

2001

2001-4 Understanding Educational Outcomes of Students from Low Income Families: Evidence from a Liberal Arts College with a Full Tuition Subsidy Program

Todd R. Stinebrickner

Ralph Stinebrickner

Follow this and additional works at: <https://ir.lib.uwo.ca/economicscibc>



Part of the [Economics Commons](#)

Citation of this paper:

Stinebrickner, Todd R., Ralph Stinebrickner. "Understanding Educational Outcomes of Students from Low Income Families: Evidence from a Liberal Arts College with a Full Tuition Subsidy Program." CIBC Centre for Human Capital and Productivity. CIBC Working Papers, 2001-4. London, ON: Department of Economics, University of Western Ontario (2001).

ISSN: 1702-0085

ISBN: 0-7714-2356-X

**Understanding Educational Outcomes of
Students from Low Income Families:
Evidence from a Liberal Arts College with
a Full Tuition Subsidy Program**

by

Todd R. Stinebrickner and Ralph Stinebrickner

Working Paper # 2001-4

December 2001



CIBC Working Paper Series

Department of Economics
Social Science Centre
The University of Western Ontario
London, Ontario, N6A 5C2
Canada

This working paper is available as a downloadable pdf file on our website
<http://www.ssc.uwo.ca/economics/centres/cibc/>

Understanding educational outcomes of students from low income families:
Evidence from a liberal arts college with a full tuition subsidy program

Todd R. Stinebrickner and Ralph Stinebrickner¹

Issues related to the schooling attainment of children from low income families arise frequently in current education policy debates. In part due to a recent increase in the disparity between the wages of college graduates and the wages of individuals with less than a college degree, there has been a specific interest in understanding why a very high percentage of children from low income families do not graduate from college and why the college graduation rates of children from low income families are substantially lower than those of children from other families. Using unique new data obtained directly from a high-quality liberal arts college that maintains a full tuition subsidy program (and large room and board subsidies) for all students, this paper provides direct evidence that reasons unrelated to the direct costs of college are very important in explaining these realities.

¹Affiliations: The University of Western Ontario and Berea College respectively. Comments very welcome. Please direct correspondence to Todd Stinebrickner at trstineb@julian.uwo.ca. We would like to thank Pam Thomas for her invaluable assistance with this project, particularly the data extraction phase. We are also grateful for helpful comments which were provided by John Bound, Caroline M. Hoxby, Jeff Smith, Dan Black, Chris Swann, Ben Scafidi, Sarah Turner, Susan Dynarski and seminar participants at a meeting of the NBER higher education group, the University of Michigan, Duke University, the University of Virginia, the University of Western Ontario, Queens University, York University, UC Irvine, McMaster, and the Upjohn Institute. The authors are grateful for funding from the Mellon foundation. The first author also received assistance from the Agnes Cole Dark fund. The second author was supported in part by Berea College through funding for a sabbatical leave.

I. Introduction

Issues related to the schooling attainment of children from low income families arise frequently in current education policy debates. While these issues span the entire spectrum of schooling levels, a recent increase in the relative wages of college graduates has contributed to a particular interest in understanding two interrelated facts associated with college attainment.² First, a very high percentage of the children from low income families who graduate from high school do not graduate from college. Second, among high school graduates, the percentage of children from low income families who graduate from college is substantially lower than the percentage of children from other families who graduate from college. Manski (1992) documents these facts using respondents from the High School and Beyond (HS&B) who were high school seniors in 1980. He finds that five and one-half years after high school graduation only .11 of respondents from families in the lowest income quintile had graduated from four year colleges, whereas .24 of respondents from families in the middle income quintile and .39 of respondents from families in the highest income quintile had graduated.

In order to graduate from college, a person must first make the decision to enter college and then must persist in college until graduation. Although much previous literature has specifically examined the college entrance decision, work such as Manski and Wise (1983), Manski(1992), and Bowen and Bok (1998) has established that examining what happens after students arrive at college is also very important if one wishes to understand the two stylized facts described in the previous paragraph.³ For example, with respect to the first stylized fact, descriptive evidence in Manski (1992) indicates that between 54% and 71% of HS&B students in the lowest income quintile who enter post-secondary

²For documentation of changes in the relative wages of college graduates, see, for example, Bound and Johnson (1992), Katz and Murphy (1992), and Murphy and Welch (1992).

³For studies of college entrance see, for example, Kane (1994) and Heckman, Lochner, and Taber (1998).

education fail to graduate from a four year college within 5.5 years. With respect to the second stylized fact, evidence in Manski and Wise (1992) indicates that between 51% and 71% of the HS&B college graduation gap between students in the lowest and highest income quintiles and between 48% and 65% of the HS&B college graduation gap between students in the lowest and middle income quintiles can be attributed to differences in college attrition rates between groups rather than to differences in college entrance rates between groups.⁴

Thus, understanding the causes of the high college attrition rates of students from low income families is important from the standpoint of understanding the low absolute and relative college graduation rates of this group that were described in the first paragraph.⁵ An explanation along traditional lines is that the low graduation rates arise largely for reasons related to the burden of paying for college. In considering the prominence that this explanation has traditionally received, it is worth noting that, although many students from low income families may not face excessively high net tuition costs due to the existence of need-based financial aid, the total direct costs for these students that arise after also factoring in the costs of room and board, books, and fees will typically be non-trivially greater

⁴In the previous two sentences, the exact percentages depend on the treatment of individuals who enter two year schools after high school graduation. Low income students are substantially less likely than other students to enter four-year institutions but are approximately equally likely to enter two-year institutions. Thus, counting individuals who enter two-year institutions when computing college entrance rates (and assuming that these students drop-out if they do not receive degrees from four-year institutions) increases both the college entrance rates and attrition rates of low income students relative to higher income students and leads to the higher number in each pair.

Using the National Longitudinal Study of the High School Class of 1972 (NLS-72), Manski and Wise (1983) found that a two standard deviation increase in family income implies a .15 increase in the probability of college persistence and a .07 increase in the probability of college entrance (holding constant other observable characteristics including parental education). Bowen and Bok (1998) also find large effects of family income on college persistence using the data from *The College and Beyond*.

⁵We stress that we use the term “high” somewhat loosely when referring to attrition rates; attrition, in and of itself, is not inherently bad and this paper does not attempt to determine what “optimal” college attrition rates would be for low income students.

than zero.⁶ Thus, the direct costs of college have the potential to be burdensome for students from poor families, especially if these families tend to be borrowing constrained.

An alternative explanation is that the high attrition rates of students from low income families arise for reasons related to a student's background or family environment that would exist even if the direct costs of college were zero.⁷ For example, students from low income families may, on average, attend lower quality elementary and secondary schools, receive less encouragement from their families to take advantage of beneficial schooling opportunities within a particular school, receive less educational instruction at home, be less likely to have parents who stress the importance of obtaining a college degree, or receive less encouragement to remain in college when academic or social difficulties arise during college.⁸ It is important to note that this family background/environment explanation is used throughout this paper to capture all reasons other than those directly related to the burden of paying for college. Thus, it also potentially includes reasons related to the interaction of borrowing constraints and financial circumstances that would be present for a student and his/her family even if direct costs were zero. For example, even in the presence of a full subsidy of direct costs, negative shocks to family income may contribute to retention differences between income groups if students from low income

⁶From a definitional standpoint, when computing the direct costs of college it is appropriate to use college room and board costs net of the costs the person would incur if he/she did not attend college (these net room and board costs may typically be positive, especially if a person tends to live at home if he/she does not attend college). However, from the standpoint of thinking about the influence of college costs on a liquidity constrained person, it may be desirable to consider all of the room and board costs since these are the costs the individual will have to find a way to pay in order to be in school. In the remainder of the paper, we abstract from this small distinction.

⁷The possibility that this type of explanation may be important in explaining differences in educational outcomes has been raised recently by Cameron and Heckman (1998), Shea (1996), and Cameron and Taber (1999). The former work suggests that "factors more basic than short-term cash constraints...determine the schooling family income relationship" and that factors such as family background "play a central role in determining schooling decisions."

⁸Many of these reasons stem from the reality that students from low income families are more likely to have parents who have not attended college. See, for example, Kiker and Condon (1981).

families are more likely to return home to help their parents in bad economic times.⁹ Similarly, students from low income families may be more likely to leave school than other students if the amount of consumption that is foregone by attending college is higher for these students and students learn about their willingness to delay consumption after making the decision to begin college.¹⁰

Learning about the importance of the “family environment” explanation and the “direct costs” explanation is of direct relevance to current policy, in part because government education policy has often been based on the belief that, in the absence of government intervention, post-secondary educational attainment may be limited for students from low income families due to borrowing constraints (Taubman, 1989).¹¹ For example, knowledge about the importance of the family environment explanation is valuable from the standpoint of understanding the extent to which expensive tuition subsidy programs by themselves may not equalize college graduation probabilities across income groups and from the standpoint of understanding the importance of also considering alternative educational policies that do not target the direct costs of college. Unfortunately, analyzing the importance of the family environment explanation or the direct costs explanation is a difficult empirical

⁹In families with lower income, the human capital of the student is likely to represent a higher proportion of total family wealth. If negative family income shocks occur, lower income families will tend to have fewer sources of wealth from which to draw and may be more likely to “cash-in” the human capital wealth of their children (especially if these families are borrowing constrained).

