

Stinebrickner, Todd R.; Stinebrickner, Ralph

**Research Report**

## The relationship between family income and schooling attainment: Evidence from a liberal arts college with a full tuition subsidy program

Research Report, No. 2000-8

**Provided in Cooperation with:**

Department of Economics, University of Western Ontario

*Suggested Citation:* Stinebrickner, Todd R.; Stinebrickner, Ralph (2000) : The relationship between family income and schooling attainment: Evidence from a liberal arts college with a full tuition subsidy program, Research Report, No. 2000-8, The University of Western Ontario, Department of Economics, London (Ontario)

This Version is available at:

<https://hdl.handle.net/10419/70370>

**Standard-Nutzungsbedingungen:**

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

**Terms of use:**

*Documents in EconStor may be saved and copied for your personal and scholarly purposes.*

*You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.*

*If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.*

The relationship between family income and schooling attainment:  
Evidence from a liberal arts college with a full tuition subsidy program

Todd R. Stinebrickner and Ralph Stinebrickner<sup>1</sup>

Researchers have long been interested in understanding why a strong relationship between family income and educational attainment exists at virtually all levels of schooling. 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 individuals from low income families are less likely to graduate from college than other students. Using unique new data obtained directly from a liberal arts school that maintains a full tuition subsidy program, this paper provides direct evidence that family environment reasons that are unrelated to the tuition costs of college are very important. The paper pays close attention to the issue of selection bias by deriving a set of seemingly very plausible conditions under which the estimator of interest is “conservative.” The findings, which suggest that non-trivial differences in educational attainment would exist even if tuition was zero for all students, have implications for expensive policy programs such as the full tuition subsidy program that was recently approved by the state of California.

---

<sup>1</sup>Affiliations: The University of Western Ontario and Berea College respectively. Comments very welcome. Please direct correspondence to Todd Stinebrickner at [trstineb@julian.uwo.ca](mailto: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, Jeff Smith, Dan Black, Chris Swann, Ben Scafidi, Sarah Turner, Susan Dynarski, and seminar participants at the University of Michigan, the University of Virginia, the University of Western Ontario, Queens University, York University, UC Irvine, the Upjohn Institute, and a meeting of the NBER higher education group. The first author is grateful for assistance from the Agnes Cole Dark fund. The second author was supported in part by Berea College through funding for a sabbatical leave.

## **Abstract**

Researchers have long been interested in understanding why a strong relationship between family income and educational attainment exists at virtually all levels of schooling. 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 individuals from low income families are less likely to graduate from college than other students. Using unique new data obtained directly from a liberal arts school that maintains a full tuition subsidy program, this paper provides direct evidence that family environment reasons that are unrelated to the tuition cost of college are very important. The paper pays close attention to the issue of selection bias by deriving a set of seemingly very plausible conditions under which the estimator of interest is “conservative.” The findings, which suggest that non-trivial differences in educational attainment would exist even if tuition was zero for all students, have implications for expensive policy programs such as the full tuition subsidy program that was recently approved by the state of California.

## **I. Introduction**

Researchers have long been interested in understanding why a strong relationship between family income and educational attainment exists at virtually all levels of schooling. 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 individuals from low income families are less likely to graduate from college.<sup>2</sup> For respondents in the High School and Beyond survey (HS&B) who were high school seniors in 1980, the raw data in Manski (1992) show that the probability of obtaining a bachelor degree from a four year college within five and a half years after high school is .39 for individuals in the highest income quintile, .24 for individuals in the middle income quintile, and .11 for individuals in lowest income quintile.

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 work has specifically examined the college entrance decision, the reality that approximately half of all matriculating students do not graduate suggests that examining what happens after students arrive at college is also important.<sup>3</sup> Evidence in Manski and Wise (1983) and Manski (1992) indicates that the relationship between family income and college entrance and the relationship between family income and college persistence are both of importance in determining the overall relationship between family income and college graduation. For example, when college entrance rates and college attrition rates are calculated including only individuals who begin college at four year schools, roughly 51% of the HS&B college graduation gap between students in the lowest and highest income quintiles and 48% of the HS&B college graduation gap between students in the lowest

---

<sup>2</sup>Among others, see Bound and Johnson (1992), Katz and Murphy (1992), and Murphy and Welch (1992).

