do not make use of specific course grades, but only GPA's.

[55] See Table A.1 for the actual values of the heterogeneity parameters upon which Table VI and the following discussion are based.

TABLE VI

RANK-ORDER OF TYPES BY "PERMANENT" UNOBSERVED TRAITS

	Type 1	Type 2	Type 3	Type 4
Full-time wage	2	3	1	4
Part-time wage	4	3	1	2
Consumption value of leisure	2	3	1	3
Consumption value of school	4	1	3	1
Perceived value of h.s. graduation	4	1	3	1
Schooling ability/motivation[a]	4	2	3	1

[a]Refers to GPA.

the perceived value of graduation, and schooling ability/motivation.[56] Indeed, the rankings of these heterogeneity components follow exactly the rankings by GPA shown in Table V.

Although type 4's are clearly "school-types," they rank last in "permanent" traits that accompany high full-time wage offers for non-high-school graduates. However, the relationship between school-related heterogeneity parameters and full-time work heterogeneity parameters is not perfectly inverted. Type 1's do not have the highest "endowment" of market skills in jobs performed by non-high-school graduates, ranking second to type 3's. The dispersion in type-specific full-time wage offers is large. On average, type 4's receive full-time wage offers (with zero work experience) of $2.59 per hour, type 2's $3.65 per hour, type 1's $4.19, and type 3's $5.29 per hour, a difference of $2.70 per hour from lowest to highest.[57] The rank-order of part-time wages is different, but the dispersion is small. On average, type 4's receive part-time wage offers (with zero work experience) of $4.52 per hour, type 2's $4.27 per hour, type 1's $3.71, and type 3's $4.63 per hour, a difference of less than $1.00 per hour.[58] Part-time wage offers exceed full-time wage offers with zero work experience for types 2 and 4, the non-dropouts, and differ very little for types 1 and 3, the dropouts. However, as is probably consistent with one's priors, full-time wage offers grow with additional work experience considerably faster than do part-time wage offers, about three times the rate.

[56]Although the magnitude of the difference in the expected value of graduation of type 2's vs. type 4's is very large, the standard error of the difference is several orders of magnitude greater. The difference between type 3's and type 4's is statistically significant at the ten percent level. The difference in the consumption value of school attendance of type 1's and 3's vs. type 4's is statistically significant at the five percent level.

[57]The difference in type 1's vs. type 4's full-time wage offer is statistically significant at the ten percent level and that between type 3's and type 4's at the five percent level.

[58]None of the differences are statistically significant.

The value of graduation ranges from a little over 1 million to about 1.7 million dollars.[59] However, as noted, these values include the value of leisure times total available hours, L, and are thus sensitive to the choice of L. But, what is relevant for decision-making is the difference between the value of graduation and the expected present value of lifetime utility if the individual does not graduate. Evaluating the latter at period 6, this differential is negative for both nongraduate types; for type 1's, it is −\$176,908 while for type 3's it is −\$35,723. For the graduates, the differential is \$577,167 for type 2's and \$798,928 for type 4's.[60]

Family Background and Types

Types have been treated as unobserved. However, given the importance of type in determining a youth's propensity to drop out, it would be useful to try and relate types to measured family background characteristics that are given at the time of high school entry. Note that each of the individuals in the sample can be assigned a set of type probabilities by applying Bayes' rule to that individual's contribution to the likelihood function (8). Family background information can then be merged with the outcome data used in the estimation and related to type probabilities.

Table VII shows the mean probabilities of the four types by selected family background characteristics. Although most of the relationships between type propensities and family background characteristics are unsurprising, there are a few, perhaps unexpected, correlations. As might be expected, type 1's are significantly over-represented among those youths with the least educated parents, those who were not living with either natural parent at age 14 or only with the youth' s natural mother, those with 4 or more siblings, and those living in a family with income less than one-half the median of the sample. However, type

[59] To check the robustness of these estimates to the linearity of the terminal value function, we reestimated the model allowing for a set of interactions between types and the state variables, G and H. To conserve on parameters, we assumed that the effect of work experience on the value of graduation is the same for the two dropout types, types 1 and 3, and zero for type 4's. In addition, the effect of cumulative GPA on the value of graduation was assumed to be zero for the dropout types, but different for the two graduation types, 2 and 4. We found that an extra 2000 hours of work experience gained while attending high school increases the value of graduation by about \$23,000 for types 1 and 3 and by only \$1,000 for type 2's. In addition, an extra point of cumulative GPA increases the value of graduation by about \$26,000 for the type 2's and by \$13,000 for the type 4's. These are all rather small effects given that the value of graduation is over \$1 million for all types. As these magnitudes suggest, the impact of allowing for these interactions on the rest of the parameters and on the ln likelihood value is inconsequential.