¹⁰The direct costs of schooling (rather than, for example, the opportunity costs of attending) are typically the relevant policy instrument for policymakers. Addressing borrowing constraints implies an effort to help a student cover the **direct** costs of schooling. Thus, while family borrowing constraints may lead to exits even if direct costs are zero (e.g., if families suffer income losses), these types of exits are appropriately attributed to what we call the “family environment” explanation because they would not be typically addressed by educational policy. That is, they would remain even under the most generous tuition/direct cost subsidy that might be considered.

¹¹Kane (1994) finds that the college entrance decisions of low income high school graduates are quite sensitive to the cost of college tuition, but also finds that a one dollar decrease in tuition has a larger effect on college attendance than a one dollar increase in the current need-based Pell Grant program. Heckman, Lochner, and Taber (1998a,b) suggest that, although the partial equilibrium effects of tax and tuition subsidies are quite large, the long run, general equilibrium effects may be much smaller.

task. As mentioned earlier, most individuals who are currently enrolled in college face total direct costs that are not zero and these costs are determined by a complex set of interactions between family income, tuition costs, financial aid grants, financial aid loans, and ability. In short, there is typically no obvious way to “control” for the effect that direct costs are having on the absolute or relative college attrition rates of students from poor families.

This difficulty has been recognized by previous literature which emphasizes that the relative importance of the two explanations is still very much an open question. For example, in the context of a discussion of college attrition, Bowen and Bok (1998) remark that “One large question is the extent to which low national graduation rates are due to the inability of students and their families to meet college costs, rather than to academic difficulties or other factors.” Similarly, when discussing differences in college attrition by family income, the findings of Mankiw and Wise (1983) lead them to “...raise the possibility that if one wanted educational attainment to be unrelated to family income, for example, low family income may have to be offset by external funds, even if income itself were not a major determinant of college attendance.”

In an effort to improve our understanding of the college outcomes of low income students, this paper takes a very simple and direct approach which is made possible by our fortunate access to unique, new data. Specifically, we analyze data from the administrative records of Berea College which is located in central Kentucky where the “Bluegrass meets the foothills of the Appalachian mountains.” As will be discussed throughout the paper, there are numerous features of the school and our data that are desirable from the standpoint of this study.¹² However, of particular interest given the nature of this

¹²One feature that is not discussed in the remainder of the paper is that Berea operates a mandatory work study program. All students work approximately ten hours a week in this program and are not allowed to add additional hours in off-campus jobs. Thus, we can rule out the possibility that attrition at this school is being caused primarily by excessive non-academic work hours during school. In addition, in Stinebrickner and Stinebrickner (forthcoming) we find no evidence that students from different income groups work different numbers of hours. As a result, we can rule out the possibility that differences in attrition between students from different income groups are caused by differences in

study is that the school provides a full tuition subsidy and large room and board subsidies to all entering students regardless of family income. Thus, the direct costs of schooling are approximately zero for the students in our data.¹³ This very unique feature, which to our knowledge only exists at perhaps one other school in the United States, allows us to directly examine two interrelated questions.¹⁴ First, to what extent are the high attrition rates of students from low income families caused by factors other than the direct costs of college. Second, to what extent are differences in attrition rates between individuals from different income groups caused by factors other than the direct costs of college.¹⁵

The paper proceeds as follows. In Section II, we discuss the data from Berea and examine descriptive Kaplan-Meier duration models that show the rate at which students leave Berea. Our findings suggest that, even though direct costs of schooling are approximately zero, roughly half of all students fail to graduate. In Section II, when we examine the attrition rates of the sample as a whole, our discussion implicitly assumes that the students at Berea can be thought of generally as a group of low income students. While the income distribution at Berea indicates that this is more or less reasonable, there exists substantial variation in the family income of the students in our sample because the poorest

non-academic work during school. Using other data, controlling for the effects that working has on attrition may be difficult

¹³According to the Berea College 1998 admissions brochure, entering students at Berea have an annual room, board, and college fee bill of only approximately \$1000. Students graduate from Berea with an average of approximately \$1000 in student loans.

¹⁴We know of only one other school in the United States that is similar in nature. College of the Ozarks in Point Lookout, Missouri is a liberal arts college which has a full tuition subsidy for all students. According to their web-site, total yearly cost for room, board, and fees are approximately \$2650 a year. We also know of two more specialized schools that also offer full tuition subsidy programs. Cooper Union in New York City offers programs in architecture, fine arts, and engineering. Webb Institute in Glen Cove, New York offers a program in naval architecture and marine engineering.

¹⁵Although students in public high schools also do not pay tuition, the high school attrition issue is somewhat different than the college attrition issue. In the former, students decide to leave school after being forced to attend school until they reach certain mandatory ages. In the latter, students have decided to leave school after deciding it was optimal to start school.

students are extremely poor while the “wealthiest” students can perhaps be thought of as lower middle class. In Section III we take advantage of this variation to examine the relationship between family income and attrition. Both Kaplan Meier survivor functions and proportional hazard survivor functions indicate that, although the direct costs of schooling are approximately zero for all students, a strong positive relationship remains between family income and the length of time that an individual remains at Berea. Thus, our primary results indicate that reasons unrelated to the direct costs of schooling appear to be very important in determining high attrition rates of students from low income families and that the differences in college outcomes by family income that have been documented in past research may remain to a large extent even under full tuition subsidy programs.

In Section IV the paper explores several possible reasons that family environment is found to have such a strong effect on college outcomes. It is worth noting again that the family environment explanation is meant to include all reasons unrelated to the direct costs of schooling, and, as a result, includes both reasons related to differences in educational preparation and encouragement and also reasons related to the interaction of borrowing constraints and financial circumstances that would be present even if the direct costs of college were zero. With respect to the former, students at Berea from lower income families are found to receive significantly lower college grades than other students (even after controlling for college entrance exam scores and other observable characteristics) and these differences explain the majority of the difference in the attrition rates between income groups. Elementary and high school quality ratings obtained from the state of Kentucky are included in an attempt to examine whether the better academic performance of the higher income students arises because they attend better schools or have better classmates. On average, students from higher income families attend better schools and school quality is found to be positively related to college grades and college persistence. Nonetheless, the effect of family income remains strong even when the school quality information is included. This suggests that parents have a strong direct effect on the academic

achievement and attainment of their children. With respect to the latter, we find no evidence that negative income shocks explain differences in educational outcomes between income groups. Nonetheless, due to the imperfect nature of our data for this particular analysis, we believe that more research is needed in order to determine the extent to which borrowing constraints and financial circumstances might influence college outcomes if direct costs were zero.

Our results strongly indicate that reasons unrelated to the direct costs of college are important in determining college attrition. However, because no students at Berea pay tuition, the situation at Berea cannot provide direct evidence about the importance that direct costs currently play in determining college attrition. In an effort to provide a rough idea of the importance of the family environment explanation relative to the direct costs explanation, the duration model of attrition is reestimated in Section V using students from the National Educational Longitudinal Study: Base Year Through Third Follow-Up (NELS-88) who entered college during the middle of the 1989-1997 period that is covered by our Berea College data. The effect of family income on attrition for students in the NELS-88 is found to be very similar to that found in the Berea data despite the fact that the NELS-88 students are attending institutions that charge tuition. Thus, although one must be very careful about the conclusions that can be drawn from this comparison, the exercise suggests that family environment factors may be the driving force in determining the strong relationship that has consistently been found between family income and college outcomes.

In Section VI, the paper concludes and discusses the implications of this work for tuition subsidy programs such as the one recently approved in the state of California.

II. Berea College history, descriptive statistics, and attrition of sample as a whole

Berea College was founded in 1855 as a one room school on land donated to founder John G. Fee by Cassius Clay, a wealthy landowner and prominent leader in the movement for gradual emancipation. According to the founder, the school was designed to be “anti-slavery, anti-caste, anti-

rum, anti-sin.” Given this background, it is not surprising that the current mission of the school is to provide an education to those of “great promise but limited economic resources.” Although Berea admits students from throughout the United States and from many foreign countries, the school has a primary focus of providing an education for students from Appalachia. In 2001, Berea was ranked first among regional liberal arts colleges in the south by U.S. News and World Report. The full tuition and room and board subsidies are made possible by a sizeable endowment.