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

and middle income quintiles can be attributed to differences in college attrition rates between the groups. When college entrance rates and college attrition rates are calculated including individuals who begin college at both two and four year schools, roughly 71% of the graduation gap between students in the lowest and highest income quintiles and 65% of the graduation gap between students in lowest and middle income quintiles can be attributed to differences in college attrition rates between the groups.<sup>4</sup>

Thus, understanding why students from low income families have worse college outcomes than other students is important from a policy standpoint. An explanation along traditional lines is that differences by family income arise largely for “tuition reasons” (e.g., liquidity constraints) that are related to the burden of paying for college. For example, when discussing differences in college outcomes by family income, the findings of Manski 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 might have to be offset by external funds...” The potential importance of this explanation was also raised more recently by Bowen and Bok (1998): “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.”

An alternative to the tuition explanation is that the relationship between family income and college outcomes arises due to “family environment reasons” that would exist even if the tuition costs of

---

<sup>4</sup>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, including students who enter two-year institutions (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.

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*.

college were zero for all students.<sup>5</sup> 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.<sup>6</sup> It is important to note that the family environment explanation also potentially includes reasons related to families' financial circumstances that would be present even if tuition costs were zero. For example, even in the presence of a full tuition subsidy, negative shocks to family income may contribute to retention differences between income groups if students from low income families are more likely to return home to help their parents in bad economic times.<sup>7</sup>

Separating the relative importance of the “family environment explanation” and the “tuition explanation” for differences in college outcomes by family income is important from a policy standpoint 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

---

<sup>5</sup>The 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.”

<sup>6</sup>Many 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).

<sup>7</sup>In 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.

due to borrowing constraints (Taubman, 1989).<sup>8</sup> Unfortunately, determining how much of the difference in college outcomes by family income would remain if tuition costs were zero is typically a very difficult empirical task because when tuition costs are not zero, college outcomes 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 of tuition costs.

In an effort to provide direct information about the importance of the family environment explanation, this paper takes advantage of unique data obtained 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 in Section III and throughout the paper, there are numerous features of the school and our data that are desirable from the standpoint of this study. However, of particular interest given the nature of this study is the fact that Berea College operates with a mission of providing an education to those who “have great promise, but limited economic resources” and provides a full tuition subsidy to all entering students (and large room and board subsidies) regardless of family income.<sup>9</sup>

This unique feature allows an opportunity to provide direct evidence about the importance of the family environment explanation described above.<sup>10</sup> In particular, if the sample of students who are

---

<sup>8</sup>Kane (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.

<sup>9</sup>According to the Berea College 1998 admissions brochure, entering students at Berea have a room, board, and college fee bill of only approximately \$1000. Students graduate from Berea with an average of approximately \$1000 in student loans.

<sup>10</sup>Roughly speaking, the direct costs of college can be thought of as the difference between the total cost of tuition, fees, books, travel to school, housing, and food if the person attends college and the total cost of housing and food if the person does not attend college. Thus, in general, the direct costs of college can be slightly positive even when tuition is zero (especially if students who do not attend college tend to live at home for a period after high school). At Berea tuition costs are zero. Moreover, the

observed entering college is “representative” of the population of interest (in a sense that will be described more fully in the next section), evidence about the importance of the family environment explanation can be obtained simply by comparing the outcomes of individuals in different income groups (and taking into account potential differences in factors such as ability). The estimates from a duration model of attrition in Section IV indicate that, even though tuition costs are zero, a strong positive relationship exists between family income and the length of time that an individual remains in college. For example, students in the bottom third of the income distribution at Berea are approximately twenty percent more likely to drop-out of school before the third year than students in the upper third of the income distribution.

Of course, it is important to keep in mind that the relationship that is observed between family income and college performance for those who enter this college may be influenced by selection bias if the sample of those who attend this college is not representative of the population of interest. In general, it will always be difficult to rule out with certainty the possibility that selection effects contribute to the observed relationship between college performance and family income. However, when tuition costs are zero, it becomes easier to understand the conditions under which selection bias is potentially problematic. A simple theoretical model in Section II shows that the presence of zero tuition costs removes one likely source of upward bias and allow the conditions, under which either an accurate or a conservative estimate of the family environment effect is obtained, to become clearly apparent. This is important because, if, as previous research suggests, it is reasonable to believe that these conditions are satisfied, the observed relationship between family income and college outcomes at Berea suggests that the family environment explanation

---

additional room and board subsidy implies that the direct costs of college attendance at Berea are also approximately zero. For simplicity, we continue to refer to the alternative to the family environment explanation as the tuition explanation (although the discussion suggests that we could also call it the “direct costs explanation”). We also note that, although opportunity costs (e.g., foregone earnings) influence college entrance and college attrition, it is the direct costs of college that are typically the focus of education policy.



is very important.

