Compensating Differentials in College and the Workplace, Learning, and the Choice of College Major

Peter Arcidiacono*

November 29, 2001

Abstract

Large differences earnings and ability differentials exist across majors. I estimate how much of the earnings differentials are due to ability and selection. I also separate out the non-monetary compensating differentials individuals receive both in the workplace and in school from the differences in returns to abilities across the majors. To separate out workplace differentials from school differentials, the dynamics of students' educational decisions are analyzed. Differences in monetary returns explain little of the ability sorting across majors; virtually all sorting is because of the compensating differentials at college and in the workplace. Large earnings premiums exist for certain majors even after controlling for selection.

Key Words: Dynamic Discrete Choice, Returns to Education, Human Capital, Compensating Differentials.

JEL: I21, C35, J24, D83

---

*Department of Economics, Duke University. Email: psarcidi@econ.duke.edu. The author thanks Eric French, John Jones, Derek Neal, Peter Norman, Jonathan Parker, John Rust, Marc Rysman, Jim Walker, and seminar participants at Duke University, RAND, UC San Diego, University of Oregon, University of Iowa, University of Hawaii, University of Tennessee, and the Federal Reserve Bank of Chicago. The author is especially thankful to John Kennan for many helpful conversations. The author gratefully acknowledges financial support from the Alfred P. Sloan Foundation, and computing resources from Jim Walker.

1 Introduction

Large earnings and ability differences exist across majors. Selection into majors depends upon the monetary returns to various abilities, the compensating differentials in the workplace, and the compensating differentials for studying particular majors in college. This paper seeks to separate out these three effects and identify how much each contributes to the sorting across majors.

To accomplish this, I propose a dynamic model of college and major choice which has three periods. In the first period, individuals choose both a college and a major or choose to enter the labor force. The first period decision is made given expectations about what choices will be made in the second period. In the second period, individuals learn more about the characteristics of each of the majors as well as how they perform in the college environment. With this new information, individuals update their decisions by changing their major and/or changing their college, or entering the labor force. In the third period, individuals work, receiving earnings based upon their past educational choices.[1] The model is flexible enough to capture the relationship between college quality and choice of major while allowing individuals to switch majors over time.

Two veins of the human capital literature that have also examined returns to schooling are relevant to the work here. The first focuses on dynamic models of years of schooling including Keane and Wolpin (1997), and Cameron and Heckman (1999, 2001). These articles focus on dynamic selection with Keane and Wolpin also stressing the importance of how schooling decisions are made given expectations about future events. The second, while not modelling the decisions to attend school, show that all schooling is not rewarded equally in the labor market. Daniel, Black, and Smith (1997), Loury and Garman (1995), and James et al. (1989) find that college quality,[2] major,[3] and grade point average all affect future earnings. Brewer and Ehrenberg (1999) model the selection into tiers of college and the corresponding returns to these tiers. This paper bridges these two literatures by disaggregating the schooling decision in a dynamic environment.

[1] Altonji (1993) proposes a similar theoretical model.

[2] The latter two use average SAT score of the school as a measure of college quality. Both also find that the relationship between college quality and future earnings, while positive, is weak. The former finds larger effects using a selectivity index.

[3] Paglin and Rufolo (1990) also find large differences in earnings across majors and suggest that ability sorting plays a large role in choice of major.

Recognizing the importance of college attendance, and in turn college quality, in determining future earnings, Fuller, Manski and Wise (1982) estimate a model of college choice. Berger (1988), Daymont and Andrisani (1984), and Turner and Bowen (2001), on the other hand, examine how people choose their college majors. No attempt has been made to integrate the college major decision with the college choice decision empirically, yet there is reason to believe that the two decisions are linked. Further, neither the literature on college choice nor the literature on choice of major has treated these decisions as dynamic for the individual. While transferring schools is a somewhat rare event, changing majors is not. I develop a dynamic model of college choice and major where students receive new information about their abilities over time. With this new information, students revise their decisions as to what schools to attend and what majors to study.

I find that math ability is important both for labor market returns and also for the sorting into particular majors. In contrast, verbal ability has little effect on labor market outcomes or on sorting. I find significant effort costs which depend upon relative math ability. These costs are convex and lead to interior optimal school qualities. Individuals trade off the costs of attending higher quality colleges with the benefits coming later in the form of both compensating differentials and higher monetary returns. Large monetary premiums exist for choosing particular majors even after controlling for selection. However, these large premiums and the differential monetary returns to ability and college quality cannot explain the ability sorting present across majors; virtually all sorting is occurring because of compensating differentials either in school or in the workplace. In contrast, the fact that schools are heterogeneous does lead to some ability sorting.

Section 2 provides the basic trends in the data. A dynamic model of college and major choice as well as the econometric techniques are described in Section 3. Section 4 provides the empirical results and estimates the premiums for different majors. Section 5 examines how well the model predicts the trends seen in the data as well as simulating how various features of the environment lead to ability sorting. Section 6 concludes the paper.

2 Education Choices and Earnings Outcomes

This section provides descriptive statistics on the earnings and characteristics of individuals who participated in the National Longitudinal Study of the Class of 1972 (NLS72).[4] I study those NLS72 participants who were accepted to a four-year institution and who reported test score information. Only those who indicated that they had been accepted to a four year college were included in the sample. I aggregate majors into four categories: Hard Sciences (including math), Business (including economics), Social Sciences/Humanities/Other, and Education. The major findings are:

1. Earnings are strongly correlated with major choice.

2. Ability sorting across majors occurs both before and during college.

3. Lucrative majors draw the high math ability students at each school.

4. Poor performance is correlated with dropping out or switching to a less lucrative major.

Table I addresses the first point listed above, displaying 1986 mean earnings data for individuals by their intended major going into college (1972 college major) and their major halfway through their college career (1974 college major).[5] Standard deviations are in parentheses.

Note the more than sixteen thousand dollar spread between the highest paying major, hard science, and the lowest paying major, education, for the 1972 choice. In fact, those who chose not to attend college actually had higher average earnings than those who chose education either in 1972 or in 1974.

The difference in earnings between the hard sciences and education increases once the students update their major decision. With the exception of education, all majors see higher earnings after the re-sorting.[6] Table II provides some insight as to why this is the case.

[4]The NLS72 is a stratified random sample which tracks individuals who were seniors in high school in 1972. Individuals were interviewed in 1972, 1973, 1974, 1976, 1979, and 1986.

[5]Individuals must have worked between 30 and 60 hours in an average week and, given that average work week, earned between five and a half and one hundred and forty eight and a half thousand dollars a year (1999 dollars). In order to keep the samples somewhat consistent, other restrictions on the data were made as outlined in the appendix.

[6]Although 1974 business majors are listed here as having higher earnings than hard science majors, under less rigorous selection rules this does not hold. This is the only case where the qualitative results change because of the selection rules.

Table 1: 1986 Earnings† by College Major

Major	1972 Choice	1974 Choice	Difference
Hard Science	50,535 (24,805)	52,315 (24,419)	1,780
Business	49,249 (26,227)	52,796 (23,015)	3,547
Social Science/ Humanities	38,955 (18,583)	43,088 (22,288)	4,133
Education	33,616 (13,589)	32,305 (10,417)	-1,311
No College	36,478 (18,016)	36,664 (18,642)	185

† Translated into 1999 dollars.

Table II shows mean math and verbal SAT scores by major as well as the corresponding means for the individual's peers for both 1972 and 1974.[7] With the differences in math abilities across majors, we begin to see that that the results in Table I may be in large part due to selection.[8] Differencing one's own ability from the ability of his peers shows that the ordering of majors by earnings is very similar to the ordering by relative math ability. This is not the case when we add the verbal score as business majors due substantially worse than their peers on the verbal portion of the SAT. In fact, while business majors have essentially the same math abilities and attend the same quality institutions as the social science and humanities majors, their verbal scores are much worse. It is interesting to note that hard science majors not only have the highest math SAT scores, but the highest verbal scores as well.

[7]For those who did not take the SAT, SAT scores were predicted using scores from a standardized test taken by the survey recipients and demographic characteristics. Results from these regressions are reported in the appendix.

[8]Although we would expect the average SAT math and verbal gaps to be zero, the actual average math and verbal gaps are -15.1 and -25.7 respectively. If we limit the sample to those who take the SAT, the problem is somewhat mitigated shrinking the math and verbal gaps to -6.3 and -18.3 respectively.