From their administrative database, Berea College made available records for the 4089 full-time students that matriculated between the fall semester of 1989 and the fall semester of 1997. We concentrate on domestic students who did not transfer to Berea from another post-secondary institution. This eliminated six hundred students. Given the emphasis of the study on family income, the 490 students that had declared independent status were also eliminated. This left 2999 students. The covariates which are used primarily in the study are a student’s sex, race, family income at the date of matriculation (in 1997 dollars), an indicator of whether the person’s permanent home is within two hours driving distance from Berea College, family size, score on the verbal portion of the American College Test (ACT), and score on the math portion of the ACT.¹⁶ Only 178 individuals had a missing value of one or more of these variables.¹⁷ Thus, the final sample consists of 2821 individuals. Although the data from Berea also include information about high school grade point averages, this information is missing

¹⁶Most individuals in our sample took the ACT exam. In cases where a student took only the Scholastic Aptitude Test (SAT), the student’s SAT scores were converted to ACT “equivalents.” In section IV we discuss the interpretation of our results given the use of math ACT and verbal ACT as ability measures.

A student’s distance from home is potentially endogenous. However, removing this variable had very little effect on the estimated importance of the other characteristics.

¹⁷The number of individuals who had missing values of each variable is: sex=17, race=64, family income=58, math ACT=54, verbal ACT=54, and family size=11. For each variable, a probit model was estimated with an indicator of “whether or not the variable was missing” for a particular person as the dependent variable and the set of other variables as the independent variables. No evidence was found that variables are missing in systematic ways.

for 418 of the 2821 students. Thus, we choose to primarily present results from models that do not include high school grades.¹⁸

The histogram of family income in Figure 1 reveals that many students come from very poor families. One-third of students have a family income in the first year of less than \$15,800. Another one-third of students have a family income in the first year between \$15,800 and \$28,020. Most of the remaining families have a family income in the first year of less than \$50,000. Family incomes are right truncated because, as will be discussed in more detail in Section III, eligibility for admission requires that a student's family income must be below a maximum level.

The college performance outcome that we primarily concentrate on is duration of college attendance. This is worthwhile if, as research such as Kane and Rouse (1995) suggests, completing some college leads to an increase in a person's earning potential. Another reason to study duration rather than the binary college graduation outcome is that, because all students who have not finished a degree by the end of the fall semester of 1997 are right censored in our data, concentrating directly on the latter would seriously limit the amount of useable data. For example, even under the assumption that no students take more than five years to graduate, it is not possible to determine graduation outcomes for individuals who matriculated after the first three years of our data. Ignoring the last five years of data is inefficient because these years contain useful information about the likelihood of college completion.

Although students can choose to leave school at any time during the school year, the data do not indicate the exact date at which a student leaves. Instead, we observe the semester in which the person leaves. Starting in the seventh semester after matriculation, some students begin to graduate. In order

¹⁸However, comparisons of our results to estimates obtained from models which include high school grades, indicate our findings regarding the relationship between income and performance are robust to whether or not high school grades are included.

to avoid the complication of modelling both the attendance duration and the exit reason, the focus of the empirical work in this paper is on student retention up until the start of the seventh semester.¹⁹ Stinebrickner (1998) shows that almost all individuals in this sample who return for the start of their fourth year (and are not censored) eventually graduate.²⁰ Therefore, beginning the seventh semester is almost synonymous with graduation.

Figure 2 shows a non-parametric Kaplan-Meier survivor function for the duration of time that an individual in our sample remains in college. The survivor function evaluated at time t represents the probability that a student will stay more than t full semesters before leaving school (i.e., he will start at least the $t+1^{\text{st}}$ semester). The probability that an individual will stay more than six full semesters (start his seventh semester) is approximately .47. Thus, approximately half of entering at students at Berea do not graduate even though the burden of paying for college has been removed as a possible cause of attrition.

The previous results show that many students do not graduate and that exits are for reasons unrelated to the direct costs of college. It is worth noting that what is likely to be of ultimate interest from a policy standpoint is whether individuals eventually receive a degree at this school or another four year institution. However, the difference between educational attainment at Berea and total post-secondary educational attainment appears to be relatively small. In correspondence with the director of institutional research at Berea College, it was learned that exit interviews taken in recent years show that only approximately .17 of exiting students express some intent to transfer to another two year or four year post-secondary institution. Further, the majority of these students never actually

¹⁹For the empirical work, individuals who persist until the seventh semester are artificially censored at this point.

²⁰The hazard rate for dropping out without graduating in the seventh semester is .037. The hazard rate for dropping out without graduating in the eighth semester is .044.

request a transfer transcript, which in most cases, is a necessary condition for actually transferring.

III. The relationship between family income and college outcomes at Berea

Largely because many of the students are extremely poor, a substantial amount of variation exists in the family incomes of the students in our sample. In this section, we take advantage of this variation to examine whether the type of positive relationship between family income and college attrition that has consistently been found in the literature remains in a situation where the potential burden associated with paying for college has been removed.²¹ Kaplan-Meier survivor functions of the sort used in Section II, indicate that this is the case. In particular, Figure 3 shows that the Kaplan-Meier survivor functions differ for individuals in the lowest third, middle third, and highest third income groups. The probability that an individual in the highest third finishes more than six full semesters is eighteen percent larger than the probability that an individual in the lowest income third finishes more than six full semesters (.516 versus .439).

The Kaplan-Meier survivor functions do not take into account the effect that covariates have on retention. Consequently, retention differences among the income groups may be the result of differences in other observed characteristics that make students less likely to remain in school. Table 1 shows descriptive statistics for the overall sample and each of the income thirds. In general, the variable means are quite similar across income groups. This result suggests that retention differences between income groups are likely to remain even after taking into account other observable characteristics. This can be verified using a proportional hazard model.

The hazard, $h_i(t)$, represents the probability that a person will leave school at time t conditional

²¹We noted earlier that the highest observed family incomes in our sample are not particularly large. It seems that if the low income individuals in this sample are found to perform differently in terms of retention (or other outcome measures) relative to higher income individuals in this sample, they would also perform differently relative to individuals with family incomes greater than the truncation point in these data. Naturally one should be cautious when attempting to draw out-of-sample conclusions.

on not having left before time t

$$(1) h_i(t) = \exp(\beta X_i + \epsilon_i + B(t))$$

where β is a set of coefficients which measure the effect of the exogenous characteristics X_i on the hazard rate, ϵ_i represents a person specific heterogeneity term, and the baseline hazard $B(t)$ indicates how the hazard rate changes with the duration of attendance. Identification of the proportional hazard model requires that the baseline hazard be separable from other covariates. The primary results in the paper come from a specification that includes a non-parametric baseline and a parametric (normal) distribution for the unobserved heterogeneity.²² However, as discussed in footnote 23, the results were found to be very similar when the model was specified with a flexible form for the unobserved heterogeneity of the type proposed by Heckman and Singer (1984).

Table 2 shows the maximum likelihood estimates of the proportional hazard model. Column one shows estimates when family income enters as a continuous variable. Column two shows estimates when the effect of income is estimated semi-parametrically by including an indicator variable for whether a person's family income places him in the lowest third income group and an indicator variable for whether the person's family income places him in the middle third income group. Column three shows estimates when the effect of income is estimated semi-parametrically and income is divided into six different groups. Column 4 shows estimates when income enters as a continuous variable and high school grades are also included.

The coefficient associated with a particular variable can be used to compute the factor by which the hazard rate would change if the variable increased by one unit, with a negative coefficient indicating that an increase in the variable would be associated with a lower probability of leaving. For example, the coefficient on Math ACT, $-.051$, indicates that the hazard rate decreases to $\exp(-.051) = .950$ of its

²²The baseline hazard is assumed to be constant within each of the semesters. The value of each of these constants is estimated.

previous value when the Math ACT score increases by one point.

Table 2 indicates that family income has a highly significant effect, even after controlling for the effect of educational background variables and other observable characteristics. Column 1 shows that a \$10,000 increase in family income leads to a hazard rate that is lower by a factor of $\exp(-.083)=.920$.²³ For a “baseline” student, figure 4 compares the predicted survivor function for a family income of \$5,000 to the predicted survivor function for a family income of \$40,000.²⁴ The probability that the person with a \$40,000 family income remains in school for more than six full terms is twenty-five percent higher than the probability that the person with \$5,000 in family income remains in school more than six full terms (.520 versus .416).²⁵ Column 2 shows that the income coefficients are also statistically significant and quantitatively large when income enters as two indicator variables. A person in the lowest income group and middle income group have hazard rates which are $\exp(.243)=1.275$ and $\exp(.201)=1.222$ as large as the hazard rate of an individual in the highest income group holding all other observable characteristics constant. Figure 5 shows how the predicted survivor function for the baseline person varies depending on whether the person is in the lowest, middle, or highest income group. Column 3 shows that, when income is divided into six groups, retention rates are

²³As mentioned earlier, very little change was found when the model was specified with a flexible form for the unobserved heterogeneity of the type proposed by Heckman and Singer (1984). For example, for the specification in column 1 of Table 2, when ϵ_i is assumed to be a discrete random variable with two possible values it was found that the estimated effect (standard error) of family income is $-.084 (.024)$ and the value of the log likelihood function is -3484.05 . The full results for this specification in the Appendix show that estimated effects are also very similar for other observable variables. Similar results were found when the number of possible values allowed for ϵ_i was increased, and, as a result, these specifications are not shown. Although the results are not shown, the results in columns 2-4 of Table 2 were also found to be robust to the specification of the unobserved heterogeneity term.