Because no students at Berea pay tuition, the paper cannot provide direct evidence about the importance of the tuition explanation. However, in an effort to provide a rough idea of the importance of the family environment explanation relative to the tuition explanation, the duration model of attrition is estimated in Section VI 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 been found between family income and college outcomes.

In Section V, the paper explores several possible reasons that family environment is found to have such a strong effect on college outcomes. Students from low income families are found to receive significantly lower college grades (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 their children.

In Section VII, 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.

## **Section II. A Simple Theoretical Model**

We are interested in the effect that family environment has on college performance for a population of interest that we refer to generically as the group of all “college eligible” high school graduates. Suppose that the probability of college graduation for a person in this population is given by  $P^G(Y,X,M)$  where  $Y$  is the family income of the student,  $X$  is a set of observable characteristics of the student (e.g., measures of ability, sex, and race), and, as will be discussed in more detail,  $M$  is the student’s “motivation.” In practice, what is typically observed (e.g., in longitudinal survey data) is the relationship between  $P^G$  and  $Y$  (perhaps after “controlling” for the effect of  $X$ ) for a sample of individuals who choose to enter college given current tuition costs and financial aid offers. In these cases, the observed relationship between  $P^G$  and  $Y$  will be influenced by the family environment effect, the tuition effect, and any selection effects that are present due to the college entrance decisions of students or due to the admission decisions of colleges. A desirable feature of our data from Berea is that the tuition effect is entirely removed. Thus, if any selection effects that are present imply that the sample will produce an accurate or downwardly biased estimate of the relationship between family income and college outcomes that exists in the population of interest, the sample estimate from Berea can be viewed as an accurate or conservative estimate of the family environment effect.

In this section, we begin by examining the college entrance decision in general and later examine the admission process and student entrance decision associated with this specific school. Consider the case of a typical student who faces non-zero tuition costs. If the student does not face liquidity constraints, and, therefore, can borrow freely (from parents or financial institutions) to finance the costs of college, a simple model of the college entrance decision would suggest that the individual decides to attend college if the

expected benefits from entering college outweigh the expected benefits from choosing not to attend,

$$(1) \hat{P}^G W^G + [1 - \hat{P}^G] W^N - C \geq W^N.$$

Or, upon rearranging,

$$(2) \hat{P}^G \geq C/[W^G - W^N]$$

where  $\hat{P}^G$  is the individual's estimate of the probability of college graduation at the time he makes his decision of whether to attend college,  $W^G$  is the person's expected discounted lifetime wages if he graduates from college,  $W^N$  is the person's expected discounted lifetime wages if he does not graduate from college, and  $C$  is the money cost of college attendance.<sup>11</sup> For simplicity, the model does not differentiate between the lifetime wages that a person receives if he does not start college and the lifetime wages that he receives if he drops out of college.<sup>12</sup> The model also generally abstracts from issues associated with the effect that college has on non-pecuniary utility, but implicitly allows the non-pecuniary utility associated with the college years to enter the college decision through its effect on the (perceived) probability of

---

<sup>11</sup>  $\hat{P}^G$  could be different than  $P^G$  if the person's estimates of his probability of graduating are incorrect. This possibility will be discussed. See Manski (1993) for a discussion of the difficulties associated with modelling high school graduates' expectations.

<sup>12</sup> This simplifying assumption implies that, for a person who drops out, the time in college does not significantly change his discounted lifetime income. The motivation for this assumption is simply to keep the model as transparent as possible while retaining the most important aspects of the college entrance decision. However, to some extent, the assumption can be motivated in this application by the reality that most students in these data who do not graduate tend to leave college quickly (so that foregone earnings are relatively small) and by evidence on the importance of sheepskin effects found by Park (1994), Jaeger and Page (1996), Hungerford and Solon (1987), Heywood (1994), and Belman and Heywood (1991, 1997) among others (so that increases in future earnings associated with completing a small amount of college are relatively small). See also Kane and Rouse (1995) who reach somewhat different conclusions about sheepskin effects. Further, any increase in future earnings which does occur while a student is obtaining "some college" would be offset to some extent by the earnings which are foregone while the person is in college.