Table 2: SAT Scores by College Major

		1972 Choice			1974 Choice		
		Own	Peer	Relative	Own	Peer	Relative
	Major	Ability	Ability	Ability	Ability	Ability	Ability
	Hard Science	566	547	19	594	560	34
		(103)	(62)	(93)	(98)	(64)	(88)
	Business	498	522	-24	528	533	-5
		(105)	(58)	(90)	(92)	(55)	(87)
SAT Math	Social Science/	500	526	-26	518	535	-17
	Humanities	(104)	(58)	(95)	(100)	(56)	(91)
	Education	458	502	-44	467	504	-37
		(95)	(50)	(89)	(95)	(54)	(84)
	No College†	430			482	514	-32
		(102)			(110)	(57)	(102)
	Hard Science	499	515	-16	523	526	-3
		(106)	(58)	(96)	(106)	(60)	(93)
	Business	444	494	-50	464	501	-38
		(96)	(57)	(85)	(93)	(52)	(90)
SAT English	Social Science/	481	499	-18	499	510	-12
	Humanities	(107)	(56)	(96)	(102)	(56)	(93)
	Education	431	477	-46	438	478	-40
		(92)	(48)	(87)	(89)	(51)	(84)
	No College	404			445	486	-41
		(98)			(106)	(53)	(97)

†No college in 1974 refers to those who attended in 1972 but not in 1974.

Selection also plays a role in who stays at college. Table II shows that for all majors both the average student abilities and the average peer abilities increase after allowing students to drop out of college or switch to a different major. The largest differences across years in average abilities were in the most lucrative majors. Drop outs had lower math and verbal scores and came from schools with lower scoring peers on average than those who chose to stay in school in a major besides education. This can explain part of the increases in average earnings for non-education majors between the 1972 and 1974 choice. Education majors in 1972 look very similar to education majors in 1974, suggesting that whatever increases in earnings we would except to see by removing drop outs is offset by individuals switching into and out of education.

Table III provides some information on who it is that is transferring into and out of particular majors or dropping out. Significant differences exist in drop out rates across majors, with hard science and business majors dropping out less often than education and social science/humanities majors. Despite the low drop out rates, hard science majors are the least likely to stay with their initial choice. This suggests that it may be easier to switch into other majors from a hard science major.

The characteristics of those who drop out are very different across initial majors. While business majors drop out at the same rate as hard science majors, the abilities of the business major drop outs and their peers are substantially below their hard science counterparts. Despite this, earnings for business major drop outs are only slightly lower than the earnings for hard science drop outs and significantly higher than social science/humanities majors. Besides having much lower verbal scores, business major drop outs look very similar to social science/humanities majors. Those who drop out of education are of significantly lower abilities, attend lower quality schools, and earn much less than all of the other drop out groups. Further, conditional on choosing a major besides education, drop outs earn more than those who switch into education. In all other cases, drop outs earn less.

Comparing cross major switches (those who choose major i then major j with those who choose j then i) yields more ability sorting. Hard science majors who switch to business have on average lower math scores than business majors who switch to the hard sciences. Similarly, those who switch from hard science to social science/humanities or education have lower math abilities than the social science/humanities and education majors who switch to the hard sciences. Excluding hard science majors, business majors trade lower math ability

Table 3: SAT and Earnings Transitions

1972 Major	Variable	1974 Major: Hard Science	Business	Soc Sci/ Hum	Education	Drop Out	Total
	Own Math Ability	602	555	543	494	537	566
	Peer Math Ability	565	539	539	495	532	547
Hard Science	Own Verbal Ability	529	468	496	461	466	499
	Peer Verbal Ability	531	504	514	472	498	515
	1986 Earnings	54333	55348	47309	33849	44473	50535
	% of '72 Hard Science	42%	8%	19%	2%	28%	100%
	Own Math Ability	579	516	506	498	455	498
	Peer Math Ability	534	530	539	515	502	522
Business	Own Verbal Ability	449	452	470	439	420	444
	Peer Verbal Ability	495	501	515	489	475	494
	1986 Earnings	42860	51862	55700	38110	40429	49249
	% of '72 Business	3%	54%	10%	3%	30%	100%
	Own Math Ability	546	533	515	475	474	500
	Peer Math Ability	530	541	535	515	513	526
Soc Sci/ Humanities	Own Verbal Ability	501	493	503	456	452	481
	Peer Verbal Ability	504	508	509	484	487	499
	1986 Earnings	40273	51211	41855	31289	33621	38955
	% of '72 Soc Sci/Hum	4%	4%	49%	7%	36%	100%
	Own Math Ability	583	537	479	460	433	458
	Peer Math Ability	544	525	513	500	495	502
Education	Own Verbal Ability	495	492	478	429	409	431
	Peer Verbal Ability	516	495	490	476	471	477
	1986 Earnings	51176	57437	36537	32315	30364	33616
	% of '72 Education	2%	3%	11%	51%	33%	100%

students for higher ability students as well. Finally, social science/humanities majors who switch to education are of lower math abilities than education majors who switch to social science/humanities.[9] There is also some evidence of comparative advantage as those who have high math scores relative to verbal scores are more likely to choose a hard science major over a social science/humanities major.

With over 30% of those attending college in 1972 dropping out by 1974 and another 18% switching majors,[10] individuals must be learning about their tastes and abilities. Table IV presents some evidence that learning about one's abilities affects one's future educational decisions. Table IV displays freshmen grade point averages[11] and standard deviations by 1972 major across four groups: those who switched to a more lucrative major (switched up), stayed in the same major, switched to a less lucrative major (switched down), and those who dropped out. In all cases those who dropped out tended to have poorer college performances

Table 4: Grades in 1972 Major by Switching Category†

	Hard Science	Business	Social Science/ Humanities	Education	Total
Switch Up		2.71 (0.39)	2.84 (0.60)	2.94 (0.56)	2.87 (0.59)
Same Major	3.03 (0.64)	2.71 (0.58)	3.00 (0.56)	2.78 (0.54)	2.93 (0.59)
Switch Down	2.69 (0.64)	2.59 (0.70)	2.82 (0.56)		2.71 (0.63)
Drop Out	2.51 (0.70)	2.44 (0.76)	2.60 (0.65)	2.51 (0.70)	2.55 (0.69)

†Switching down refers to switching to a less lucrative (low SAT math gap) major, while switching up refers to switching to a more lucrative (high SAT math gap) major.

than those who decided to remain in college. Also, those who switched down performed

[9] The same pattern is observed for verbal scores with one exception: business majors who switch to the hard sciences score worse than hard science majors who become business majors.

[10] By definition, disaggregating the major categories further would lead to even more switching.

[11] This is a categorical variable which is taken from the survey. Students were asked to give their average g.p.a. at the time of the survey. Midpoints of the categories were used in all mean calculations.

worse than those who stayed in the same major or switched up. Comparing switching up to keeping the same major leads to a more muddled picture. In two of the three cases those who switched up had performed better than those who stayed the same, though this pattern is not observed in the largest category. Perhaps the learning that occurs may be partially major specific. Those who perform poorly know to try a different major or drop out. If performing well is more important in the lucrative majors, poor performers who stay in school will find one of the lower paying majors more attractive. Those who perform well have an incentive to stay in the same major to take advantage of the major-specific skills. However, to the degree that the discovered abilities are general, students have an incentive to switch to majors where their abilities may be put to better use.

3 Model and Estimation

The trends in the data suggest a dynamic model of college decision-making. I model the college education process as consisting of three periods. In the first two periods individuals decide among a variety of schooling options or choose to enter the workforce. All individuals work in the third period, reaping the benefits of their past educational decisions. A broad outline of the model is summarized below.

1. In period 1, individuals are given a choice set from which they can choose both a college and a major or enter the labor market. The choice set is the set of schools where the individual was accepted. The labor market is an absorbing state.

2. After the first period decision, those who chose a schooling option receive new information about their abilities (through grade point averages) and how well they like particular fields (through preference shocks).

3. In period 2, those who pursued a schooling option again choose from the same schooling options as in period 1 or enter the labor market.

4. After the second period decision, individuals who chose a schooling option in the second period again receive new information about their abilities. They then enter the labor market in period 3.

Periods 1 and 2 roughly correspond to the individual's first two and last two years of college. Period 3 includes all years after college. The model involves estimating parameters of two types: utility function parameters and transition parameters. Transition parameters are only used in forming expectations about uncertain future events. These include the parameters of the grade generating process, through which individuals learn about their abilities and the corresponding value of pursuing particular educational paths, and the parameters of the earnings process which dictate expectations individuals have about future financial outcomes. I first discuss the transition parameters and for the moment assume that errors are independent across all stages of the model. I then relax this assumption later in the paper.

3.1 The Labor Market

Once individuals enter the workforce they make no other decisions: the labor market is an absorbing state. Earnings are a function of observed ability, A, where A is individual specific. I assume that the human capital gains for attending the jth college operate through the average ability of the students at the college, $\overline{A}_j$. In some majors individuals may acquire more human capital than in other majors, leading to earnings differentials across majors. I assume that log earnings for a particular year are given by:

$$\ln(W_{jkt}) = \gamma_{w1k} + \gamma_{w2k}A + \gamma_{w3k}\overline{A_j} + \gamma_{w4k}Z_w + g_{kt} + \epsilon_t \tag{1}$$

where individual i subscripts are suppressed. Z_w is a vector of other characteristics which may affect earnings, k indicates major, and g_{kt} is the growth rate on earnings. The shocks (the ϵ_t's) are assumed to be distributed $N(0, \sigma_w^2)$.