²⁴The baseline person was given the mean values of the continuous covariates and was given median values for the indicator variables.

²⁵Note that little difference is observed in the income coefficient in column four when high school grades are included.

quite similar for the bottom three income groups but increase significantly over the upper half of the income distribution at Berea.

It is important to note that the unmeasured determinants of family income and schooling attainment may be different in our sample than it would be in various populations of interest. The types of biases that may be present from viewing the relationship between family income and college attrition at Berea as an estimator of the relationship in larger populations of interest is discussed in detail in Stinebrickner and Stinebrickner (2000).²⁶

It is worthwhile to keep in mind that the income variable used in the preceding analysis, family income at the time of matriculation, is a noisy measure of the desired variable, permanent family income. In particular, one might be concerned with the possibility that non-classical measurement error could be generated through the income threshold that is used to determine which students are eligible for admission. At low levels of permanent income, it seems likely that few households will experience shocks such that their first year family incomes make them ineligible for admission. As a result, average first year family income for the lower income groups might be expected to be very similar to average permanent family income. However, for individuals with levels of permanent income that are slightly below the income threshold, positive shocks will make them ineligible for admission, and, for individuals with levels of permanent income that are above the income threshold, negative shocks will be needed for them to enter the sample. As a result, average first year family income for the high income group might be expected to be lower than average permanent family income. If this scenario

²⁶For example, as discussed in Stinebrickner and Stinebrickner (2000), admission decisions at Berea are made among the group of “income-eligible” applicants on the basis of student quality without reference to an applicant’s family income. As a result, there is no obvious reason to believe that the admission policy at Berea is creating a relationship between family income and college attrition among attendees that is unlike what would be present among the set of all applicants to the school. Other biases can also potentially arise from students’ decisions of whether to apply to Berea and whether to attend Berea if accepted.

is true, it is possible that the associated bias could lead to an undesirable overstatement of the effect of income on duration in the model where income enters as a continuous variable. However, our investigations into this matter did not produce evidence that the income threshold causes this type of problem.²⁷

IV. Interpretation of Berea results -reasons for attrition differences between income groups.

Section III indicates that statistically significant and quantitatively large differences exist in retention rates between income groups at Berea even in the presence of the full subsidy. From the standpoint of designing effective policy programs, it would be desirable to determine which of the possible reasons related to family environment that were discussed in the introduction are most plausible. In this section, we attempt to examine this issue using our data from Berea.

Without additional information, conclusions about the reasons behind the differences by family income depend to a large extent on the interpretation of our measures of ability, the ACT math and verbal exams. For example, consider one extreme in which the ACT exams are essentially types of IQ tests that predominantly measure a person's inherent ability at birth and are largely unaffected by a person's formal and informal educational environments while growing up. In this scenario, from the standpoint of graduation probabilities for those who matriculate, the size of the difference between

²⁷For 2001 individuals in our sample, a family income value is observed in both the first and second year. Only seven of these individuals have a family income in the second year that is greater than \$70,000. Further, the average change in family income between the first year and the second year is -4888 for students classified as high income based on first year family income, -2228 for students classified as middle income, and 2313 for students classified as low income. For 1375 individuals in our sample, a family income value is observed in both the first, second, and third years. Averaging the second and third year family incomes shows that only five individuals have an average that is greater than \$70,000. Further, the average change in family income between the first year and the average of the second and third years is -4557 for students classified as high income (based on first year family income), -2277 for students classified as middle income, and 2817 for students classified as low income. The results for individuals whose income is observed for four years is similar.

income groups can reasonably be thought of as the full disadvantage of being born into a low income family. That is, the disadvantage which is attributable to any of the reasons discussed in the introduction: differences in educational opportunities and preparation, differences in parental support and encouragement during a person's college career, or differences in family responses to income shocks that are unrelated to the costs of college.

In reality it is certainly true that ACT scores to some extent also capture the amount of learning that takes place during a student's youth. However, to the extent that this endogeneity exists, from the standpoint of graduation probabilities for those who matriculate, it seems reasonable to believe that the size of the retention differences between income groups is a conservative estimator (understates) of the true lifetime disadvantage of being born into a low income family. The reason for this is simply that the test scores of students from low income families will tend to understate these students' inherent ability if students from low income families suffer on average from inferior learning environments when young. Thus, the retention differences between students from high income families and low income families would be found to be even larger if it was possible to control for "true" ability levels at birth rather than the potentially endogenous ACT scores. However, even if the scores are potentially endogenous, we would be able to rule out the possibility that differences in educational opportunities for youth in different income groups cause the income differences in retention if we believed that these test scores are able to fully capture the aspects of learning/ability that are relevant for college. However, this is not necessarily the case. For example, the math portion of the ACT exam certainly measures something about a person's quantitative background and ability, but is likely to only indirectly indicate whether an individual had the opportunity to take a calculus class while in high school.

Thus, while the previous paragraphs seem to suggest that our estimates of the influence of family income on college completion will be somewhat conservative if test scores depend endogenously on family income, they do not shed much light on the reasons for the income differences in retention rates

that remain even in the presence of the full tuition subsidy. In the remainder of this section we attempt to more directly explore the plausibility of possible explanations.

Grades and academic preparation

The previous discussion raised the possibility that students from lower income families may tend to fare worse academically in college even conditional on college entrance exam scores. Semester by semester college grade regressions suggest that this is true. For example, Table 3 shows the results of a regression of first semester college grade point average (GPA) on observable personal characteristics for all individuals who finished their first term. The coefficient on family income, .040, implies that the GPA of an individual with a family income of \$40,000 is on average .16 higher than an individual with \$0 of family income holding other observable characteristics constant.²⁸ Further, the coefficient is statistically significant with a t-statistic of 10.0. Although the regression results are not shown, the effect of family income on term grade point average is also significant (at a .10 level of significance or lower) for four of the subsequent five semesters with point estimates of .023, .036, .041, .014, and .028.²⁹ The pooled regression involving all grades in all years produces a point estimate of .033 and an associated t-statistic of 4.013.

From the standpoint of retention, whether these effects are quantitatively large depends on the nature of the relationship between grades and retention. Thus, to get a sense of the extent to which these grade differences between income groups “explain” the family income differences in retention that were found earlier, the duration model was estimated including cumulative grade point average as a

²⁸The mean and standard deviation of students’ grade point averages in the first period are 2.464 and .856 respectively.

²⁹p-values are .048, .002, .001, .277, and .048 respectively. The grade differences are consistent with (but somewhat larger) than the findings of Betts and Merrall (forthcoming) who analyze the relationship between first-semester grades and family background for students at the University of California at San Diego. Bowen and Bok (1998) also find a positive relationship between socioeconomic status and college grades.

time-varying covariate.³⁰ The coefficient on cumulative grade point average in Table 4 indicates that poor grades are a very significant predictor of exits from school. If college GPA increases by a full point, the hazard rate decreases to $\exp(-1.630) = .195$ of the previous value. A comparison of Table 4 with the first column of Table 2 shows that the effect of income decreases substantially (from $-.083$ to $-.037$) and becomes statistically insignificant when grades are taken into account.³¹

Although it is clear that college grades are strongly related to exits from school and that lower income students receive lower average college grades, one must be very cautious about what conclusions are drawn from this information. Certainly, one plausible story is that grades are essentially exogenous to the drop-out decision, in which case it would be reasonable to conclude that differences in college grade performance between income groups are caused by unmeasured differences in academic preparation or study skills between income groups. However, another possibility is that students who are unhappy at college and/or are planning to drop-out may receive lower grades simply because they are less focussed on their studies than they otherwise would be.³²

³⁰The relevant cumulative grade point average for semester t was assumed to be the cumulative grade point average of the person at the end of the semester. This timing assumption allows the student's current class performance to influence his current behavior and also ensures that a grade point average was available for a person for the first semester. However, under this timing assumption, the cumulative grade point average is not available for a person during semester t if the person leaves in the middle of this period and does not receive grades. Table 4 shows estimates which are obtained by filling in any missing grades at time t with the person's cumulative grade point average at the end of time $t-1$. The simple approach taken here represents an effort to replace the small amount of missing data in a reasonable manner without adding complexity to the empirical work. For some discussions of dealing with missing data in a more general manner see Little and Rubin (1987), Lavy, Palumbo and Stern (1998) and Stinebrickner (1999).