[60] Because we do not observe graduate types after period 5, we calculated the value of lifetime utility if not a graduate assuming the maximum work experience, $H = 50{,}000$, in the steady state. This understates somewhat the differential value of graduating. The calculations for the nongraduates done in the same way would yield differentials of −\$518,865 for type 1's and −\$472,369 for type 3's.

TABLE VII
THE RELATIONSHIP OF TYPE TO SELECTED FAMILY BACKGROUND CHARACTERISTICS[a]

	Type 1	Type 2	Type 3	Type 4
Sample (Pct.); (543)	5.2	77.0	13.4	4.5
Mother's Schooling				
All non-missing (524)	4.8	78.2	12.4	4.6
Non-HS Grad (111)	10.9	62.6	23.9	2.7
HS Grad (284)	3.9	83.0	10.1	2.9
Some College (70)	2.6	77.6	8.9	10.9
College Grad (59)	0.0	84.9	6.5	8.6
Father's Schooling				
All non-missing (514)	4.3	78.6	12.4	4.7
Non-HS Grad (132)	8.9	70.8	16.9	3.4
HS Grad (202)	3.6	77.6	15.5	3.2
Some College (69)	2.9	83.7	10.2	3.2
College Grad (111)	0.0	86.7	2.7	9.7
HH Structure at Age 14				
All non-missing (556)	4.7	77.4	13.4	4.5
Live with Mother Only (70)	8.4	65.5	18.8	7.3
Live with Father Only (24)	4.2	70.8	20.9	4.1
Live with Both Parents (451)	4.0	79.9	11.8	4.2
Live with Neither Parent (11)	8.8	65.6	25.5	0.0
Number of Siblings				
All non-missing (556)	4.7	77.4	13.4	4.5
0 (23)	8.1	76.9	6.8	8.2
1 (121)	4.0	80.1	11.1	4.8
2 (145)	2.9	82.7	9.7	4.6
3 (128)	2.7	74.5	19.9	3.0
4 + (139)	8.4	72.3	14.3	5.0
Region at age 14				
All non-missing (550)	4.7	77.2	13.5	4.6
South (138)	4.9	74.1	17.1	4.0
Non-south (412)	4.6	78.2	12.3	4.8
Parental Income: 1978				
All non-missing (465)	4.5	76.4	14.4	4.6
$Y <= \frac{1}{2}$ Median (71)	15.0	61.6	19.3	4.2
$\frac{1}{2}$ Median $< Y <=$ Median (170)	4.1	73.2	18.4	4.2
Median $<= Y < 2$ Median (196)	1.8	83.7	10.7	3.8
$Y >= 2$ Median (28)	0.0	81.7	4.5	13.7

[a]Numbers in parentheses are sample sizes.

1's are also over-represented among youths who had no siblings. Again as expected, type 4's are significantly over-represented among youths whose mother's are college educated, whose father's are college graduates, and whose family income is above twice the median. However, they are also over-represented among youths who lived only with their natural mother at age 14 and who had no siblings. Thus, youths who have no siblings are more likely to be at both extremes with respect to school-related permanent heterogeneity traits.

An important question is whether modeling heterogeneity nonparametrically in fact captures most of the heterogeneity in measured family background characteristics (family income in 1978, mother's schooling, family structure at age, southern residence at age 14, city residence at age 14, and number of siblings) that might be relevant for school-work decisions. Table VIII presents regressions of three school outcomes, first only on family background characteristics and then with type probabilities as additional regressors. The first two school outcomes, highest grade completed through twelfth grade and average GPA in high school are within-sample outcomes used in the estimation of the model. The third school outcome is an out-of-sample forecast of the amount of schooling obtained by age 27 for those who graduated from high school. In each case, we test whether family background variables are jointly statistically significant.[61] For the two within-sample school outcomes, we find that family back-

TABLE VIII

REGRESSIONS OF SCHOOL OUTCOMES ON FAMILY BACKGROUND CHARACTERISTICS AND TYPE PROBABILITIES[a]