If  $C$  represents only the tuition cost of college attendance, the model above has the undesirable property of suggesting that all students should enter college if tuition costs are zero. However, this problem is removed if we imagine that  $C$  captures the non-pecuniary or pecuniary costs of moving to a new location to attend school (e.g., the cost of finding new friends).

graduation. Those who find college unenjoyable will tend to be less likely to graduate.

Motivation,  $M$ , is considered explicitly as a separate determinant of  $P^G$  to allow us to distinguish performance differences between income groups that are observed simply because of sample selection from performance differences that are observed because of inherent differences that exist between income groups in the population of interest. In order to facilitate the discussion of the cases where selection bias is present, we define  $M$  as to capture everything about an individual's propensity or desire to succeed in college that is uncorrelated with family income in the population. Thus, if the low income students in a sample of college attendees tend to have lower levels of  $M$  than the high income students in the sample of college attendees, selection bias will lead to an overstatement of the relationship between family income and college performance that exists in the population of interest.<sup>13</sup> On the other hand, if the low income students in a sample tend to have levels of  $M$  that are at least as high as those of the high income students in a sample, selection bias will lead to an accurate or downwardly biased estimate of the relationship between family income and college performance that exists in the population. We are particularly interested in the conditions under which the latter situation occurs, because, as discussed earlier, this case implies that the relationship between family income and college performance that is observed in our sample at Berea will be an accurate or conservative estimate of the family environment effect.

Assume for the time being that the right side of equation (2) does not vary by family income and condense notation by renaming the right side of equation (2)  $C^*$ . Equation (2) simply says that an individual enters college if his estimate of the probability of graduation is greater than a threshold  $C^*$  that

---

<sup>13</sup>In this case, the college entrance decision creates a positive correlation between family income and motivation. As a result, because motivation is not observed in empirical work, the relationship between family income and performance that is observed for individuals who enter college will capture both the effect of income on performance that is present in the population and a portion of the effect that unobserved motivation has on performance.

does not depend on family income.<sup>14</sup> This implies that the presence and nature of selection effects depend on how estimates of the probability of graduation vary with  $Y$ , holding  $X$  and  $M$  constant. Given previous literature about the relationship between  $P^G$  and  $Y$ , it seems unlikely that individuals from low income families (who are less likely to have attended high quality elementary and secondary schools and are less likely to have parents and friends who have graduated from college) would have higher estimates of the probability of graduation than individuals (with the same values of  $X$  and  $M$ ) from higher income families. Instead, it seems that the plausible case to consider is  $\partial \hat{P}^G / \partial Y \geq 0$ .<sup>15</sup>

A person from a “high income” family with income  $Y^H$  will choose to enter college if

$$(3) \quad \hat{P}^G(Y^H, X, M) \geq C^*.$$

To see that the marginal low income student who enters college has a level of motivation that is at least as high as that of the marginal high income person that enters college, define  $M^H$  to be the motivation level such that equation (3) holds with equality, in which case the person with characteristics  $X$  and income  $Y^H$  is indifferent between attending and not attending college. Then, the fact that a student from a lower income family,  $Y^L < Y^H$ , will only enter college if he has a motivation level that is at least as high as  $M^H$  follows immediately from  $\partial \hat{P}^G / \partial Y \geq 0$ ,

$$(4) \quad \hat{P}^G(X, Y^L, M^H) \leq \hat{P}^G(X, Y^H, M^H) = C^*.$$

Therefore, under the assumption that  $C^* = C / (W^G - W^N)$  is constant across income groups, the simple

---

<sup>14</sup>Note that, although observable characteristics  $X$  are likely to influence  $C^*$ , we do not explicitly discuss this possibility because it does not directly influence the conclusions which are drawn from our theoretical model.