³¹Individuals who leave in the middle of the first semester are deleted from the analysis because no information about their grades are observed. Although this is not ideal, the number of students that leave in the middle of a semester is relatively small.

³²Another possible explanation is that students from low income families may be forced to spend more time working in non-academic jobs. This can be examined directly using the administrative data because Berea College operates a mandatory work-study program, does not allow students to work off-campus, and maintains students' work records as part of its administrative database. By taking advantage of the existence of random assignment to freshmen jobs, Stinebrickner and

While it is typically difficult to credibly separate the relative importance of the two effects, there does seem to be evidence that the latter endogeneity explanation is not the driving force. Although the full results are not shown, the estimated effect and t-statistic of family income on the cumulative grade point average at the beginning of the seventh semester for the 963 sample members who remained in school until their seventh semester are .028 and 2.9 respectively.³³ Thus, income is a significant predictor of cumulative grade point average even though members of this group are very likely to graduate, and, as a result, should be relatively unaffected by the type of grade endogeneity described above. Further, because the effect of income on grades for this group is found to be roughly constant across semesters, it does not appear that the difference in cumulative grades in the seventh semester is simply due to the low income individuals in this group adjusting more slowly to college.³⁴ Thus, the grade data does seem to provide at least some evidence that low income individuals are at a disadvantage because of their educational backgrounds, even after taking into account college entrance exam scores.

As discussed earlier, students from low income families could be less prepared for college because they receive inferior formal educational instruction or because they receive inferior educational instruction at home. In an effort to provide some information about the relative importance of the two possibilities, we concentrate on the 1188 students in the sample from the state of Kentucky. The benefit of studying this group is that, for each school year between 1991-1992 and 1996-1997, a measure of

Stinebrickner (forthcoming), find a quantitatively large (and statistically significant) negative effect between working and grade performance. However, because no relationship between family income and hours-worked is found, the results indicate that differences in employment do not explain differences in grades that exist between different income groups.

³³The regressors are the same as in Table 3.

³⁴The point estimates on the income coefficient in a semester by semester regression of semester grades on observable characteristics for those individuals who persist until seventh semester are .025, .021, .022, .027, .021, and .028 respectively. The associated p values are .054, .083, .080, .025, .112, and .048.

school quality was constructed under the Kentucky Instructional Results Information System (KIRIS) in order to “provide financial incentives for districts, schools, and teachers to make progress toward specific goals.”³⁵ We obtain a single school quality measure for each district by averaging the six yearly ratings. This single measure has a mean of 41.107 and a standard deviation of 2.807. The district ratings used here combine ratings for elementary, secondary, and high schools and involve weighted averages of things such as student test scores, attendance rates, retention rates, dropout rates, and the rate at which students “successfully” transition after graduation from high school. Thus, the rating for a particular district will capture both school quality and the ability/home learning environment of students in the district.

A \$10,000 increase in family income is estimated to increase a person’s school rating by .342 with an associated t-statistic of 5.941. Thus, students from lower income families do attend schools with somewhat lower ratings. Although the full results are not shown, the estimated coefficient and t-statistic associated with the school quality variable from a pooled regression involving all semester grades in all years for the KY subsample are .017 and 3.034 respectively.³⁶ Thus, higher school ratings are related to significantly higher college grades. Nonetheless, the estimated effect of family income on college grades remains large for the KY subsample when the school quality measure is included; the point estimate and t-statistic are .038 and 3.160 respectively.³⁷

Although this would seem to suggest that the effect of family income on attrition will remain

³⁵See *The Briefing Packet of the Kentucky School and District Accountability Results for Accountability Cycle 3*, Kentucky Department of Education, Wilmer S. Cody, Commissioner, 1998.

³⁶The set of observable characteristics includes the school quality variable and the set of variables in Table 3. There are 1117 individuals and 4569 total observations. Standard errors are adjusted to reflect the pooled nature of the data.

³⁷When the school quality measure is not included, the point estimate and t-statistic for the KY subsample are .043 and .3666 respectively.

strong even after taking into account the school ratings, there are plausible non-grade avenues through which the school ratings could affect student attrition. For example, students from schools with higher ratings will tend to have more friends or classmates who are also attending college, and, as a result, may view leaving college before completion to return home as less desirable than other students. In an attempt to capture both the effect that good schools have on college grades and these other types of effects, we estimated the duration model for the KY subsample including both the observable characteristics in column 1 of Table 2 and the school quality measure. Because the estimated effects of the majority of the variables are very similar to those in column 1 of Table 2, the entire set of results is not included. However, the coefficient and t-statistic associated with the school quality variable was found to be $-.034$ and 1.98 . Thus, students from inferior schools do leave more quickly even after taking into account the other observable characteristics. However, the estimated effect of family income remains very important and school quality is not found to account for much of the difference in attrition between students from different income situations. When the school quality measure is not included, the point estimate and t-statistic associated with family income for the KY subsample are $-.115$ and -3.40 respectively. When the school quality measure is included, the point estimate and t-statistic decline to $-.106$ and 3.11 respectively. Adding additional controls to the specification did not change the results in any important way.³⁸ Certainly it is possible that a non-trivial amount of measurement error exists in the school ratings.³⁹ Nonetheless, the results seem to suggest that families have a very strong direct

³⁸For example, when the duration specification in column 4 of Table 2 (which includes high school gpa) was estimated with the addition of the school quality variable and a variable measuring the interaction of school quality and high school gpa, no evidence was found that this interaction term was important (the point estimate and t-statistic associated with the interaction term were $.011$ and $.700$ respectively). As a result, in this case the findings regarding family income stayed the same as described earlier in this paragraph.

³⁹For example, it is likely that the ratings were not designed to exclusively measure college preparation.

effect on their children.

Negative income shocks to family income

In the introduction, an argument was made that negative income shocks may have differential retention effects on low income students, even if direct costs are zero. Until this point, the analysis has utilized the student's family income in the year that a student matriculated to college. However, the data also contain additional information about family income in the subsequent years of attendance. Thus, by adding a time varying "change in family income" variable, Δ income, and an interaction of this variable with the family income variable to the proportional hazard specification in the first column of Table 2 it is possible to some extent to examine whether negative income shocks influence the attendance decision when a full subsidy is in place and to examine to what extent the effect of income shocks differ by income. This exercise did not produce any evidence that negative income shocks have a significant effect on the timing of exits.⁴⁰

Although this suggests that differences between income groups are not being caused by differential responses to income shocks, it is important to note that measurement error, created by the timing of when family income is measured, will lead to parameter bias in this analysis. The income data come directly from the yearly FAFSA form which must be submitted by the student sometime between January 1 and April 15 and corresponds to a family's W2 income form from the previous year. We make the assumption that the first reported yearly income is the relevant income for the fall semester of the person's first year. We make the assumption that the second reported income is the relevant income for the spring semester of the first year and the fall semester of the second year. Similarly, the third reported income corresponds to the fourth and fifth semesters and the fourth reported income corresponds to the

⁴⁰The estimated effect (standard error) of the Δ income variable in this specification was -.010 (.023) and the estimated effect (standard error) of family income was -.098 (.028). Adding an interaction term between family income and Δ income did not change the nature of the results. The sample size was 2446.

sixth semester.

Essentially, the income values that we associate with the two semesters in a given calendar year come from the family's W2 form from the previous calendar year. To the extent that we are interested in the effect of income shocks, it seems likely that this timing assumption will lead to a downward bias for the estimator of the effect of income shocks with the extent of the bias depending on the extent to which income shocks tend to cause students to exit very quickly.⁴¹

V. A comparison to students who pay tuition

The total relationship between family income and college attrition that has been found in the literature is the sum of the portion due to the family environment explanation and the portion due to the direct costs explanation (assuming that selection bias has been accounted for or is not important). Section III suggests that the family environment effect is important. However, because tuition is zero for everyone at Berea, the Berea data cannot provide direct evidence about the importance of the direct costs explanation. In an effort to provide an idea of the size of the total relationship between family income and college outcomes (and, thus, a rough idea of the importance of the family environment explanation relative to the direct costs explanation), we estimate the duration model of attrition using data from the NELS-88. These data sustain continuing trend comparisons with the National Longitudinal Study of the High School Class of 1972 and the High School and Beyond (1980), which

⁴¹Under this timing assumption, an income will be missing in any case in which a person begins the second semester of a particular school year, fails to fill out the FAFSA form during the semester, and leaves school sometime before the start of the next school year. This will also potentially lead to a bias in the coefficient on income if those who fit this description are more likely to be leaving because of changes in family income. This source of bias can be completely removed by assuming that the correct income for the two semesters in a school year is the income that was reported on the FAFSA form in the January-April period before the school year started. However, this change increases the income measurement error due to the timing of income measurements. In practice, making this change made very little difference in the model estimates.

were used by Manski and Wise (1983) and Manski (1992), and are a logical choice because students who made normal progress through high school and entered college soon after graduation would have matriculated during the middle of the sample period covered by our Berea data.