	Highest Grade Comp. Thru HS		Highest Grade Point Average[b]		Completed Schooling: Age 27	
	(1)	(2)	(1)	(2)	(1)	(2)
Family Income in 1978/1000	.013 (.0046)	−.00035 (.00135)	.00093 (.00068)	.0010 (.0062)	.00069 (.0012)	.00049 (.0010)
Mother's Highest Grade Completed	.046 (.027)	.012 (.011)	.091 (.031)	.055 (.030)	.222 (.069)	.236 (.060)
Lived in Nuclear Family at Age 14	.173 (.143)	−.094 (.053)	.103 (.144)	.039 (114)	.101 (.376)	.064 (.351)
Lived in South at Age 14	−.185 (.129)	−.047 (.040)	.437 (.280)	.515 (.272)	.269 (.317)	.291 (.302)
Lived in City at Age 14	−.325 (.109)	−.010 (.036)	−.161 (.177)	.068 (.161)	.030 (.273)	.096 (.260)
Number of Siblings	−.075 (.032)	−.0066 (.010)	−.022 (.041)	.0052 (.039)	−.124 (.087)	−.128 (.082)
Age at High School Entry	−.332 (.143)	−.049 (.044)	−.176 (.146)	−.054 (.130)	.098 (.213)	.102 (.203)
Probability Type 2	—	3.74 (.191)	—	1.75 (.225)	—	3.58 (.432)
Probability Type 3	—	1.68 (.212)	—	.769 (.233)	—	—
Probability Type 4	—	3.68 (.204)	—	3.61 (.368)	—	5.06 (.483)
Constant	15.7 (2.07)	8.90 (.716)	3.59 (2.07)	.661 (1.93)	11.2 (3.17)	7.25 (3.09)
R-Squared	.135	.900	.048	.196	.124	.231
H_0: No Family Background Effects	reject	not reject	reject	not reject	not reject	not reject

[a] Robust standard errors in parentheses.
[b] Includes course grades of "F".

[61] All hypothesis tests are based on a 5% significance level.

ground characteristics are jointly significant when type probabilities are excluded, but jointly insignificant when they are included. Perhaps more surprising is that for the out-of-sample school outcome variable those characteristics are jointly insignificant regardless of the inclusion of type probabilities. In all cases, type probabilities are themselves jointly significant.[62] Evidently, the finite-mixture of four types incorporates the predictive information about school outcomes that is contained in these family background characteristics. Further, in other regressions, we found that these family background characteristics actually explain only a small proportion, less than 10%, of the variation in any of the type probabilities. Thus, there are clearly unmeasured traits critical for predicting high school performance that are not captured by the family background variables in the NLSY.

Not only are these unmeasured traits related to high school completion, but they also are related to criminal activity (presumably captured as an element of the value of leisure parameter). Table IX presents regressions of three measures of criminal activity on family background characteristics and type probabilities: whether or not the youth shoplifted more than once in the year preceding the 1981 interview, whether the youth ever sold drugs during that period and whether they obtained at least one-quarter of their income in that period from criminal activities. In all cases, the higher the probability that the youth was a type 2 or 4, i.e., types with high probabilities of graduating from high school, the lower was their propensity to engage in criminal activity. On the other hand, family background characteristics were jointly insignificant in predicting shoplifting behavior and only living in a city at age 14 in the case of selling drugs and living in the south at age 14 in the case of criminal income were statistically significantly related to crime.[63]

The Causes of Dropping Out

Why do youths drop out of high school? The model provides for a number of possible reasons for a youth to drop out of high school: disliking school, placing a high value on leisure, having low ability/motivation, facing good market opportunities, and having low expectations of the payoff to graduation. It is difficult to quantify their relative contributions because the model is highly nonlinear and because there is no obvious metric to judge the relative sizes of alternative exogenous changes that would induce dropout behavior. For both reasons, we adopt as a metric the baseline characteristics of type 2 youths, the

[62] These results are robust to whether we use a tobit specification to account for truncation or an ordered logit specification.

[63] Estimated type probabilities are either less than .001 or greater than .999 in over 85% of the cases. Thus, youths are identified with a single type with a very high probability. In that case, using type probabilities as regressors is almost the same as using dummy variables for whether the youth was a high school dropout. Therefore, it may not be so surprising that these type probabilities dominate family background factors in explaining school outcomes. One might not have anticipated that they also dominate in explaining criminal behavior.

TABLE IX

LOGIT REGRESSIONS OF CRIMINAL BEHAVIORS ON FAMILY BACKGROUND CHARACTERISTICS AND TYPE PROBABILITIES[a]

	Shoplifted More Than Once in Past Year	Sold Drugs in Past Year	At Least $\frac{1}{4}$ of Income in Past Year From Crime
Family Income in 1978/1000	.014	.011	.017
	(.011)	(.013)	(.019)
Mother's Highest Grade Completed	−.010	.009	−.184
	(.067)	(.076)	(.114)
Lived in Nuclear Family at Age 14	−.440	−.430	−.407
	(.292)	(.349)	(.512)
Lived in South at Age 14	−.982	−.470	−1.77
	(.352)	(.380)	(.728)
Lived in City at Age 14	.373	1.34	−.113
	(.295)	(.513)	(.540)
Number of Siblings	−.068	.012	−.073
	(.075)	(.074)	(.110)
Probability Type 2 or 4	−1.15	−1.61	−2.45
	(.534)	(.572)	(.687)
Probability Type 3	−.434	.049	−.967
	(.592)	(.610)	(.719)
Constant	−.0045	−1.62	1.95
	(1.00)	(1.20)	(1.69)
(Pseudo) R-Squared	.054	.126	.142
Sample probability	.228	.140	.067
H_0: No Family Background Effects	reject[b]	reject[c]	not reject