I use the SAT math and verbal scores as my measures of observable ability. The corresponding school averages[12] are then used for the measure of school quality. Average abilities need not be interpreted as peer effects but as the best measure of college quality available. Crucial to identification of the coefficient on earnings in the utility function is that an exclusion restriction exists. The exclusion restriction I use is the college premium between 1973 and 1975 across states as calculated from the Current Population Survey (CPS).[13] Other

[12] These data are taken from the colleges themselves. The Basic Institutional Source File has information taken from the 1973-74 Higher Education Directory, the 1973-74 Tripariate Application Data file, the 1972-73 HEGIS Finance Survey and the 1972 ACE Institutional Characteristics File.

[13] Sparsely populated states are aggregated in the CPS, so instead of 50 data points there are actually only 22. I regress log earnings for those who are 22 to 35 years of age on an age quadratic for both men an women. I then pull out the gender and age-specific effects and average across regions to obtain the college premiums.

variables used in the log earnings regression are grade point average and gender.

3.2 Learning

While grade point averages are expected to have a positive effect on future earnings, individuals learn about their abilities through them as well. A signal on unobserved ability is given in the realization of first period college performance. Performance in the first period, G_1, is a function of the major chosen, k, individual and choice specific characteristics, Z_j, as well as a noisy signal of the unobserved ability A_{uk} which is partially major specific. Specifically let A_{uk} follow:

$$A_{uk} = \phi_1 + \phi_{2k}$$

I assume ϕ_1 is distributed $N(0, \sigma_{u1}^2)$ while ϕ_{2k} is distributed $N(0, \sigma_{u2}^2)$ for all k. Further the absolute ability component (ϕ_1) and the major specific component ϕ_{2k} are independent from one another. Since this is ability which individuals were not able to forecast, it is independent from Z_j in the first period.

Specifically, performance in the first period takes the following form:

$$G_{1|jk} = \gamma_{1k} + \gamma_{2k} Z_j + A_{uk} + \eta_1 \tag{2}$$

where γ_1 is a vector of coefficients to be estimated and η_1 is a white noise component distributed $N(0, \sigma_1)$.

Each individual takes the difference between his actual performance and expected performance as his signal on the unobserved ability. I am assuming here that the econometrician and the individual have the same information set when predicting first period performance, an assumption which will be relaxed when controls for unobserved heterogeneity are implemented later in the paper. Let the signal given on the unobserved ability be called A_{u1k}, where $A_{u1k} = A_{uk} + \eta_1$.

The unobserved ability also affects second period performance, which in turn affects the present value of lifetime earnings in the final period. I assume the second period grade process has the same parameters up to a parameter on the intercept term and on the ability terms. That is, the relative value of particular abilities within majors does not change across periods. Performance in the second period then takes the following form:

I use the same restrictions on extreme observations as in the previous section.

$$G_{2|jk} = \gamma_3\gamma_{1k} + \gamma_4(\gamma_2 Z_j) + \gamma_5 S + \gamma_4 A_{uk} + \eta_2 \quad (3)$$

where S indicates that the individual switched majors and η_2 is again a white noise component. Note that Z_j may be different from the Z_j in the first period as individuals may transfer schools. Individuals use the information they receive from first period performance to forecast second period performance.

$$G_{2|jk} = \gamma_3\gamma_{1k} + \gamma_4(\gamma_2 Z_j) + \gamma_5 S + \gamma_4 \frac{\sigma^2_{u_1} + \sigma^2_{u_2}(k_1 = k_2)}{\sigma^2_{u_1} + \sigma^2_{u_2} + \sigma^2_1} A_{u1k} + \eta \quad (4)$$

η is then a sum of normally distributed variables which are independent from Z_j.[14] The errors from this regression are heterosckedastic as those who switch majors are expected to have higher variances on their error terms. The signal to noise ratios for those who stay in the same major versus those who switch then makes it possible to identify the importance of major specific ability versus absolute ability.

For my performance measure, I use the individual's college grades during the year immediately after the student has made the period 1 and period 2 choices. Hence, the grades used will be those reported in 1973 and 1975. All school variables are based upon the choices made in October of 1972 and 1974. The variables which make up Z_j are the abilities of the individual (both verbal and math), the average abilities of the individual's peers and information about the individual's performance in high school.

3.3 Choice of College and Major

Individuals may choose a school from a set J where colleges themselves are not important; it is only the characteristics of the colleges that are relevant to the model. That is, one does not receive utility from attending Harvard but from attending a school that has faculty and students with particular characteristics. Those who decide to attend college must also choose a major from the set K. The same set of majors exist at all colleges. When making the

[14] To see this, consider the regression $A_{uk} = \gamma A_{u1k} + \epsilon$. The error from this regression is, by construction, orthogonal to A_{u1k}. Z_j is then also orthogonal to ϵ since the only correlation Z_j had with A_{uk} was through the sum A_{u1k}. γ is then the signal to noise ratio: how much information the draw on A_{u1k} is providing on A_{uk}.

college and major decisions, individuals take into account the repercussions these decisions have on future earnings.

The NLS72 has data on the top three schooling choices of the individual in 1972 and on whether or not the individual was accepted to each of these schools. It also has data on the schooling choice made in 1972 and 1974 and I restrict the data set to those student's who attend one of their top three choices in both periods. Unfortunately, the NLS72 does not have data on whether an individual was considering any other four year institutions. Hence, I may only be partially observing the choice set.[15] I aggregate majors into four categories as in the previous section. The maximum number of choices available in periods 1 and 2 is then thirteen: four majors for each of three schools and a work option.

I assume that utility is separable over time. Utility of being in the workforce is given by the expected discounted sum of the log of yearly earning as well as non-monetary compensating differentials:

$$u_{wjk} = \alpha_{w1} X_{wjk} + \alpha_{w2} \sum_{t=t'}^{T} \beta^{t-t'} E\left(\log\left[W_{jkt}\right]\right) \tag{5}$$

where T is the retirement date and t' is the year the individual enters the workforce. The first term here represents a compensating differentials component to working in a particular field. For example, an individual who is not good at math may not want a math-intensive job beyond the fact that he may be compensated less because of his poor math skills. The expression for utility can then be rewritten as:[16]

$$u_{wjk} = \alpha_{w1} X_{wjk} + \alpha_{w2} \left[\gamma_{w1k} + \gamma_{w2k} A_i + \gamma_{w3k} \overline{A_j} + \gamma_{w4k} Z_w + \sum_{t=t'}^{T} \beta^{t-t'} E\left(\log\left[\exp(g_{tk} + \epsilon_t)\right]\right)\right]. \tag{6}$$

Define the flow utility u_{tjk} as the utility received while actually attending college j in major k at time t. This flow utility includes the effort demanded by choosing major k at school j as well as any compensating differentials which may take place. The flow utility for pursuing a particular college option is then:

$$u_{tjk} = \alpha_{c1k} X_j - c_{tjk} + \epsilon_{tjk} \tag{7}$$

[15] Not observing other schools in the choice set does not appear to be important as those students who applied to at least three schools are less than 15% of all NLS72 participants who applied to college.

[16] One of the advantage of choosing the log utility specification is that errors in growth rates result in changes in the coefficients on the constant and gender terms in the utility specification but do not affect other parameter estimates. While the NLS72 has good data on yearly earnings for 1973 through 1979 and also for 1986, we have little information on the growth rates by major late into the life cycle.

where X_j is a vector of individual and school characteristics which affect how attractive particular education paths are.[17] These include such things as the cost of the school, college quality as a consumption good, and whether particular sexes have preferences for particular majors. The individual's unobserved preferences for the schooling options is given by the ϵ_{tjk}'s.

Each of the majors varies in their demands upon the students. I assume that each major requires a fixed amount of work which varies by the individual's ability, A, ability of one's peers, $\overline{A}_j$, the ability that is learned about in college, A_{uk}, and the major chosen, k. Let A_M and A_E represent math and verbal ability; with total ability given by $A_T = A_M + A_E$. The cost of effort, c_{jk}, is assumed to follow:

$$c_{tjk} = \alpha_{c2k}(A_M - \overline{A}_{jM}) + \alpha_{c3k}(A_V - \overline{A}_{jV}) - \alpha_{c4}(A_T - \overline{A}_{jT})^2 \tag{8}$$

Note that the psychic cost function allows the costs to majoring in particular fields to vary by relative ability in the linear term, but not in the squared term. This cost of effort may lead to optimal qualities that are on the interior: even if an individual was allowed to attend all colleges, the individual may choose to not attend the highest quality college because of the effort required. With different levels of effort required by different majors, optimal college qualities may vary by major. Individuals are then trading off the cost of obtaining the human capital with the future benefits.