<sup>15</sup>If individuals’ beliefs reflect the true relationship between  $P^G$  and  $Y$  then  $\partial \hat{P}^G / \partial Y = 0$  if  $\partial P^G / \partial Y = 0$  and  $\partial \hat{P}^G / \partial Y > 0$  if  $\partial P^G / \partial Y > 0$ . It is also possible that  $\partial \hat{P}^G / \partial Y = 0$  even if  $\partial P^G / \partial Y > 0$ . For example, a potential student from a low income family may assess the probability of college graduation simply by comparing his college entrance exam scores or high school grades to the distribution of these scores reported by a potential college, and may not realize that his college preparation may have been less thorough than the education received by individuals from higher income families.

model suggests that the relationship between income and college performance that is observed for individuals who select into college (conditional on other observable characteristics) will tend to serve as an accurate estimate of the relationship that is present in the population of interest when the first expression in equation (4) holds as a strict equality and will tend to serve as a downwardly biased estimate of the relationship that is present in the population of interest when the first expression in equation (4) holds as a strict inequality. Further, although the assumption that  $C^*$  is constant was useful for illustrative purposes, a sufficient condition for the result is that the  $C^*$  for low income students is at least as high as the  $C^*$  for high income people. This is the case because lowering  $C^*$  induces individuals with lower probabilities of graduation to enter college. In this scenario, on the margin where individuals are indifferent between attending and not attending college, low income individuals with similar observed characteristics would tend to have higher (or the same) levels of unobserved motivation.

When tuition costs are non-zero it may not seem plausible to assume that  $C^*=C/(W^G-W^N)$  is at least as high for low income students because, holding ability constant, a student from a low income family is likely to face lower tuition costs at a particular institution due to the presence of need-based financial aid. On the other hand, at Berea where tuition costs are zero,  $C$  is essentially constant across individuals. In this case, if the simple theoretical model reasonably depicts the college entrance decision, the observed relationship between family income and college performance at Berea will be an accurate or conservative estimator of the family environment effect if the return to a college education,  $W^G-W^N$ , that is anticipated by students, is increasing or constant across the income groups represented in the Berea data.<sup>16</sup> Previous literature suggests that it is reasonable to believe that this condition is satisfied. According to Patrinos

---

<sup>16</sup>Further, even if  $W^G-W^N$  is decreasing across income groups, the model suggests that problematic selection bias will only exist if the amount of upward bias that this creates is greater than the amount of downward bias that was shown to exist if individuals from low income families make the college entrance decision under the realization that they may be less likely to graduate.

(1995), previous research has found that no relationship exists between returns to schooling and family socioeconomic background in the United States.<sup>17</sup> Further, studies at Berea College have shown that the majority of the subset of students at Berea who come from the Appalachian region return to the Appalachian region after graduation. This will tend to lower the returns to schooling of the poorer students at Berea if, as suggested by Renkow (1996), the returns to schooling are lower in rural areas where high skilled jobs are difficult to find. Finally, our results in section V indicate that the cumulative grade point averages of students at Berea increase with family income. Evidence in the literature that a positive relationship typically exists between college grades and future earnings (see e.g., Loury and Garman (1995)) suggests that this will tend to increase the returns to education of high income individuals relative to low income individuals.

The model describes education from the traditional viewpoint of an investment good. However, if education is also a consumption good (and is normal), children from higher income families will tend to demand more education, and, therefore, will be more likely than children from lower income families to enter college if everything else is equal.<sup>18</sup> For reasons directly analogous to those described above, this would imply that students in the sample from high income families may have lower values of  $M$  than students in the sample from low income families with identical observable characteristics. Thus, any selection bias from this source would tend to make resulting estimates more conservative.

The preceding portion of this section has focused on whether selection bias is created as individuals

---

<sup>17</sup>For example, Cohn and Kiker (1986) reach this conclusion using the Michigan Panel Study of Income Dynamics (PSID). Using data from the 1980 census, Card and Krueger (1992) find “no evidence that parental income or education affects average state-level rates of return.” However, because the census does not include a direct measure of parental education and income, Card and Krueger (1992) use the median level of income and education in a person’s state. See also an earlier study by Hauser (1973).