[a] Robust standard errors in parentheses.
[b] Not reject if exclude "lived in south at age 14" variable.
[c] Not reject if exclude "lived in city at age 14" variable.

modal high school graduation type. To provide a quantitative assessment of the significance of different factors contributing to dropout behavior, we impose on the nongraduate types, 1 and 3, the heterogeneity parameters of type 2's and calculate the change in their performance. Table X produces the results of these experiments.

As is evident from the table, adopting any single trait of type 2's would have no effect on the high school graduation rate of type 1's. In addition, only one trait when adopted, school ability/motivation, would substantially alter completed schooling levels, from 8.2 years to 10.6 years. The story is a bit different for type 3's. Adopting type 2's full-time wage offer endowment (a reduction in the wage offer by about 40 percent throughout the life cycle) or type 2's perceived value of graduation (an increase of $227,871 or 15 percent) increases the high school graduation rate from .6 to 17 percent. Alternatively, adopting type 2's value of leisure (a reduction from $17.53 per hour to $5.70 per hour) increases type 3's graduation rate to 10.2 percent while adopting type 2's consumption value of school attendance (an increase of $12,322, from −$619 to

TABLE X

EFFECT HETEROGENEITY IN TRAITS ON SCHOOL COMPLETION LEVELS

	Type 1		Type 3	
	Percent H.S. Graduates	Ave. Schooling Nongraduates	Percent H.S. Graduates	Ave. Schooling Nongraduates
Baseline	0.0	8.2	0.6	9.9
If type 1, 3 had type 2 trait:				
Full-time wage offer	0.0	8.1	17.1	9.8
Part-time wage offer	0.0	8.2	0.7	9.9
Consumption value of leisure	0.0	8.1	10.2	9.7
Consumption value of school	0.0	8.5	6.4	10.6
Perceived value of graduation	0.0	8.2	17.0	10.0
School ability or motivation	0.0	10.6	11.8	10.7
Val. graduation + abil/motiv.	100.0	—	100.0	—

$11,703 per year in the case that they don't work while attending school) increases the graduation rate to 6.4 percent.

Perhaps, most interesting is the effect of adopting type 2's school ability/motivation (a fall in the probability of failing at least one course from 64 percent to less than one percent, in the case of attending grade 9 and not working). With this change, the graduation rate of type 3's increases to 11.8 percent, which indicates that the dropout behavior of type 3's is not due only to their low ability/motivation. If type 3's made the same schooling-work choices as type 2's, their graduation rate would be the same as for type 2's, 100 percent. As the last row of the table demonstrates, if in addition to having the same ability/motivation as type 2's, type 3's also had the same expected value of graduation, the graduation rate of type 3's would in fact replicate that of type 2's even though they still had other different permanent traits. Of course, the reason type 3's have a low expected value of graduation may be because they have low ability and/or motivation.

Work and Performance in High School

The joint hypothesis that hours of work does not adversely affect course grades is rejected by the model.[64] It is therefore possible that a policy designed to reduce hours worked while attending high school will increase graduation

[64] A Wald test of the hypothesis that $\theta_3 = \theta_4 = \theta_5 = 0$ yields a chi-square value of 19.5. The critical value with 3 degrees of freedom is 11.3 at the one percent level.

rates and GPA's. However, policies that limit hours of work may also affect the decision to attend school. We therefore simulate the impact of four policies that increasingly constrain work-school choices, first on school attendance rates and then on school performance measures.

The first constraint (C1) does not permit youths to work and attend school simultaneously. The second (C2) is more constraining, not permitting youths to work during any of the first four years after entering high school regardless of their school attendance. The next constraint (C3), still more restrictive, not only does not permit youths to work during the first four years after entry but also forces them to attend school, i.e., either extends school leaving ages or perfectly enforces existing ones. The final constraint (C4) is the same as the last but extends the constraint to the fifth year after entry for those who did not graduate in four years.

Table XI shows the effects of these constraints on the school attendance decisions of the nongraduating types. The baseline case is given in the first column for each type. Notice that with both of the first two constraints attendance rates fall relative to the baseline for both types. Faced with these constraints, youths with these traits would prefer either the working or leisure alternative in the case of the first constraint or the leisure alternative in the case of the second, relative to attending school. The third constraint forces attendance to be 100 percent in the first four years. However, attendance drops to only 11 percent in period five for type 1's and 20 percent for type 3's. The fourth constraint imposes attendance for all five years.