After making the second period college decision, there are no decisions left and the individual enters the workforce. The expected present discounted value of lifetime utility conditional of choosing a college option in the second period, v_{2jk}, is then given by:

$$v_{2jk} = E_2(u_{2jk} + \beta u_{wjk}) \tag{9}$$

where β is the discount rate. Individuals then choose the option which yields the highest present value of lifetime utility. Note that the unobservable preference term ϵ_{2jk} is embedded in u_{2jk} and is known to the individual but not the econometrician. What is unknown to the individual is grades in the second period and the time path of earnings. The expected present value of choosing to enter the workforce is then the sum of the expected log of lifetime earnings, u_{wo}.[18]

[17] I have no data on major specific variables, hence there is no k subscript.

[18] Any compensating differentials in the workplace are then relative to the no college compensating differentials.

To the extent that compensating differentials appear through variables that are both in X_j and Z_j, the compensating differential in college will not be separately identified from the compensating differential in the workplace using just second period decisions. However, also modeling the first period decision makes it possible to separate the two parts out.

In the first period, individuals taken into account how their actions will affect the value of their future choice. Let V_2 indicate the best option in the second period. Individuals then choose the v_{1jk} which yields the highest utility where v_{1jk} is given by:

$$v_{1jk} = E_1(u_{1jk}) + \beta E_1(V_2|d_1 = j, k) \tag{10}$$

This second expectation is taken with respect to both shocks to ability and shocks to preferences. Individuals get to optimize again after the realization of these shocks, but there is a cost to not knowing this information *a priori*.

By integrating out the new information on one's abilities in the expectation of future utility and assuming that the new information is uncorrelated with the unobservable preferences, equation (11) results.

$$v_{1jk} = E_1(u_{1jk}) + \beta \int E_1 \left(V_2|A_{u1}, d_1 = j, k\right) \pi_1(dA_{u1k}|X_j, d_1 = j, k) \tag{11}$$

Note that there is still an expectation operator in front of the future utility component because individuals receive draws on their unobservable preferences after making their first period decisions. Even if the new information on ability was known to the individual, the second period decision would still be stochastic because of the evolution of the unobservable preference parameters (the ϵ's).

In order to actually estimate models of this type, some assumptions need to be made on the distribution of the unobserved preferences. Specifically, let the ϵ_{tjk}'s in each time period be taken from a generalized extreme value distribution which yields nested logit probabilities in a static model: schooling options in one nest, work options in the other. That is, ϵ_{tjk} has a component which does not vary across schooling options. Let the variance for the cross-school component at time t be given by μ_{2t}. The variance on ϵ_{tjk} itself be given by μ_{1t} where μ_{1t} must be greater than μ_{2t}.[19]

[19] All parameters in discrete choice models are relative to the variance term where the variance term is usually normalized. Here, I normalize with respect to μ_{11}

With the added assumption that the unobservable preference terms are uncorrelated over time,[20] closed form solutions for the conditional expectations of future utility exist. Specifically, the present value of indirect utility for attending school j in major k during period 1 is now given by equation (12).[21]

$$v_{1jk} = E_1(u_{1jk}) + \beta\mu_{12} \int \ln \left[\left(\sum_j \sum_k \exp\left[(v_{2jk} | A_{u1k}, d_1 = j, k) \right] \right)^{\frac{\mu_{22}}{\mu_{12}}} + \exp(u_{wo}) \right] \pi_1(dA_{u1k}) \tag{12}$$

Recall that A_{u1k} is found through the first period performance regression given in equation (2), where A_{u1k} was assumed to be a normally distributed random variable. In order to evaluate this expression Rust (1987) showed that, by discretizing the values A_{u1k} can take, it is possible to obtain consistent estimates of the parameters of the utility function.[22] With $p_1(\cdot)$ being the discretized version of $\pi_1(\cdot)$, equation (13) results.[23]

$$v_{1jk} = E_1(u_{1jk}) + \beta\mu_{12} \sum_m \ln \left[\left(\sum_j \sum_k \exp\left[E_1\left(v_{2jk} | A_{u1m}, d_1 = j, k \right) \right] \right)^{\frac{\mu_{22}}{\mu_{12}}} + \exp(u_{wo}) \right] p_1(A_{u1m}) \tag{13}$$

With the assumptions made on the distribution of the unobservable preferences and the earnings and grade processes, the probability of an individual choosing school j and major k in period t take a nested logit form:[24]

$$Pr(d_t = j, k) = \frac{\exp(v'_{tjk})}{\sum_j \sum_k \exp(v'_{tjk}) + \left(\sum_j \sum_k \exp(v_{tjk}) \right)^{1-\mu_{2t}} \exp(u_{wo})}$$

where the sums are taken over all possible options available to the individual. The expected net present value of indirect utility for attending school j in major k during period t is then

[20] An assumption which is made more palatable in the next section.

[21] See McFadden (1981) for the result.

[22] Keane and Wolpin (1994) present an alternative method which does not involve discretizing the error term. Their method involves approximating the integrals with different functions. Keane and Wolpin (1997) use this method in their model of career decisions. For good reviews on solution methods for dynamic discrete choice problems see Eckstein and Wolpin (1989) and Rust (1994, 1996).

[23] Here, A_{u1k} does not depend upon X_j as the expectations are on the forecast error which are independent from X_j in the first period.

[24] Rust (1987, 1988) showed this explicitly for the multinomial logit case and his general model produces the nested logit specification as a specific case.

given by v'_{tjk}. It is 'net' because I am differencing all of the present value of indirect utility functions in period t by the present value of pursuing the work option in period t. It is 'expected' for two reasons. First, both the researcher and the individual only have expectations regarding the value of future decisions. It is also 'expected' because the unobservable preferences are unobserved to the researcher, and I am defining v'_{tijk} such that it does not include these unobservable preferences. Since only part of the indirect utility is observed in v'_{tijk}, the decision an individual makes from the researcher's standpoint is random.

3.4 Restrictions on Compensating Differentials

Given three ability variables (SAT math, SAT verbal, and high school class rank), two college quality variables, compensating differentials in the workplace and at college, and four majors, forty compensating parameters exist just from ability and college quality. All of these measures are highly correlated. I make the following assumptions to reduce the number of parameters that need to be estimated:

1. Net of the effort cost, the ability compensating differential in college is proportional to the ability compensating differential in the workplace.

2. Net of the effort cost, the college quality compensating differentials are proportional to the ability compensating differentials.

These assumptions reduce the number of parameters from forty to twenty-two. The effort costs, which previously could not be separated out from the other compensating differentials, are now identified. Identification comes from patterns of behavior which are not proportional. No restriction is placed on the sign of the effort costs; the model may produce estimates which are inconsistent with the theory. I test these restrictions by estimating the unrestricted and restricted models and using a likelihood ratio test.

3.5 Estimation Strategy

With independent errors across the earnings, grades, and choice processes, the log likelihood function is the sum of three pieces:

$L_1(\gamma_w)$- the log likelihood contribution of earnings.

$L_2(\gamma_g)$- the log likelihood contribution of grade points averages.

$L_3(\alpha, \gamma_g, \gamma_w)$- the log likelihood contribution of college and major decisions.

with 143 parameters to estimate.

It is possible to estimate all the parameters in the indirect utility function, the performance equations, and the log earnings equations using full information maximum likelihood. However, this would be computationally burdensome. Note that consistent estimates of γ_w and γ_g can be found from maximizing L_1 and L_2 separately.[25] With the estimates of γ_w and γ_g, consistent estimates of α can be obtained from maximizing L_3.[26]

3.6 Serial Correlation of Preferences and Unobserved Ability

One of the assumptions which seems particularly unreasonable is that the unobservable preferences parameters are uncorrelated over time. That is, if one has a strong unobservable preference for engineering initially, he is just as likely as someone who has a strong unobservable preference for education initially to have an unobservable preference for education when it comes time to choose a college and a major in the second period. We would suspect that this is not the case. Further, it is unreasonable to assume that there is no unobserved (to the econometrician) ability which is known to the individual.[27]

Mixture distributions provide a way of controlling for serial correlation and selection. Assume that there are R types of people with π_r being the proportion of the rth type in the population.[28] Types remain the same throughout all stages, individuals know their type, and preferences for particular fields and college quality may then vary across types. An example would be if the parameters of the utility function do not vary across types except for the constant term. This would be the same as having a random effect which is common across everyone of a particular type. The log likelihood function for a data set with I observations is then given by:

---

[25] See Rust and Phelan (1997) and Rothwell and Rust (1997).