There are 2823 students in the NELS-88 who entered a bachelor degree program at a four year (private or public) college in the fall of 1992 or the fall of 1993. In order to make the NELS-88 sample more similar in nature to the Berea sample, we remove the 684 individuals in the data who have a family income that is above \$85,800 in 1997 dollars. After also removing individuals with missing values of family income or the other characteristics that are used in the analysis, we are left with 1468 individuals.⁴²

The analogs to column 1 and column 2 of Table 2 are shown in Table 5.⁴³ In both cases, the estimated effects of family income are very similar to those in Table 2, despite the fact that the students

⁴²The variables used in this analysis are indicators for male and black, a math standardized test score, a reading standardized test score, and family income. Of the “income-eligible” students, 383 have a missing family income, 399 have a missing standardized reading score, and 397 have a missing standardized math score.

⁴³The family income variable in the NELS-88 is categorical. To estimate the analog of column 2 of Table 2 (in which family income is categorical), we define the low income variable to include all incomes less than 17160 and we define the high income variable to include all incomes greater than 40040.

To estimate the analog of column 1 of Table 2 (in which family income enters as a continuous variable), because the exact income value is not observed, we first fit a lognormal distribution to the categorical data (using both the individuals in the sample used to estimate Table 5 and the individuals that were removed from the final sample due to higher family incomes). We then compute the likelihood contribution for person i by integrating the likelihood contribution conditional on the person’s family income (from Appendix A) over the appropriate distribution of the person’s family income given the estimated lognormal distributions and the person’s family income category. We found very similar results (i.e., point estimate and t-statistic of -.071 and -2.95 respectively) when we estimated the continuous case by simply making the assumption that the family income of each person in a particular income class is equal to the midpoint of that class.

Estimation utilizes sampling weights. See Grogger and Neal (2000) for a discussion of the NELS-88 sampling scheme and the importance of using weights.

in the NELS-88 are attending tuition charging institutions.⁴⁴ Thus, although one should be very careful about the conclusions that can be drawn from this comparison, the exercise suggests that family environment factors may be the driving force in determining the strong relationship between family income and educational outcomes. However, it is important to note that the effect of direct costs under current tuition/financial aid programs is influenced by both the possibility that low income students find paying direct costs burdensome for reasons such as liquidity constraints and the reality that students from low income families are likely to face lower direct costs due to the existence of need-based financial aid. The former suggests that low income students will tend to leave school more quickly. The latter suggests that low income students will leave school less quickly. Thus, even if the total effect of direct costs on college outcomes is close to zero, this does not imply that liquidity constraints are unimportant for low income students.

It seems possible that many of the possible reasons related to the family environment effect would be as closely related to parental education backgrounds as they would be to family income per se. Unfortunately, information regarding parental education is not available in our Berea data. However, this information is available in the NELS-88. When we estimated the specification in column 1 of Table 5 with an additional indicator of whether the student had at least one parent who had obtained

⁴⁴ Figure 4 shows that the predicted survivor function at t=3 (probability of finishing more than 3 full semesters) at Berea is .67 if the person has family income of \$40,000 and is .58 if the person has family income of \$5,000. For the NELS-88 sample, these numbers are .72 and .63 respectively.

The upper income bound of \$85,800 was chosen to try to make the sample somewhat similar to the Berea sample while still retaining a reasonable sample size. Because the income data is categorical, the next option would be to draw the cutoff at \$57,200. Estimating the model with this cutoff produced an estimated effect and t-statistic associated with family income of -.047 and -1.37 respectively. We also tried restricting the sample to individuals attending colleges with similar enrollments at Berea. Estimating the model with individuals at schools with a total enrollment of less than 5441 (n=482) produced an estimated effect and t-statistic associated with family income of -.105 and -2.43 respectively. Estimating the model with an enrollment cutoff of 1898 (n=175) produced an estimated effect and t-statistic associated with family income of -.138 and -1.92 respectively. Trying to simultaneously restrict both enrollment size and further restrict family income was not possible due to the small sample size that resulted.

at least a college degree we found that this variable was statistically significant with a point estimate of $-.36$ and a t -statistic of -3.02 . The estimated effect of family income declined to $-.062$ but remained statistically significant with a t -statistic of -2.859 .

VI. Conclusion

Bowen and Bok (1998) raise the possibility that low national graduation rates may be due to the inability of students and their families to meet college costs. The high overall attrition rates at Berea College, where all students receive a full tuition subsidy (and pay an average of approximately \$1,000 for room, board, and college fees), suggest that college exits often occur for reasons that are unrelated to the direct costs of college. It is important to note that it is necessary to be cautious about drawing inference about more general populations of interest on the basis of a single school. Nonetheless, given the statistically significant and quantitatively large relationship that is found between family income and performance at Berea, the discussion of the possible sources of bias that might be present in our Berea estimator that appears in Stinebrickner and Stinebrickner (2000), and the comparison to outcomes in the NELS-88, this work suggests that reasons related to family environment are the most important determinants of the differences in college outcomes by family income that have consistently been found in the literature.

The results in this paper suggest that non-trivial differences in educational attainment would exist even if direct costs were zero for all students. However, it is important to note that equality in educational outcomes between income groups could potentially be achieved by removing (reducing) the direct costs for only low income students. This is a feature of the California tuition subsidy program that will begin full-scale operation in the fall of 2001. However, if equality in educational outcomes is achieved in cases like these, it may be important for policymakers to realize that this does not necessarily occur simply because these programs address imperfections in capital markets which imply

that low income individuals face liquidity constraints. It seems likely that lowering the costs of college will make staying in school (or entering school) optimal for some non-liquidity-constrained low income individuals who would have otherwise found it optimal to leave (or not enter) after realizing that they were not well-prepared academically for college. From the standpoint of wisely using educational budgets, more research is necessary to understand the potential effectiveness of programs which would improve the educational opportunities of low income individuals before they reach college or would address non-cost difficulties that low income individuals face after arriving at college.

References

- Belman, Dale and Heywood, John S. "Sheepskin Effects in the Returns to Education: An Examination on Women and Minorities." *Review of Economics and Statistics*, November 1991, 73(4), pp. 720-24.
- Belman, Dale and Heywood, John S. "Sheepskin Effects by Cohort: Implications of Job Matching in a Signalling Model." *Oxford Economic Papers*, October 1997, 49(4), pp. 623-37.
- Betts, Julian and Merrall, Darlene. "The Determinants of Undergraduate Grade Point Average: The Relative Importance of Family Background, High School Resources, and Peer Group Effects." *The Journal of Human Resources* (forthcoming).
- Bound, John and Johnson, George. "Changes in the Structure of Wages in the 1980's: An Evaluation of Alternative Explanations, *The American Economic Review*, June 1992, pp. 371-392.
- Bowen, William and Bok, Derek. *The Shape of the River*. Princeton, New Jersey; Princeton University Press, 1998.
- Cameron, Stephen V and Heckman, James J. "Life Cycle Schooling and Dynamic Selection Bias: Models and Evidence for Five Cohorts of American Males." *Journal of Political Economy*, April 1998, pp. 262-333.
- Cameron, Stephen V and Taber, Christopher. "Borrowing Constraints and the Returns to Schooling." unpublished manuscript, 1999.
- Card, David and Krueger, Alan. "Does School Quality Matter? Returns to Education and the Characteristics of Public Schools in the United States." *Journal of Political Economy*, February 1992, pp. 1-42.
- Cohn, E. and Kiker, B.F. "Socioeconomic Background, Schooling, Experience and Monetary Rewards in the United States. *Economica*, November 1986, 53, pp. 497-503.
- Daniel, Kermit, Black, Dan, and Smith Jeffrey. "College Quality and the Wages of Young Men." 1997, Unpublished Manuscript
- Daniel, Kermit, Black, Dan, and Smith Jeffrey. "College Characteristics and the Wages of Young Women." 1995, Unpublished Manuscript
- Grogger, Jeff, and Neal Derek. "Further Evidence on the Benefits of Catholic Secondary Schooling." *Brookings-Wharton Papers on Urban Affairs*, 2000, vol 1. pp. 151-193.
- Hauser, R.M. "Socioeconomic Background and Differential Returns to Education." in L. Solomon editor, *Does College Matter*. New York: Academic Press. 1973.
- Haveman, Robert H. and Wolfe, Barbara. "Schooling and Economic Well-Being: The Role of Nonmarket Effects." *Summer* 1984, 19(3).