Table XII shows the effects of these experiments on graduation rates, school completion levels, and GPA (exclusive of failures) for each of the four types. The baseline is given in the first row of the table. A quick perusal of the table reveals that all of the constraints have only a trivial effect on GPA's for all types. We, therefore, concentrate on their effects on completed schooling levels, which further restricts attention to types 1 and 3.

TABLE XI

THE EFFECTS OF EMPLOYMENT AND SCHOOL LEAVING CONSTRAINTS ON SCHOOL ATTENDANCE RATES (PCT.) BY PERIOD[a]

Period	Type 1					Type 3				
	Base	C1	C2	C3	C4	Base	C1	C2	C3	C4
1	100	100	100	100	100	100	100	100	100	100
2	80.7	54.8	59.9	100	100	91.4	61.0	74.6	100	100
3	55.3	30.3	31.1	100	100	82.0	40.6	55.1	100	100
4	31.1	11.4	16.7	100	100	57.6	19.4	36.9	100	100
5	10.5	3.5	6.6	11.0	100	18.2	3.1	13.9	19.9	100

[a] Base is the baseline case. C1 imposes the constraint that working while attending high school is not permitted. C2 imposes the constraint that working during the first four years of high school eligibility is not permitted, regardless of school attendance. C3 imposes the constraints that (i) working during the first four years of high school eligibility is not permitted and (ii) school attendance during those years is mandatory. C4 imposes the constraints that (i) working is not permitted until either high school graduation or the five years of high school eligibility is exhausted and (ii) school attendance during those years is mandatory.

TABLE XII

THE EFFECTS OF EMPLOYMENT AND SCHOOL LEAVING CONSTRAINTS ON SCHOOL COMPLETION LEVELS AND GRADES BY TYPE[a]

	Type 1			Type 2			Type 3			Type 4		
	Pct. Grad	Ave. Sch. < 12	GPA	Pct. Grad	Ave. Sch. < 12	GPA	Pct. Grad	Ave. Sch. < 12	GPA	Pct. Grad	Ave. Sch. < 12	GPA
Base	0.0	8.2	1.53	100	—	2.39	0.6	9.9	1.75	100	—	3.99
C1	0.0	8.1	1.52	100	—	2.44	0.4	9.1	1.83	100	—	3.99
C2	0.0	8.1	1.52	100	—	2.44	1.3	9.5	1.75	100	—	3.99
C3	0.0	8.4	1.53	100	—	2.44	3.1	10.3	1.75	100	—	3.99
C4	0.0	8.6	1.55	100	—	2.44	17.7	10.7	1.75	100	—	3.99

[a]Base is the baseline case. C1 imposes the constraint that working while attending high school is not permitted. C2 imposes the constraint that working during the first four years of high school eligibility is not permitted, regardless of school attendance. C3 imposes the constraints that (i) working during the first four years of high school eligibility is not permitted and (ii) school attendance during those years is mandatory. C4 imposes the constraints that (i) working is not permitted until either high school graduation or the five years of high school eligibility is exhausted and (ii) school attendance during those years is mandatory.

Imposing C1, as was seen in Table XI, reduces school attendance of both types 1 and 3, which leads to a slight increase in the dropout rate and a reduction in the average schooling attainment of dropouts by .8 years. Constraint C2 also reduced attendance, but somewhat less than the C1 constraint. However, in this case establishing a strict no-work constraint for four years, slightly increases graduation rates while reducing the average schooling of dropouts by .4 years. Requiring four years of high school attendance without working leads to a modest increase in graduation rates among type 3's (and no increase among type 1's) and an increase in average schooling among dropouts of about one-half year. Finally, requiring five years of attendance, for those who do not graduate in four years, without working provides the maximum increase in graduation rates that are feasible given the traits of the two types. Even here, no type 1's graduate and only 17.7 percent of the type 3's graduate. However, the average schooling levels of dropouts increase by close to one year.[65]

[65]We performed one further check of model specification, namely we re-estimated the model assuming only two rather than four types. As might be expected, the estimates now revealed one distinct dropout type and one distinct graduation type, each of which was a combination of the two types that made up the previous dropout types (1 and 3) and graduation types (2 and 4). The fit of the model was considerably worse. In particular, whereas both in the actual data and in the simulated data based on four types, over 98% of those who graduated from high school did so in four years, in the simulated data based on only two types, the comparable figure was less than 80%. In addition, the ln likelihood value fell to −9,445 (from −9,166) with only 20 fewer parameters, although applying a strict likelihood ratio test is not appropriate. In terms of substantive conclusions, we found that as opposed to the small, but discernible, effect of working on grade point average when there are four types, there was no effect at all of working on grade point average when there are only two types.