[26] The standard errors are not consistent, however, unless the covariance matrix of the parameters is *block diagnol* as estimates of transition parameters are being taken as the truth. Full information maximum likelihood with one Newton step would produce consistent estimates of the standard errors. See Davidson and MacKinnon (1993) for how using this two step method affects the standard errors. Rust and Phelan (1997) note that in other work the two stage estimation procedure has had little effect on the standard errors.

[27] See Willis and Rosen (1979) for the importance of selection in education.

[28] See Keane and Wolpin (1997) and Eckstein and Wolpin (1999) for other examples of using mixture distributions to control for unobserved heterogeneity in dynamic discrete choice models.

$$L(\alpha, \gamma_g, \gamma_w) = \sum_{i=1}^{I} \ln \left(\sum_{r=1}^{R} \pi_r \mathcal{L}_{1ir} \mathcal{L}_{2ir} \mathcal{L}_{3ir} \right) \tag{14}$$

Here, the α's and γ's can vary by type and $\mathcal{L}$ refers to the likelihood (as opposed to the log likelihood).

Now the parts of the log likelihood function are no longer additively separable. If they were, a similar technique could be used as in the case of complete information: estimate the model in stages with the parameters of previous stages being taken as given when estimating the parameters of subsequent stages. Using the EM algorithm,[29] I am able to return the additive separability.

Note that the conditional probability of being a particular type is given by:

$$Pr_i(r|\mathbf{X_i}, \alpha, \gamma, \pi) = \frac{\pi_r \mathcal{L}_{1ir} \mathcal{L}_{2ir} \mathcal{L}_{3ir}}{\sum_{r=1}^{R} \pi_r \mathcal{L}_{1ir} \mathcal{L}_{2ir} \mathcal{L}_{3ir}} \tag{15}$$

where $\mathbf{X_i}$ refers to the data on the decisions and the characteristics of the individual.

The EM algorithm has two steps: first calculate the expected log likelihood function given the conditional probabilities at the current parameter estimates, second maximize the expected likelihood function holding the conditional probabilities fixed. This process is repeated until convergence is obtained. But the expected log likelihood function here is now additively separable.

$$\sum_{i=1}^{I} \sum_{r=1}^{R} Pr_i(r|\mathbf{X_i}, \alpha, \gamma, \pi) \left(L_{1ir}(\gamma_w) + L_{2ir}(\gamma_g) + L_{3ir}(\alpha, \gamma_g, \gamma_w) \right) \tag{16}$$

Taking the conditional probabilities as given, I can get estimates of γ_w from maximizing the L_{1r}'s times the conditional probabilities. Similarly, estimate γ_g from maximizing the conditional probabilities times the L_{2r}'s. I then only use the L_{3r}'s to find estimates of α— not needing the L_{3r}'s to obtain estimates of γ_w and γ_g. Note that all of the parts of the likelihood are still linked through the conditional probabilities where the conditional probabilities are updated at each iteration of the EM algorithm. Arcidiacono and Jones (2000) show this method produces consistent estimates of the parameters with large computational savings.

[29] See Dempster, Laird, and Rubin (1977)

4 Empirical Results

This section provides the results from estimating the parameters of the performance equations, the log annual earnings equations, and the structural parameters of the utility function. Although the results of the model with unobserved heterogeneity are interdependent, I present the estimation of each equation separately.

4.1 Performance Regressions

The results of estimating the parameters of the first period performance equation are given in Table 5. The first column displays the coefficient estimates without unobserved heterogeneity, while the second gives estimates with unobserved heterogeneity approximated by two types. Two additional restrictions are placed on the coefficients. First, the coefficient on math (verbal) college quality is constrained to be proportional to the coefficient on math (verbal) ability. The sign, however, is not constrained. Second, the coefficient on do not know is class rank is constrained to be proportional to the effect of high school class rank. The assumption is then that the coefficient on do not know class rank yields what we would expect their class rank to be.

All of the ability coefficients are positive, with smaller coefficients for education. Without unobserved heterogeneity, math ability is particularly useful in the hard sciences, while verbal ability is particularly useful in the social sciences/humanities. Once the mixture distribution is added, the differences in ability coefficients within a major dissipate. High school class rank positively effects grade point averages, with those where we do not know their class rank having an expected class rank at the sixty-ninth percentile. This number is comparable to the observable mean class rank for the data.

Without the mixture distribution, the coefficient on math college quality is negative one: a one point increase in both math ability and math college quality yields no change in expected grade point averages. The coefficient on verbal college quality is negative a half, suggesting that grade inflation is more common at schools that have a disproportionately high verbal college quality. This result becomes magnified when the mixture distributions are added: schools with high math college qualities have grade deflation (the coefficient is less than negative one) with high verbal quality colleges having grade inflation.

Females receive higher grades than their male counterparts. Larger effects are found in business with smaller, but still positive, effects in the social sciences/humanities. Adding unobserved heterogeneity has little effect on the female coefficients. The results with unobserved heterogeneity show that type 2's receive substantially higher grades in all subjects.

Table 6 displays the results of the second period performance regression. Adding the mixture distribution here only affects grade point averages through the expected grade point average; that is, the predicted values from the first period regression. The expected grade point average was positive and slightly increased with the controls for unobserved heterogeneity. Both with and without the mixture distribution, the coefficient on the shock was positive while the corresponding coefficient on the shock times switching was negative. Hence, information is being conveyed in the first period shocks, a portion of which is major specific. The fixed cost of switching majors on g.p.a. was negative, but small and insignificant whether or not controls for unobserved heterogeneity were included.

Using the coefficient estimates, it is possible to back out the signal to noise ratio for those who stay in the same major and those who switch. For those who stay in the same major the estimated signal to noise ratio is .54 and .51 without and with unobserved heterogeneity respectively. These numbers decrease to .45 and .40 if the individual switched majors. That the numbers are smaller when unobserved heterogeneity is added makes sense: more about what the individual knows is in expected g.p.a. rather than in the shock, while the transitory portion is still present in the shock. Hence, without unobserved heterogeneity we would be overestimating the informational content of the shock.

4.2 Log Earnings Regressions

Estimates of the log earnings equations are given in Table 7. A key to later to identifying the coefficient on earnings in the utility function is to have a variable which is only in the log earnings regression. As previously discussed, I use average state earnings for both workers who graduated from college and those who did not graduate from college as the exclusion restriction. The coefficient is positive and significant, though the magnitude does drop when unobserved heterogeneity is added.

The ability and school quality variables are all constrained to be greater than zero. Throughout, it is the math ability and college quality which matters: the constraint on verbal ability and college quality almost always binds. The highest returns for math ability are seen for hard science majors, while math school quality is most important for social sci-

Table 5: First Period Performance Regressions (1973 G.P.A.)

		One Type		Two Types	
		Coefficient	Stand. Error	Coefficient	Stand. Error
	Hard Science	0.1502	0.0443	0.1156	0.0212
SAT Math	Business	0.1060	0.0504	0.0983	0.0328
Interactions	Soc/Hum	0.0944	0.0330	0.1012	0.0170
(00's)	Education	0.0592	0.0431	0.0697	0.0280
	Hard Science	0.0816	0.0425	0.1136	0.0260
SAT Verbal	Business	0.1100	0.0611	0.1170	0.0363
Interactions	Soc/Hum	0.1590	0.0312	0.1510	0.0181
(00's)	Education	0.1053	0.0514	0.0935	0.0342
	Hard Science	1.0474	0.1972	1.0632	0.1921
HS Class Rank	Business	0.9004	0.2449	0.9065	0.2455
Interactions	Soc/Hum	0.7607	0.1472	0.7283	0.1442
	Education	1.1111	0.2404	1.1516	0.2435
	Hard Science	0.1586	0.0695	0.1441	0.0665
Female	Business	0.1890	0.1128	0.2160	0.1105
Interactions	Soc/Hum	0.0763	0.0525	0.0727	0.0520
	Education	0.1796	0.1040	0.1859	0.1005
	Hard Science	1.7394	0.2496	1.5799	0.2229
Constant	Business	1.7607	0.2244	1.6681	0.2296
	Soc/Hum	1.8962	0.1942	1.8608	0.1831
	Education	1.6176	0.2315	1.4863	0.2086
Coefficients	Don't Know Rank	0.6906	0.0499	0.6907	0.0484
Common	Math Quality	-1.0544	0.6001	-1.7190	0.3133
Across Majors	Verbal Quality	-0.4786	0.5594	0.0074	0.2788
	Hard Science			0.3822	0.0593
Type 2	Business			0.2360	0.0932
Interactions	Soc/Hum			0.1406	0.0486
	Education			0.2749	0.0809

Table 6: Second Period Performance Regressions (1975 G.P.A.)