- Heckman, J. and Singer, B. "A Method for Minimizing the Impact of Distributional Assumptions in Econometrics Models for Duration Data" *Econometrica*, vol. 52, (1984): 271-320.
- Heckman, James, Lochner, Lance, and Taber, Christopher. "Tax Policy and Human-Capital Formation." *American Economic Review*, May 1998, 88(2), pp. 293-297.
- Heckman, James, Lochner, Lance, and Taber, Christopher. "General-Equilibrium Treatment Effects: A study of Tuition Policy." *American Economic Review*, May 1998, 88(2), pp. 381-386.
- Heywood, John S. "How Widespread are Sheepskin Returns to Education in the U.S.?" *Economics of Education Review*, September 1994, 13(3), pp. 227-34.
- Hungerford, Thomas, and Solon, Gary. "Sheepskin Effects in the Returns to Education," *Review of Economics and Statistics*, 69(1), pp. 175-177.
- Jaeger, David A. and Page, Marianne E. "New Evidence on Sheepskin Effects in the Returns to Education." *Review of Economics and Statistics*, November 1996, 78(4), pp. 733-40.
- Johnson, William R. "Distributional Issues in the Public Support of Higher Education." Unpublished Manuscript.
- Kane, Thomas J. "College Entry by Blacks since 1970. The Role of College Costs, Family Background, and the Returns to Education, *Journal of Political Economy*, October 1994, 102(5), pp. 878-911.
- Kane, Thomas J. and Rouse, Cecilia. "Labor Market Returns to Two and Four Year College." *American Economic Review*, June 1995, 85(3), 600-614.
- Katz, Lawrence and Murphy, Kevin. "Changes in Relative Wages, 1963-1987: Supply and Demand Factors." *The Quarterly Journal of Economics*, February 1992, 107(1), pp. 35-78.
- Keane, Michael and Wolpin, Kenneth. "The Effect of Parental Transfers and Borrowing Constraints on Educational Attainment." unpublished manuscript, 1998.
- Kiker, B.F. and Condon, C.M. (1981). "The Influence of Socioeconomic Background on the Earnings of Young Men." *Journal of Human Resources*, 1981, 16, pp. 442-448.
- Lavy, V., Palumbo, Michael and Stern Steven, "Simulation of Multinomial Probit Probabilities and Imputation of Missing Data," *Advances in Econometrics*, eds., Thomas Fomby and R. Carter Hill, JAI Press, 1998.
- Lazear, Edward. "Education: Consumption or Production?" *Journal of Political Economy*, June 1977, 85, 569-598.
- Little, R., and D. Rubin, *Statistical Analysis with Missing Data*, New York; Wiley and Sons, 1987.
- Loury, Linda Datcher and Garman, David. "College Selectivity and Earnings," *Journal of Labor Economics*, April 1995, 13(2), pp. 335-350.
- Manski, Charles. "Parental Income and College Opportunity." Democratic Study Center Report, Washington, DC. August, 1992.

- Manski, Charles. "Adolescent Econometricians: How Do Youth Infer the Returns to Schooling?" in Clotfelter, Charles T. and Rothschild, Michael eds. *Studies of Supply and Demand in Higher Education*. A National Bureau of Economic Research Project Report. Chicago and London; University of Chicago Press, 1993, pp. 43-57.
- Manski, Charles and Wise, David. *College Choice in America*. Cambridge, MA and London England; Harvard University Press, 1983.
- Murphy, Kevin M and Welch, Finis. "The Structure of Wages." *The Quarterly Journal of Economics*, February 1992, 107(1), pp. 35-78.
- Park, Jin Heum. "Estimation of Sheepskin Effects and Returns to Schooling Using the Old and the New CPS Measures of Educational Attainment." Princeton Industrial Relations Section Working Paper: 338, December 1994, pp. 1-9.
- Patrinos, Harry A. "Socioeconomic Background, Schooling, Experience, Ability and Monetary Rewards in Greece." *Economics of Education Review*: 14(1), pp. 85-90 (1995).
- Renkow, Mitch. "Income Non-convergence and Rural-Urban Earnings Differentials: Evidence from North Carolina: 62(4), pp. 1017-1028 (1996).
- Shea, J. "Does Parent's Money Matter?" 1996, University of Maryland Working paper.
- Stern, S., and Todd, P. "A Test of Lazear's Mandatory Retirement Model," 1998, Unpublished Manuscript (1995).
- Stinebrickner, Ralph. "A Report on Retention at Berea College," 1998
- Stinebrickner, Todd R. "Estimation of a Duration Model in the Presence of Missing Data," *The Review of Economics and Statistics*, 1999.
- Stinebrickner, Todd and Stinebrickner, Ralph, "Understanding the Relationship Between Family Income and Schooling Attainment." University of Western Ontario working paper, 2000.
- Stinebrickner, Todd R. and Stinebrickner, Ralph. "Working During School and Academic Performance," forthcoming, *Journal of Labor Economics*.
- Taubman, Paul. "Role of Parental Income in Education Attainment." *The American Economic Review*: May 1989, 79(2), pp. 57-61.

Table 1 - Data description- full Berea sample and Berea sample divided into income thirds
n=2821

	Full Sample mean std dev.	Lowest 1/3 income mean std dev.	Middle 1/3 income mean std. dev.	Highest 1/3 income mean std. dev.
income/10000	2.245 1.359	.767 .527	2.201 .345	3.770 .792
male	0.454	0.446	0.439	0.476
black	0.1	0.131	0.087	0.082
verbal ACT	22.172 4.361	21.887 4.335	22.143 4.180	22.487 4.544
math ACT	20.410 3.859	20.123 3.828	20.385 3.854	20.723 3.877
distance from home - close	0.39	0.388	0.38	0.415
number in family	4.050 1.516	3.728 1.515	3.827 1.31	4.595 1.560

Table 2 Proportional hazard model of attrition - Berea sample

	n=2821	n=2821	n=2821	n=2442
variable	estimate SE	estimate SE	estimate SE	estimate SE
male	.242* (.067)	.240* (.067)	.243 (.067)	.068 (.073)
black	-.157 (.104)	-.149 (.104)	-.152 (.105)	-.306* (.123)
high school gpa				-.616* (.094)
verbal ACT	-.017* (.008)	-.017* (.008)	-.017 (.008)	-.015 (.009)
math ACT	-.051* (.010)	-.051* (.010)	-.051* (.010)	-.030* (.010)
family size	.004 (.020)	.001 (.020)	.016 (.023)	.014 (.023)
distance from home - close	-.204* (.065)	-.202* (.065)	-.202* (.065)	-.118 (.072)
income/10000	-.083* (.024)			-.095* (.027)
indicator for income in bottom 1/3		.243* (.079)		
indicator for income in middle 1/3		.201* (.078)		
indicator for income in bottom 1/6			.323* (.112)	
income in 2 nd 1/6			.329* (.112)	
income in 3 rd 1/6			.291* (.112)	
income in 4 th 1/6			.222* (.112)	
income in 5 th 1/6			.127* (.112)	
var. of heterogeneity	.297 (.605)	.284 (.078)	.310 (.469)	.511 (.298)
t=1	-2.223* (.184)	-2.335* (.199)	-2.426* (.218)	-2.187* (.183)
t=2	-1.699* (.142)	-1.813* (.160)	-1.900* (.180)	-1.595* (.142)
t=3	-2.381* (.133)	-2.497* (.149)	-2.581* (.168)	-2.263* (.144)
t=4	-2.129* (.126)	-2.246* (.140)	-2.328* (.158)	-1.988* (.143)
t=5	-2.784* (.159)	-2.902* (.168)	-2.982* (.183)	-2.609* (.179)
t=6	-2.822* (.170)	-2.940* (.179)	-3.020* (.190)	-2.733* (.195)
log likelihood	-3484.936	-3485.801	-3484.61	-3009.676

* t statistic greater than 2.0

The first three columns are models estimated without high school grade point averages.

The fourth column is the model estimated with high school grade point averages. Sample size is smaller in fourth column because individuals with missing high school grades are not included.

Table 3
 Regression of College Grades in First Semester - Berea sample
 for all students who completed first semester, n=2661

	GPA first semester
constant	.792* (.103)
male	-.233* (.031)
black	-.177* (.051)
verbal act	.040* (.004)
math act	.039* (.004)
income/10000	.040* (.004)
	n=2312 R2=.246

Table 4 Proportional hazard model of attrition including cumulative college grade point average - Berea sample
n=2649

	n=2649
	estimate SE
male	.001(.071)
black	-.305* (.119)
verbal ACT	.026* (.009)
math ACT	.005 (.010)
number in family	-.0004 (.024)
distance from home - close	-.154* (.072)
income/10000	-.037 (.026)
cumulative college gpa at time t	-1.630* (.052)
t=1	.660* (.167)
t=2	2.372* (.78)
t=3	1.894* (.192)
t=4	2.315* (.194)
t=5	1.771* (.216)
t=6	1.751* (.227)
log like value	-2606.958

* represents t statistic greater than two.