VI. CONCLUSIONS

In this paper, we formulated and empirically implemented a model of grade progression through high school in which youths make sequential decisions about school attendance and work. The model was estimated on longitudinal data that included information about school attendance, hours worked, wages, and course grades. The estimates of the model were used to quantify the importance of alternative reasons for dropping out of high school and to assess the effect of employment restrictions on graduation rates and grades.

The results can be summarized as follows: (i) Youths who drop out of high school have different traits than those who graduate—they have lower school ability and/or motivation, they have lower expectations about the rewards from graduation, they have a comparative advantage at jobs that are done by nongraduates, and they place a higher value on leisure and have a lower consumption value of school attendance. (ii) Without altering their ability/motivation, if dropouts were forced to remain in school for five years after entry (only four years if they graduate in that time) without working, then their graduation rate would increase only to 13 percent. Legislation prohibiting work would have very little impact on graduation rates. In addition, such legislation would have only a negligible impact on the GPA of high school graduates. (iii) Increasing the ability/motivation of dropouts to coincide with that of the modal type high school graduate would by itself increase the graduation rate to only 9 percent. Even with augmented ability/motivation, there is still a strong incentive, given their other traits, to drop out before graduating. If in addition to augmenting ability/motivation, their expected valuation of graduation was also made to coincide with the modal type of high school graduate, their graduation rate would be 100 percent.[66] (iv) Policies that do not alter traits with which youths come to high school will have very limited success in improving school outcomes.[67]

Dept. of Economics, Tel Aviv University, Tel Aviv, Israel 69978
and
Dept. of Economics, University of Pennsylvania, 3718 Locust Walk, Philadelphia, PA 19104, U.S.A.

Manuscript received August, 1997; final revision received November, 1998.

[66] It is of course possible that policies that succeeded in augmenting either one of these might affect the other. We have not modeled the behavioral process determining the levels of these attributes at the time of high school entry. This is obviously the next step in improving our understanding of school success.

[67] A number of other recent studies of school attainment also have demonstrated the importance of "endowments" that are themselves related to family background characteristics, e.g., Cameron and Heckman (1999) and Keane and Wolpin (1997).

APPENDIX

TABLE A.1

MODEL ESTIMATES

	Parameter	Asymptotic Standard Error
Full-time wage function		
α_{04}^{F}	.694	.305
$\alpha_{01}^{F} - \alpha_{04}^{F}$	.482	.296
$\alpha_{02}^{F} - \alpha_{04}^{F}$	.344	.290
$\alpha_{03}^{F} - \alpha_{04}^{F}$	.714	.300
α_{1}^{F}	.336 E − 04	.111 E − 05
Part-time wage function		
α_{04}^{P}	1.21	.232
$\alpha_{01}^{P} - \alpha_{04}^{P}$	−.197	.240
$\alpha_{02}^{P} - \alpha_{04}^{P}$	−.560 E − 01	.211
$\alpha_{03}^{P} - \alpha_{04}^{P}$	.243 E − 01	.233
α_{1}^{P}	.119 E − 04	.144 E − 04
Course grade function		
θ_{04}^{F}	2.40	1.59
$\theta_{01}^{F} - \theta_{04}^{F}$	−6.54	1.60
$\theta_{02}^{F} - \theta_{04}^{F}$	1.53	1.55
$\theta_{03}^{F} - \theta_{04}^{F}$	−4.50	1.54
θ_{1}	1.08	.040
θ_{2}	.730 E − 01	.067
θ_{3}	2.99	.120
θ_{4}	.999 E − 02	.064
θ_{5}	−.697 E − 01	.054
θ_{6}	−.136	.073
θ_{7}	−.102	.094
θ_{04}^{D}	−.335	27.3
$\theta_{01}^{D} - \theta_{04}^{D}$	−.180	27.3
$\theta_{02}^{D} - \theta_{04}^{D}$	.938 E − 01	27.2
$\theta_{03}^{D} - \theta_{04}^{D}$	−.275	27.3
θ_{04}^{C}	−.644	1.55
$\theta_{01}^{C} - \theta_{04}^{C}$	−.825	2.12
$\theta_{02}^{C} - \theta_{04}^{C}$	−.425 E − 01	1.50
$\theta_{03}^{C} - \theta_{04}^{C}$	−.417 E − 01	1.52
θ_{04}^{B}	−1.62	1.81
$\theta_{01}^{B} - \theta_{04}^{B}$	.474	1.85
$\theta_{02}^{B} - \theta_{04}^{B}$	−5.34	1.75
$\theta_{03}^{B} - \theta_{04}^{B}$	−.447	1.79
Utility of leisure		
b_{04}^{n}	4.85	4.21
$b_{01}^{n} - b_{04}^{n}$	9.12	3.70
$b_{02}^{n} - b_{04}^{n}$	−.227	.941
$b_{03}^{n} - b_{04}^{n}$	11.9	3.57
b_{1}^{n}	1.00	1.20