	One Type		Two Types	
	Coefficient	Stand. Error	Coefficient	Stand. Error
Constant	0.4532	0.0550	0.4022	0.0345
Expected G.P.A.	0.8274	0.0347	0.8700	0.0207
First Period Shock	0.4459	0.0275	0.4400	0.0188
Shock*Switch Majors	-0.0757	0.0480	-0.0944	0.0327
Switch	-0.0297	0.0255	-0.0246	0.0180

ence/humanities majors. Adding the mixture distribution lowers the return to college quality for hard science majors while keeping the other college quality coefficients close to the case without unobserved heterogeneity.

Without the controls for unobserved heterogeneity, college grades are found to be an important contributor to future earnings. This is particularly the case for business majors: going from a 2.5 to a 3.0 yields an over thirteen percent increase in yearly earnings. For the other majors, a similar increase in grade point average would yield around a five percent increase in earnings. With the exception of education majors, these effects diminish substantially when the mixture distribution is added. Now, going from a 2.5 to a 3.0 in business yields less than a eight percent increase in yearly earnings. In fact, the coefficient on grades actually becomes negative for hard science majors. This may be due to an aggregation problem as biology majors may receive higher grades but lower earnings than the other hard science majors.

Types 2's received significantly higher earnings in all fields except for education. The type 2 coefficient for education is mitigated, however, by the positive effect type 2 has on grades in education. The implied correlations with grades suggests that the unobserved ability to perform well in school translates into higher earnings not only if the individual attended college, but also if the individual chose the no college option. Also included, but not reported, are private school interacted with field, sex interacted with field, year dummies interacted with college, and sex and year dummies interacted with college.

Whether premiums exist for particular majors is difficult to see given all of the interactions

Table 7: Log Earnings Regressions[†]

		One Type		Two Types	
		Coefficient	Stand. Error	Coefficient	Stand. Error
	State Average Earnings	0.5938	0.0779	0.2925	0.0344
SAT Math Interactions (00's)	Hard Science	0.0425	0.0361	0.0506	0.0159
	Business	0.0198	0.0391	0.0217	0.0172
	Soc/Hum	0.0165	0.0165	0.0203	0.0100
	Education	0.0000	0.0370	0.0000	0.0163
	No College	0.0304	0.0126	0.0310	0.0055
SAT Verbal Interactions (00's)	Hard Science	0.0000	0.0317	0.0000	0.0140
	Business	0.0000	0.0391	0.0000	0.0173
	Soc/Hum	0.0000	0.0224	0.0151	0.0099
	Education	0.0005	0.0376	0.0000	0.0165
	No College	0.0000	0.0131	0.0000	0.0058
Math School Quality Interactions	Hard Science	0.0617	0.0948	0.0119	0.0422
	Business	0.0032	0.1339	0.0015	0.0589
	Soc/Hum	0.0557	0.0926	0.0647	0.0407
	Education	0.0000	0.0370	0.0000	0.0586
Verbal School Quality Interactions	Hard Science	0.0000	0.1087	0.0000	0.0481
	Business	0.0010	0.1385	0.0000	0.0609
	Soc/Hum	0.0000	0.0960	0.0000	0.0422
	Education	0.0140	0.1403	0.0196	0.0618
Grades	Hard Science	0.0617	0.0515	-0.0808	0.0233
	Business	0.2742	0.0547	0.1578	0.0245
	Soc/Hum	0.1076	0.0402	0.0283	0.0179
	Education	0.1264	0.0603	0.1722	0.0269
Type 2 Interactions	Hard Science			0.5411	0.0223
	Business			0.4379	0.0228
	Soc/Hum			0.4468	0.0152
	Education			-0.3144	0.0229
	No College			0.4470	0.0089

[†]Sex and private school interacted with major and sex and college interacted with year dummies were also included along with major-specific constant terms.

and the effect of ability and college quality through grades. Premiums for choosing one of the college majors over the no college option are displayed in Table 8 for both an average male and an average female. For the case with two types, I use the mean probabilities of being each of the types (.5025 and .4875 for type 1's and type 2's respectively). Also displayed are the net percentage increases over the no college sector of increases in math ability and college quality. I do not analyze the effect of verbal ability and college quality as the constraint that the coefficients be greater than zero almost always binds.

Significant premiums exist for both the average male and female ranging from a high of 27.6% (females in business, controlling for heterogeneity) to a low of -1.2% (males in education, whether or not we control for unobserved heterogeneity). The largest premiums are found in the hard science and business majors, implying that the gap in earnings across fields is not entirely driven by high ability individuals choosing the more lucrative fields. Adding unobserved heterogeneity had mixed effects on the premiums. Larger premiums existed for the hard science majors, but smaller premiums for the social science/humanities majors.

The total returns to math ability, both through grades and directly, are higher for hard science and business majors than in the no college sector. This is not the case for social science/humanities and education majors: from an earnings standpoint, increases in math ability make these two majors less attractive compared to the no college option. The returns to math college quality are positive for hard science and social science/humanities majors but negative for business and education majors. Even though the direct effects of college quality are constrained to be greater than zero, college quality can still have a negative effect through grades. This negative effect is stronger than the positive direct effect for business and education majors. The mixture distributions substantially lowered the returns to math college quality for hard science majors while increasing the returns for social science/humanities majors.

4.3 Estimates of the Utility Function

I next use the estimates of the performance and log earnings equations to obtain the second stage maximum likelihood estimates of the utility function parameters. Table 9 displays the maximum likelihood estimates for the parameters which are major specific. Sex and high school class rank interacted with major, along with major specific constant terms, were also

Table 8: Premiums for Different Majors†

		Hard Science	Business	Soc Sci/Hum	Education
	Males	19.7%	15.9%	9.4%	-1.2%
One Type	Females	15.0%	24.4%	13.0%	5.2%
	Change in Premium				
	+100 SAT Math	2.1%	1.8%	-0.4%	-2.3%
	+100 Math Quality	5.2%	-2.7%	4.5%	-0.8%
	Males	22.1%	14.4%	5.4%	-1.2%
Two Types	Females	22.0%	27.6%	5.7%	6.0%
	Change in Premium				
	+100 SAT Math	1.0%	0.6%	-0.8%	-1.9%
	+100 Math Quality	2.8%	-2.5%	6.0%	-2.1%

†Premiums are relative to No College

included.

The first two sets of rows display the ability compensating differentials for each field beyond the effort costs required in school. High math ability is more attractive for hard science and business majors, while high verbal ability is more attractive for social science/humanities majors. Controlling for unobserved heterogeneity had a very little effect on these coefficients.

The efforts costs, displayed in the next two rows, show that math ability is particularly useful in school. All of the math effort costs (in the form of relative math ability) are positive and significant, with a larger coefficient for hard science majors. While the magnitudes of these coefficients are reduced when the mixture distribution is added, they are still all positive and significant. On the other hand, the estimates reveal no significant verbal effort costs. Math ability, as in the log earnings regressions seems to be much more important than verbal ability when predicting trends in major choice and returns to schooling.

Positive shocks to performance made staying in school more attractive, with stronger effects in the hard sciences and in social science/humanities. The magnitudes of all the coefficients fall when the mixture distribution is added. This makes sense: previously one's type would be somewhat included in this performance shock. With an individual's type removed from the performance shock the information conveyed in the shock is not as relevant. Switching to a different major was very costly, though the costs were much smaller for switching into social science/humainities.

Table 10 displays the utility coefficients which are common across majors. The coefficient on log earnings is positive and significant, though falls by more than half with the mixture distribution. Transferring schools is very costly, with a coefficient very similar to the coefficients on switching majors. The monetary cost of school acts as a deterrent to choosing a schooling option, with the effect stronger for those who come from a low income household. This suggests that low income households are either liquidity constrained or that there parents are paying a lower portion of their college expenses. Estimates of the yearly discount factor are 101% and 73% for the models without and with unobserved heterogeneity respectively. Both coefficients are significantly different from zero.

Squared relative ability is negative and significant, suggesting that interior optimal school qualities may be a possibility. Further, much of the compensating differentials and monetary returns to college quality exist after the individual finishes college. This suggests that the optimal first period college quality may be lower than the optimal second period college quality. Fixing verbal ability at the college verbal ability, I calculate the optimal gap between one's own math ability and that of the college if individuals were just maximizing the first period flow utility.[30] Without unobserved heterogeneity, the optimal math gaps ($A_M - \overline{A}_M$) are 19, -3, 61, and 45 for hard science, business, social science/humanities, and education respectively. This implies that first period flow utility is generally higher when individuals attend colleges with students who have lower math abilities than their own. These results are somewhat tempered when the mixture distribution is added with the optimal gaps now at 14, 0, 42, and 35. The estimates imply that individuals want to attend higher quality colleges not for the flow utility, but for the future utility.