<sup>18</sup>For example, Haveman and Wolfe (1984) suggest that “education appears to be consumed for its intrinsic value, and possibly to broaden forms of entertainment enjoyed.” Among others, see also Lazear (1977).

decide whether to attend college. However, as noted earlier, the composition of our sample at Berea is also determined by the decision of whether to apply to Berea, the admission decision at Berea, and the final student decision of whether to attend Berea if accepted. It does not appear that problematic sources of selection bias arise due to the college's admission decisions. To be eligible for admission to Berea, the family income of a potential student must be below an income threshold. However, after determining the "income-eligible" pool of applicants, admission decisions are made on the basis of the quality of a person's application without reference to the person's family income. As a result, there is no reason to believe that low income students receive favorable treatment that would allow them to be admitted with lower levels of  $M$ .<sup>19</sup> However, as will be discussed in Section IV, it is necessary to consider how the admission income threshold could influence the interpretation of the results given the existence of only a noisy measure of permanent income.

With respect to choosing between Berea and other schools, it seems likely that holding  $X$  and  $Y$  constant, individuals with higher levels of  $M$  are less likely to view Berea as their top choice. One reason for this is that individuals with higher  $M$  are more likely to receive scholarships at other schools that lower the cost difference between Berea and their other alternatives. Another reason for this is that individuals with higher  $M$  may benefit more from attending institutions that offer types of academic programs that are not offered at Berea. For the sake of illustration, holding  $X$  constant, suppose that there exists a motivation threshold  $M^*(Y)$  such that a person with family income  $Y$  will attend Berea if  $M < M^*(Y)$  and will attend a different institution if  $M \geq M^*(Y)$ . For reasons directly analogous to those discussed earlier, selection bias

---

<sup>19</sup>Admissions counselors are likely to observe indicators of "motivation" (e.g., extracurricular activities) that are not observed by the econometrician. Thus, even if motivation is uncorrelated with family income in the population of interest, a sample in which low income students have lower levels of motivation could be created by the admissions process if low income students were admitted with less desirable values of these indicators (holding other observable characteristics constant).



would imply that our sample produces a conservative (accurate) estimate of the family environment effect if  $M^*(Y)$  is decreasing (monotone) in  $Y$ . There are several reasons to believe that this may be the case. First, borrowing money to attend other institutions may be difficult or more expensive for students from low income families if liquidity constraints exist. Second, students from higher income families are more likely to receive subsidies from their parents to cover the tuition costs at other institutions.<sup>20</sup> Third, learning about Berea may be more costly for students from high income families if the income cutoff at Berea implies that these individuals have less contact with other individuals who have attended Berea. Finally, individuals from higher income families may potentially get less non-pecuniary benefits from attending Berea if the income cutoff implies that some of a person's high school friends cannot attend Berea or if a higher income student feels that it is undesirable to attend a school where many students come from much poorer backgrounds. Thus, it seems plausible to believe that the choice of whether to attend Berea or a different school does not lead to the problematic situation in which students in our sample from low income families tend to have lower levels of  $M$  than individuals in our sample from high income families.

### **III. Data**

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

---

<sup>20</sup>This implies that, from the student's perspective, the effective cost difference between Berea and alternatives may be quite small. This assumes that, after high school graduation, parents do not simply give their children an unrestricted cash gift for an amount that is constant regardless of the post-secondary institution that the student attends. Using a dynamic programming model, Keane and Wolpin (1998) suggest that parental subsidies play an important role in determining educational attainment.

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.<sup>21</sup> In section IV we discuss the interpretation of our results given the use of math ACT and verbal ACT as ability measures. Only 178 individuals had a missing value of one or more of these variables.<sup>22</sup> 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 for 418 of the 2821 students. Thus, we choose to primarily present results from models that do not include high school grades.<sup>23</sup>

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 eligibility for admission requires that a student's family income must be below a maximum level. Therefore, even the highest observed family incomes 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

---

<sup>21</sup>Most 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."

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.

<sup>22</sup>The 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.

<sup>23</sup>However, 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.

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.

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 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.<sup>24</sup> Stinebrickner (1998a) shows that almost all individuals in this sample who return for the start of their fourth year (and are not censored) eventually graduate.<sup>25</sup> Therefore, beginning the seventh semester is almost synonymous with graduation.

---

<sup>24</sup>For the empirical work, individuals who persist until the seventh semester are artificially censored at this point.

<sup>25</sup>The 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.

Figure 2 shows a non-parametric Kaplan-Meier survivor function for the duration of time that an individual 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). Thus, the probability that an individual will stay more than six full semesters (start his seventh semester) is approximately .47. 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 function does 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 difference between income groups may remain even when other observable characteristics are controlled. This is explored formally in the next section using a proportional hazard model.