TABLE A.1

MODEL ESTIMATES (CONTINUED)

	Parameter	Asymptotic Standard Error
Utility of attending school		
b_{04}^{sn}	1.18 E + 04	7.97 E + 03
$b_{01}^{sn} - b_{04}^{sn}$	−1.29 E + 04	6.23 E + 03
$b_{02}^{sn} - b_{04}^{sn}$	1.91 E + 02	1.56 E + 05
$b_{03}^{sn} - b_{04}^{sn}$	−1.18 E + 04	6.18 E + 01
b_1^{sn}	−2.16 E + 03	871
b_2^{sn}	697	595
b_3^{sn}	3.48 E + 03	1.92 E + 03
b_4^{sn}	−1.14 E + 03	2.10 E + 03
b_0^{sp}	−6.01 E + 03	2.05 E + 03
b_1^{sp}	1.15 E + 03	378
b_2^{sp}	−2.50 E + 03	1.43 E + 03
b_3^{sp}	4.03 E + 03	1.26 E + 03
b_4^{sp}	1.88 E + 03	1.23 E + 03
b_0^{sf}	−1.49 E + 04	3.69 E + 03
b_1^{sf}	2.53 E + 03	659
b_2^{sf}	−4.27 E + 03	6.50 E + 03
b_3^{sf}	8.36 E + 03	1.86 E + 03
b_4^{sf}	4.14 E + 03	1.06 E + 03
PDV of HS graduation		
γ_{04}	1.66 E + 06	3.98 E + 03
$\gamma_{01} - \gamma_{04}$	−5.74 E + 05	3.53 E + 09
$\gamma_{02} + \gamma_{04}$	9.40 E + 04	2.99 E + 06
$\gamma_{03} - \gamma_{04}$	−1.34 E + 05	7.95 E + 04
γ_1	1.23 E + 03	9.66 E + 03
γ_2	.903	4.78
Hours offer functions		
$t <= 5$		
θ_{013}^h	2.35	.514
θ_{014}^h	.210	1.21
θ_{015}^h	−1.84	10.9
θ_{023}^h	.883	.583
θ_{024}^h	.123	.471
θ_{113}^h	−.210	.112
θ_{114}^h	−.108 E − 01	.962
θ_{115}^h	−.378 E − 01	4.39
θ_{123}^h	.570 E − 01	.691
θ_{124}^h	.401 E − 01	.146
$t >= 6$		
θ_{013}^h	−5.46	95.4
θ_{014}^h	−4.03	6.65
θ_{015}^h	−.895	1.54
θ_{023}^h	.127 E − 01	.796
θ_{024}^h	.559	.499

TABLE A.1

MODEL ESTIMATES (CONTINUED)

	Parameter	Asymptotic Standard Error
Job-finding costs		
c^{np}	1.14 E + 03	637
c^{nf}	3.48 E + 03	927
Type proportions		
π_1	.422 E − 01	.022
π_2	.764	.045
π_3	.142	.023
π_4	.518 E − 01	—
Variance-covariance matrix[a]		
$\sigma^2_{\epsilon^n}$	1.99	.425
$\sigma^2_{\epsilon^p}$	.600	.045
$\sigma^2_{\epsilon^f}$	.515	.063
$\sigma^2_{\epsilon^{sn}}$	4.49 E + 03	1.36 E + 04
$\sigma^2_{\epsilon^{sp}}$	4.14 E + 03	1.22 E + 03
$\sigma^2_{\epsilon^{sf}}$	495 E + 03	1.91 E + 03
$\rho_{\epsilon^{sn}\epsilon^{sp}}$	−.159	.052
$\rho_{\epsilon^{sn}\epsilon^{sf}}$	−.939	.848
$\sigma^2_{\eta^p}$	.101	.093
$\sigma^2_{\eta^f}$	.536 E − 01	.378
Ln likelihood		−9,166

[a] The standard errors are those of the Cholesky decomposition parameters.

REFERENCES

ANDREWS, DONALD W. K. (1988): "Chi-Square Diagnostics for Econometric Models: Theory," *Econometrica*, 56, 1419–1453.

BELLMAN, RICHARD (1957): *Dynamic Programming*. Princeton: Princeton University Press.

CAMERON, STEVEN, AND JAMES J. HECKMAN (1993): "The Nonequivalence of High School Equivalents," *Journal of Labor Economics*, 11, 1–47.

——— (1999): "Can Tuition Policy Combat Rising Wage Inequality," in *Financing College Tuition*, ed. by Marvin H. Kosters. Washington, D.C.: AEI Press.