Sample size is smaller than the sample size in Table 2 because the 172 students who did not stay in school long enough to receive first semester grades are not used in the analysis.

Table 5 Proportional hazard model of attrition - NELS-88

	n=1468	n=1468
variable	estimate SE	estimate SE
male	.308 (.084)*	.319* (.084)
black	-.132 (.124)	-.085 (.125)
verbal standardized test	-.002 (.006)	-.002 (.006)
math standardized test	-.047* (.006)	-.047* (.006)
income/10000	-.092* (.021)	
indicator for income in bottom 1/3		.347* (.132)
indicator for income in middle 1/3		.227* (.090)
t=1	-.652* (.214)	-1.141* (.221)
t=2	.237 (.218)	-.255 (.229)
t=3	-.509* (.226)	-1.005* (.232)
log likelihood	-1224.441	-1226.716

* t statistic greater than 2.0

Appendix: Proportional hazard model of attrition with Heckman-Singer heterogeneity- Berea sample

	n=2821
variable	estimate SE
male	.237* (.067)
black	-.146 (.106)
verbal ACT	-.018* (.008)
math ACT	-.056* (.010)
family size	.009 (.021)
distance from home - close	-.230* (.068)
income/10000	-.084* (.024)
type 1 for ϵ_i #	-4.178 (3.056)
type 2 for ϵ_i #	0.000
probability ϵ_i =type 1#	.955
probability ϵ_i =type 2#	.045
t=1	1.547 (3.283)
t=2	2.473 (3.092)
t=3	1.779 (3.069)
t=4	2.020 (3.068)
t=5	1.357 (3.069)
t=6	1.313 (3.056)
log likelihood	-3484.05

* t statistic greater than 2.0

Results are analogous to those in column 1 of Table 2, but specification involves Heckman-Singer flexible form for unobserved heterogeneity.

#type 2 for ϵ_i is normalized to zero. To ensure that $\Pr(\epsilon_i = \text{type 1})$ and $\Pr(\epsilon_i = \text{type 2})$ are nonnegative and sum to one we define $\Pr(\epsilon_i = \text{type 1}) = e^\gamma / (e^\gamma + e^0)$, $\Pr(\epsilon_i = \text{type 2}) = e^0 / (e^\gamma + e^0)$ and estimate γ . Point estimate (std error) of γ are 3.048* (.854).

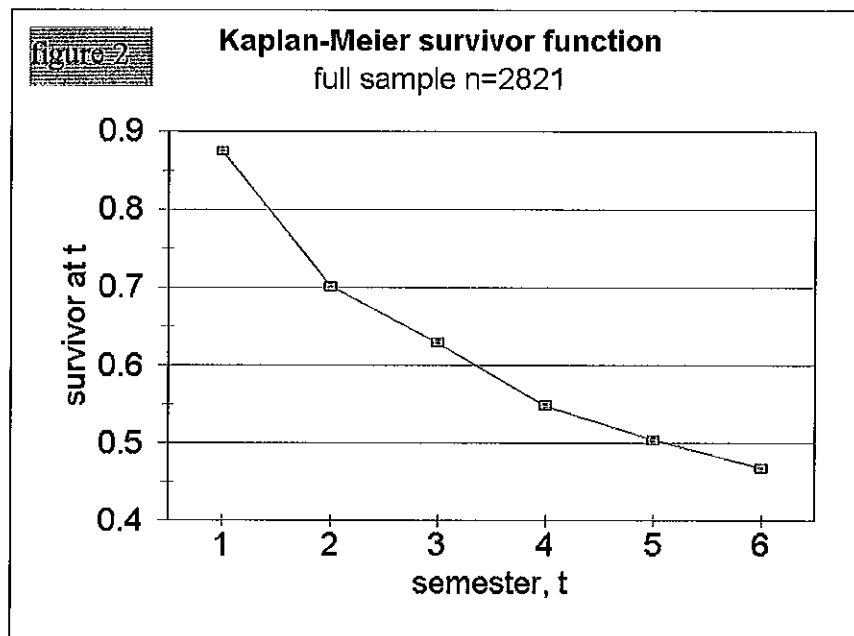
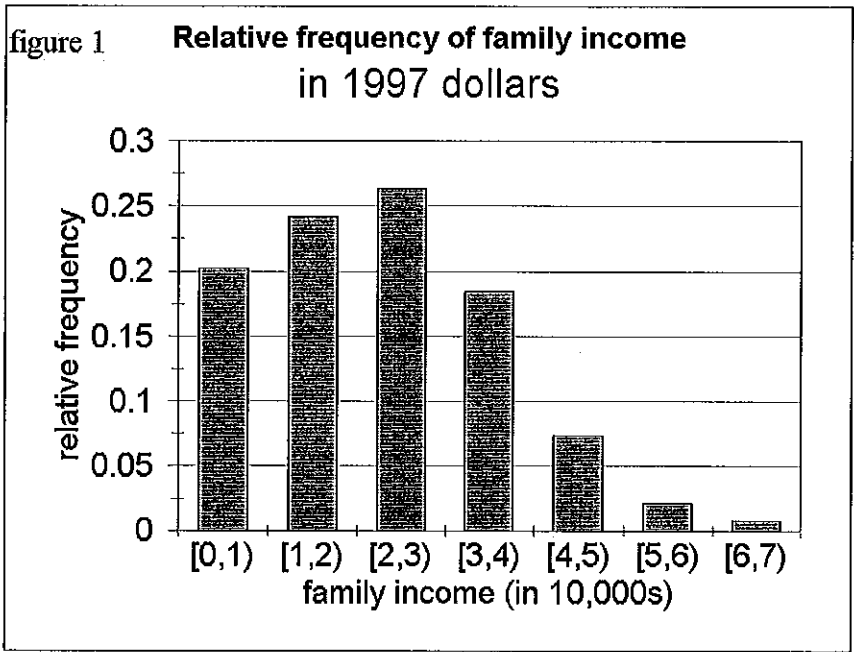


figure 3

Kaplan-Meier survivor functions
for three income groups

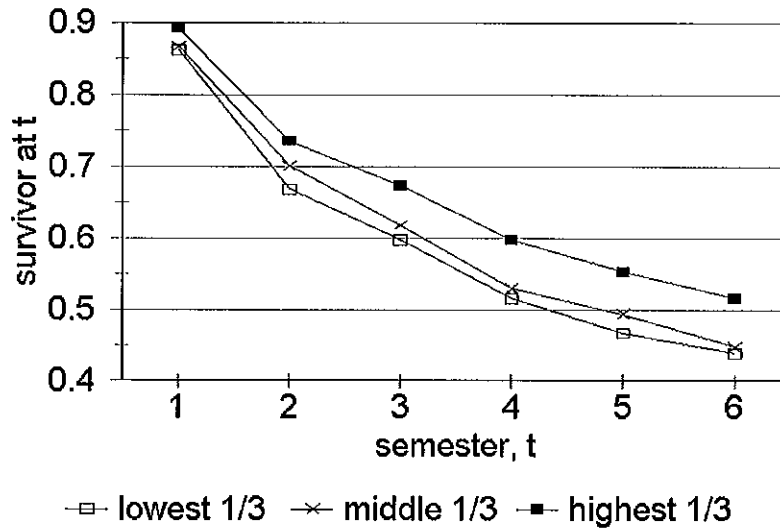


figure 4

Predicted Survivor Functions
for baseline person by income

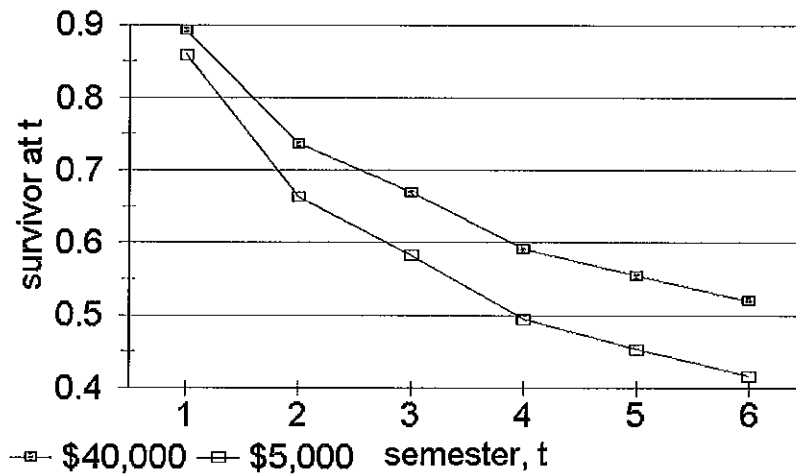


figure 5

Predicted survivor functions
for three income groups