Compensating differentials based upon abilities do exist in the workplace. The estimates of the proportion of the ability compensating differential at school that is received in the workplace is much larger than one. If the coefficient had been one, then the compensating differential in the workplace, beyond the monetary returns, would be zero. The quality compensating differential is positive and significant suggesting that college quality serves as

[30] Included in this calculation is the fact that college quality serves as a consumption good as shown by the second row in the second set of Table 10.

Table 9: Major-Specific Utility Function Parameters†

		One Type		Two Types	
		Coefficient	Stand. Error	Coefficient	Stand. Error
SAT Math Interactions (00's)	Hard Science	0.0511	0.0327	0.0482	0.0107
	Business	0.0296	0.0214	0.0276	0.0095
	Soc/Hum	0.0003	0.0072	0.0007	0.0065
	Education	0.0066	0.0102	0.0038	0.0088
SAT Verbal Interactions (00's)	Hard Science	-0.0051	0.0087	-0.0003	0.0078
	Business	-0.0075	0.0108	-0.0052	0.0090
	Soc/Hum	0.0320	0.0210	0.0315	0.0081
	Education	0.0018	0.0100	0.0039	0.0094
Relative Math Ability	Hard Science	0.2434	0.1002	0.1522	0.0443
	Business	0.1118	0.0657	0.0702	0.0398
	Soc/Hum	0.1232	0.0452	0.0861	0.0314
	Education	0.1155	0.0515	0.0801	0.0376
Relative Verbal Ability	Hard Science	0.0182	0.0385	0.0066	0.0306
	Business	-0.0250	0.0459	-0.0105	0.0363
	Soc/Hum	0.0646	0.0571	0.0283	0.0320
	Education	-0.0159	0.0436	-0.0147	0.0361
Performance Shock	Hard Science	3.3314	0.5818	3.2739	0.5287
	Business	1.9456	0.6320	1.7806	0.6018
	Soc/Hum	3.0493	0.5626	2.6733	0.4846
	Education	2.2684	0.6228	1.8546	0.5935
Switching Costs	Hard Science	-2.0541	0.1878	-2.0075	0.1892
	Business	-2.6595	0.2039	-2.6802	0.2017
	Soc/Hum	-0.8508	0.1377	-0.8307	0.1363
	Education	-2.5851	0.1984	-2.5692	0.1986
Type 2 Interactions	Hard Science			0.0808	0.0242
	Business			0.0882	0.0195
	Soc/Hum			0.1221	0.0178
	Education			0.1625	0.0164

†Sex and high school class rank interacted with major, along with major-specific constant terms, were also included.

a consumption good both in and out of the workplace. Recall that the reason there is just one number here for the quality compensating differential was because of the restrictions placed on the compensating differential for abilities and college quality. A likelihood ratio test that the restrictions hold cannot be rejected at the 90% level.

The nesting parameters are both relative to the variances of the no college error. These nesting parameters measure the cross-school component of the variance. In particular, had these coefficients been estimated to be one, than a multinomial logit would have resulted. That the actual estimates are less than one suggests that the preferences for schooling options are correlated.

Table 10: Utility Coefficients Common Across Majors

	One Type		Two Types	
	Coefficient	Stand. Error	Coefficient	Stand. Error
Log Earnings	2.2378	0.9775	0.8866	0.1450
Transfer Schools	-2.1362	0.1348	-2.1630	0.1358
Private School	0.1415	0.0542	0.1022	0.0322
School in Same State	-0.0323	0.0375	-0.0258	0.0282
Net Cost ($000's)	-0.0764	0.0287	-0.0526	0.170
Low Income*Net Cost ($000's)	-0.1235	0.0333	-0.0889	0.0183
Relative Ability Squared	-0.0103	0.0057	-0.0070	0.0035
Discount Factor	1.0081	0.0792	0.7312	0.0846
Ability Comp. Differential at Work	6.1049	4.0973	9.6913	0.1529
Quality Comp. Differential	4.0071	2.4365	2.5600	0.7105
First Period Nesting Parameter	0.6160	0.1057	0.6571	0.0865
Second Period Nesting Parameter	0.6466	0.0979	0.6496	0.0789
Variance on First Period Decision†	0.3651	0.1332	0.2680	0.0965

†Variance on the second period decision is is normalized to one.

5 Model Fit and Simulations

This section shows how well the model matches the trends in the data as well as performing simulations to see what factors are leading to the ability sorting across majors. Table 11 compares the actual data on ability and college quality distributions to what the model predicts for both the first and second period choices. If the restrictions on the ability and college quality coefficients are wrong, then the estimated ability and college quality distributions will be wrong as well.[31] The data are indexed by 'D', with estimates without and with unobserved heterogeneity indexed by '1T' and '2T' respectively.

In all cases, the models with and without the mixture distributions predict the trends in the data very well. The models often hit the observed mean exactly. The only case where the distribution is slightly off is for business majors. For the first period choice, I overpredict by five and four points math ability and math college quality. Similarly, for the second period choice I underpredict math ability by six points. Both here and for the rest of this section, adding unobserved heterogeneity did not improve the predictions.

While the model does a good job predicting the means, it is also important to see how well the model predicts the transitions. Table 12 displays the transition matrix for the data and for the models both for math ability and for the percentage choosing each field. I focus on math ability because it has much larger effects on both earnings and choice of major than the verbal ability. The model predicts the percentage of people in each cell very well; including matching the higher drop out rates found for the social science/humanities majors. Recall that it was much easier to switch into social science/humanities than the other majors. If an individual initially chose social science/humanities and did not like it, there is no low cost switch for him to make besides dropping out. While the model predictions often predict the observed distributions exactly, both models underpredict the drop out rates of education majors by three percent.

Math ability transitions are also given. These do not match near as well, and this may in part be because of the small cell sizes. Looking along the diagnol, where most of the observations are, shows that the models predict well the ability levels of those who stay in the same major. It also predicts well the trend of decreasing math abilities along the rows. That is, higher math ability students are more likely to choose hard science over business, business

[31] I do not compare model predictions for the percentage of people choosing each major as these have to match in a nested logit framework when constant terms for each major and period are estimated.

Table 11: Model Fit- Abilities†

			1972 Choice		1974 Choice	
			Own	Peer	Own	Peer
	Major		Ability	Ability	Ability	Ability
		D	566	547	594	560
	Hard Science	1T	565	545	594	562
		2T	564	545	595	562
		D	498	522	528	533
	Business	1T	503	526	522	531
		2T	503	526	522	532
		D	500	526	518	535
SAT Math	Social Science/	1T	499	525	519	535
	Humanities	2T	499	524	519	536
		D	458	502	467	505
	Education	1T	459	502	467	503
		2T	459	503	467	503
		D	499	515	523	526
	Hard Science	1T	499	514	519	528
		2T	498	514	519	528
		D	444	494	464	501
	Business	1T	447	496	462	500
		2T	447	496	460	501
		D	481	499	499	510
SAT Verbal	Social Science/	1T	481	500	499	510
	Humanities	2T	480	499	501	510
		D	431	477	439	478
	Education	1T	432	477	439	478
		2T 431	477	438	478	

†'D' refers to the actual data, '1T' to the estimates with one type, and '2T' to the estimates with two types.

over social science/humanities, and social science/humanities over education. However, the levels of the non-diagnol elements are off. Both models overpredict the abilities of those hard science majors who stay in school and, consequently, underpredict the abilities of hard science majors who drop out. In contrast, the model predicts average math ability of business drop outs to be nineteen points higher than in the data.

Given that the model matches the data, I can use the model to simulate how the math ability distributions would vary given a different environment. The first simulation assumes that all individuals attend the same school. This simulation is designed to answer how much of the observed differences in math ability across majors is due to individuals attending schools of different quality levels. The second simulation turns off the returns to math and verbal ability as well as the returns to math and verbal college quality. The results of the simulation will then show how much of the ability sorting is due to differences in returns to abilities and college qualities. Note that these simulations are not taking into account general equilibrium effects; this only designed to illustrate how much of the current sorting is due to heterogeneous schools and returns to abilities. Table 13 gives the results of the simulations as well as the estimates under the current environment.

The primary effect of everyone attending the same school is to lower the average abilities of those choosing hard sciences while raising those who choose education. Without unobserved heterogeneity, the gap between the average math abilities of second period hard science majors and the math abilities of second period business, social science/humanities, and education majors falls by 25%, 9%, and 17% respectively. Falls of the same magnitude are present in the first period choice. With unobserved heterogeneity, the average math ability does not fall as much. Hence, the gaps in math abilities between hard sciences and the other majors due to heterogeneous schools are now 18%, 8%, and 16% respectively.

While the effects are substantial for the policy simulation where everyone attends the same school, turning off the monetary returns to math ability and college quality has little effect on average math abilities across majors. The only effect turning off the monetary returns is a small drop, two to four points, in average abilities for all second period majors when the mixture distribution is used.