COLEMAN, JAMES S. (1984): "The Transition from School to Work," in *Research in Social Stratification and Mobility*, Vol. 3, ed. by Donald J. Treiman and Robert V. Robinson. Greenwich, CT: JAI Press.

D'AMICO, RONALD (1984): "Does Employment During High School Impair Academic Progress?" *Sociology of Education*, 57, 152–164.

ECKSTEIN, ZVI, AND KENNETH I. WOLPIN (1989a): "Dynamic Labour Force Participation of Married Women and Endogenous Wage Growth," *Review of Economic Studies*, 56, 375–390.

——— (1989b): "The Specification and Estimation of Dynamic Stochastic Discrete Choice Models," *Journal of Human Resources*, 24, 562–598.

——— (1990): "Estimating a Market Equilibrium Search Model from Panel Data on Individuals," *Econometrica*, 58, 783–808.

EHRENBERG, RONALD G., AND DANIEL R. SHERMAN (1987): "Employment While in College, Academic Achievement and Postcollege Outcomes," *Journal of Human Resources*, 23, 1–23.

GILLESKIE, DONNA B. (1998): "A Dynamic Stochastic Model of Medical Care Use and Work Absence," *Econometrica*, 66, 1–46.

GREENBERGER, ELLEN, AND LAURENCE STEINBERG (1986): *When Teenagers Work: The Psychological and Social Costs of Adolescent Employment*. New York: Basic Books.
HECKMAN, JAMES J., AND BURTON SINGER (1984): "A Method for Minimizing the Impact of Distributional Assumptions in Econometric Models for Duration Data," *Econometrica*, 52, 271–320.
HOOD, A., AND C. MAPLETHORPE (1986): "Bestow, Lend or Employ: What Difference Does it Make?" *New Directions for Institutional Research*, 25, 61–73.
KEANE, MICHAEL P., AND KENNETH I. WOLPIN (1994): "The Solution and Estimation of Discrete Choice Dynamic Programming Models by Simulation and Interpolation: Monte Carlo Evidence," *Review of Economics and Statistics*, 76, 648–672.
——— (1997): "Career Decisions of Young Men," *Journal of Political Economy*, 105, 473–522.
——— (1998): "The Effect of Parental Transfers and Borrowing Constraints on Educational Attainment," Mimeo, University of Pennsylvania.
——— (forthcoming): "Eliminating Race Differences in School Attainment and Labor Market Success," *Journal of Labor Economics*, forthcoming.
LIGHT, AUDREY (1995): "High School Employment," Report: NLS 95-27, U.S. Department of Labor Bureau of Labor Statistics.
MARSH, HERBERT W. (1991): "Employment During High School: Character Building or a Subversion of Academic Goals?" *Sociology of Education*, 64, 172–1989.
MILLER, ROBERT (1984): "Job Matching and Occupational Choice," *Journal of Political Economy*, 92, 1086–1120.
MORTIMER, JEYLAN T., MICHAEL D. FINCH, SEONGRYEOL RYU, MICHAEL J. SHANAHAN, AND KATHLEEN T. CALL (1996): "The Effects of Work Intensity on Adolescent Mental Health, Achievement and Behavioral Adjustment: New Evidence from a Prospective Study," *Child Development*, 67, 1243–1261.
NATIONAL CENTER FOR EDUCATIONAL STATISTICS (1992): *Overview and Inventory of State Requirements for School Coursework and Attendance*. Washington, DC: U.S. Department of Education.
NEAL, DEREK A., AND WILLIAM R. JOHNSON (1997): "The Role of Pre-Market Factors in Black-White Wage Differences," *Journal of Political Economy*, 104, 869–895.
PAKES, ARIEL (1986): "Patents as Options: Some Estimates of the Value of Holding European Patent Stocks," *Econometrica*, 54, 755–784.
ROSENZWEIG, MARK R., AND KENNETH I. WOLPIN (1994): "Are There Increasing Returns to the Intergenerational Production of Human Capital," *Journal of Human Resources*, 29, 670–693.
RUST, JOHN (1987): "Optimal Replacement of GMC Bus Engines: An Empirical Model of Harold Zurcher," *Econometrica*, 55, 999–1033.
——— (1992): "Structural Estimation of Markov Decision Processes," forthcoming in *Handbook of Econometrics*, Vol. 4, ed. by R. Engle and D. McFadden. Amsterdam: North Holland.
WOLPIN, KENNETH I. (1984): "An Estimable Dynamic Stochastic Model of Fertility and Child Mortality," *Journal of Political Economy*, 92, 852–874.
——— (1987): "Estimating a Structural Search Model: The Transition from School to Work," *Econometrica*, 55, 801–817.