Table 12: Model Fit- Transitions

1972 Major	Variable		1974 Major Hard Science	Business	Soc Sci/ Hum	Education	Drop Out
		D	602	555	543	494	537
	Own Math Ability	1T	601	561	555	519	526
		2T	602	560	555	518	523
Hard		D	42%	8%	19%	2%	28%
Science	% of '72 Major	1T	43%	6%	18%	4%	30%
		2T	43%	6%	18%	4%	30%
		D	579	516	506	498	455
	Own Math Ability	1T	558	516	510	475	474
		2T	562	516	511	475	472
		D	3%	54%	10%	3%	30%
Business	% of '72 Major	1T	3%	54%	12%	3%	29%
		2T	3%	54%	11%	3%	29%
		D	546	533	515	475	474
	Own Math Ability	1T	560	517	513	478	474
		2T	563	517	514	477	471
Soc Sci/		D	4%	4%	49%	7%	36%
Humanities	% of '72 Major	1T	4%	5%	49%	6%	36%
		2T	4%	5%	49%	6%	36%
		D	583	537	479	460	433
	Own Math Ability	1T	537	494	491	457	443
		2T	541	495	492	457	441
		D	2%	3%	11%	51%	33%
Education	% of '72 Major	1T	2%	3%	12%	53%	30%
		2T	2%	3%	13%	52%	30%

Table 13: Simulations of the Math Ability Distribution Under Different Environments

		1972 Choice			1974 Choice		
		Baseline	Same School	No Monetary Returns on Ability	Baseline	Same School	No Monetary Returns on Ability
One Type	Hard Science	565	554	564	594	579	591
	Business	503	506	499	522	525	518
	Soc Sci/Hum	499	496	500	519	511	519
	Education	459	466	463	467	474	473
Two Types	Hard Science	564	554	563	595	582	592
	Business	503	504	500	522	522	518
	Soc Sci/Hum	499	496	496	519	512	515
	Education	459	466	458	467	474	465

6 Conclusion

Large earnings and ability differences exist across majors. Selection into majors depends upon the monetary returns to various abilities, the compensating differentials in the workplace, and the compensating differentials for studying particular majors in college.

In order to separate out these components, I estimated a dynamic model of college and major choice. Individuals made an initial college and major decision conditional on expectations on what they would do the in the future. After the initial choice, individuals received about their preferences and, through their grades, about their abilities. With this new information, individuals updated their decisions by changing their major and/or changing their college, or entering the labor force. Estimates of the model revealed that positive ability shocks made staying in school attractive, especially for those interested in the hard sciences.

Math ability is found to be important both for labor market returns and also for the sorting into particular majors. In contrast, verbal ability has little effect on labor market outcomes or on sorting. Significant effort costs exist, with the effort being a function of the person's math ability relative to his peers. These costs are convex and lead to interior optimal school qualities. Individuals trade off the costs of attending higher quality colleges with the benefits coming later in the form of both compensating differentials and higher monetary returns.

Large monetary premiums exist for choosing hard science and business majors even after controlling for selection. However, these large premiums and the differential monetary returns to ability and college quality cannot explain the ability sorting present across majors; virtually all sorting is occurring because of compensating differentials either in school or in the workplace. In contrast, the fact that schools are heterogeneous does lead to some ability sorting.

References

[1] **Altonji, Joseph.**"The Demand for and Return to Education When Education Outcomes Are Uncertain." *Journal of Labor Economics*,11 (1:1993).

[2] **Berger, Mark.**"Predicted Future Earnings and Choice of College Major."*Industrial and Labor Relations Review*, 41 (3:1988).

[3] **Cameron, S. and J. Heckman.** "The Dynamics of Educational Attainment for Black, Hispanic, and White Males." *Journal of Political Economy*, 109 (3:2001).

[4] **Cameron, S. and J. Heckman.** "Life Cycle Schooling and Dynamic Selection Bias: Models and Evidence for Five Cohorts of American Males." *Journal of Political Economy*, 106 (2:1998).

[5] **Cramer, J.S.**"Predictive Performance of the Binary Logit Model in Unbalanced Samples." *Journal of the Royal Statistical Society, Series D*, forthcoming 1999.

[6] **Daniel, K.; Black, D.; and J. Smith.** "College Quality and the Wages of Young Men." working paper 1997.

[7] **Davidson, Russell and James G. MacKinnon.** *Estimation and Inference in Econometrics.* Oxford Universtiy Press (1993).

[8] **Daymont, Thomas and Paul Andrisani.** "Job Preferences, College Major, and the Gender Gap in Earnings." *Journal of Human Resources*, 19 (1984).

[9] **Dempster, A.P.; Laird, M. and Rubin, D.B.** "Maximum Likelihood from Incomplete Data via the EM Algorithm." *Journal of the Royal Statistical Society*, (1:1977).

[10] **Brewer, D. and R. Ehrenberg.** "Does it Pay to Attend an Elite Private College? Cross Cohort Evidence on the Effects of College Quality on Earnings." *Journal of Human Resources*, (1999).

[11] **Everitt, B.S. and D.J. Hand.** *Finite Mixture Distributions.* Chapman and Hall (1981).

[12] **Freeman, Richard B.** "The Demand for Education." *Handbook of Labor Economics, Volume 1*, (1986).

[13] **Fuller, David; Manski, Charles and David Wise.** "New Evidence on the Economic Determinants of Postsecondary Schooling Choices." *Journal of Human Resources*, 17 (1982).

[14] **Grogger, Jeff and Eric Eide.** "Changes in College Skills and the Rise in the College Wage Premium." *Journal of Human Resources*, 30 (2:1995).

[15] **Heckman, James J.; Lochner, Lance and Christopher Taber.** "General-Equilibrium Treatment Effects: A Study of Tuition Policy." *American Economic Review*, 88 (2:1998).

[16] **Heckman, J. and B. Singer.** "A Method for Minizimizing the Impact of Distributional Assumptions in Econometric Models for Duration Data." *Econometrica*, 52 (2:1984).

[17] **James, Esteller; Nabeel, Alsalam; Conaty, Joseph and Duc-Le To.** "College Quality and Future Earnings: Where Should You Send Your Child to College?" *American Economic Review: Papers and Proceedings*, 79 (May 1989).

[18] **Keane, Michael P. and Kenneth I. Wolpin.** "The Solution and Estimation of Discrete Choice Dynamic Programming Models by Simulation and Interpolation: Monte Carlo Evidence." *Review of Economics and Statistics*, 76 (1994).

[19] **Keane, Michael P. and Kenneth I. Wolpin.** "The Career Decisions of Young Men." *Journal of Political Economy*, 105 (3:1997).

[20] **Loury, Linda Datcher and David Garman.**"Affirmative Action in Higher Education." *American Economic Review*, 83 (2:1993).

[21] **Loury, Linda Datcher and David Garman.** "College Selectivity and Earnings." *Journal of Labor Economics*, 13 (2:1995).

[22] **McFadden, Daniel.** "Econometric Models of Probabilistic Choice." *Structural Analysis of Discrete Data with Econometric Applications*, (Edited by Manski, Charles and Daniel McFadden). (1981).

[23] **Newey, Whitney and Daniel McFadden.** "Large Sample Estimation and Hypothesis Testing." *Handbook of Econometrics, Volume 4.* (1994).

[24] **Paglin, Morton and Anthony M. Rufolo.** "Heterogeneous Human Capital, Occupational Choice, and Male-Female Earnings Differences." *Journal of Labor Economics*, 8 (1:1990).

[25] **Rothschild, Michael and Lawrence White** "The Analytics of the Pricing of Higher Education and Other Services in Which Consumers are Inputs" *Journal of Political Economy*, 103 (3:1995).

[26] **Rothwell, Geoffrey and John Rust.** "On the Optimal Lifetime of Nuclear Power Plants." working paper, June 1996.

[27] **Rust, John.** "Optimal Replacement of GMC Bus Engines: An Empirical Model of Harold Zurcher." *Econometrica*, 55 (5:1987).

[28] **Rust, John.** "Structural Estimation of Markov Decision Processes." *Handbook of Econometrics, Volume 4.* (1994).

[29] **Rust, John.** "Numerical Dynamic Programming in Econometrics." *Handbook of Computational Economics*, (1996).

[30] **Rust, John and Christopher Phelan.** "How Social Security and Medicare Affect Retirement Behavior in a World of Incomplete Markets." *Econometrica*, 65 (4:1997)

[31] **Titterington, D.M., Smith A.F., and U.E. Makov.** *Statistical Analysis of Finite Mixture Distributions*, Wiley (1985).

[32] **Turner, S. and W. Bowen.** "Choice of Major: The Changing (Unchanging) Gender Gap." *Industrial and Labor Relations Review*, 52 (2:1999).

[33] **Willis, Robert J. and Sherwin Rosen.** "Education and Self Selection." *Journal of Political Economy*, 87 (5:1979).