#### **IV. A proportional hazard model of attendance duration**

A proportional hazard model is used to examine the relationship between the available observable characteristics described in the previous section and the duration of attendance spells. The hazard,  $h_i(t)$ , represents the probability that a person will leave school at time  $t$  conditional on not having left before time  $t$

$$(6) \ h_i(t) = \exp(\beta X_i + \epsilon_i + B(t))$$

where  $\beta$  is a set of coefficients which measure the effect of the exogenous characteristics  $X_i$  on the hazard

rate,  $\epsilon_i$  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 model is specified with a non-parametric baseline and a parametric (normal) distribution for the unobserved heterogeneity.<sup>26</sup> The estimation of the model using maximum likelihood is discussed in Appendix A.

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 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  $\exp(-.083) = .920$ . For

---

<sup>26</sup>The baseline hazard is assumed to be constant within each of the semesters. The value of each of these constants is estimated.

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.<sup>27</sup> 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).<sup>28</sup> 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 quite similar for the bottom three income groups but increase significantly over the upper half of the income distribution at Berea.

It is worthwhile to note 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. If the measurement error associated with the former is uncorrelated with the latter, any existing bias will lead to an attenuation of the effect of income on duration in the model where income enters as a continuous variable. However, one must also consider 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

---

<sup>27</sup>The baseline person was given the mean values of the continuous covariates and was given median values for the indicator variables.

<sup>28</sup>Note that little difference is observed in the income coefficient in column four when high school grades are included.

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 is true, any existing bias would lead to an undesirable overstatement of the effect of income on duration.

Nonetheless, it is important to stress that, even if this type of non-classical measurement error is present, the conclusions from the models in which income enters as a series of indicator variables would remain unchanged. In these models, the primary effect of the type of measurement error described above would be to change how we interpret the “high income” group. The measurement error would not be expected to cause an overstatement of the coefficients associated with the income group indicator variables, and, as a result, the interpretation of the results from these models would continue to be that individuals in the lower income groups tend to leave school much more quickly than individuals in the higher income groups.

Further, whether (and to what extent) the type of non-classical measurement error described above is present depends on several factors such as how many individuals are “close” to the income threshold and how much inter-temporal variation exists in family income. Fortunately, it is possible to provide a rough examination of this issue because family income is observed for all years that a student attends Berea (and, as discussed in Section V, is also sometimes observed in the year after exit) and students are not forced to leave school if their family income crosses the income threshold after the first year. For 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. Thus, although this does not provide perfect information about permanent income, no empirical evidence exists to suggest that the income effect in the continuous income models would be biased upwards by the type of non-classical measurement error described above.<sup>29</sup>

The previous results show that, even in the presence of a full tuition subsidy, low income individuals are more likely to drop-out of this college before completion, even after controlling for other factors which also influence attendance duration. In reality, 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. At least for these students, 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 approximately .17 of exiting students express some intent to transfer to another post-secondary institution. However, the majority of these students never actually request a transfer transcript, which in most cases, is a necessary condition for actually transferring. Further, if students from high income families are more

---

<sup>29</sup>Similar income effects were found when the model was estimated including family income as a time-varying covariate.



likely to transfer than students from low income families, the income findings in this study represent a lower bound on the true differences between income groups.<sup>30</sup>

## **V. Interpretation of Results -Reasons for attendance duration differences between income groups.**

The previous section indicates that statistically significant and quantitatively large differences exist in retention rates between income groups even in the presence of the full tuition subsidy. From the standpoint of designing effective policy programs, it would be desirable to determine which of the possible reasons for these differences discussed in the introduction are most plausible. In this section, we attempt to examine this issue.

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 which 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 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.

However, in reality it is certainly true that ACT scores to some extent also capture the amount of

---

<sup>30</sup>The results here would also understate differences in quality adjusted educational attainment between income groups if high and low income individuals are equally likely to transfer but high income individuals tend to transfer to better schools (or four year schools instead of two year schools) than low income students.

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 section 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.<sup>31</sup> 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.<sup>32</sup> 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 time-varying covariate.<sup>33</sup> The coefficient on cumulative grade point average in Table 4 indicates that poor grades are a

---

<sup>31</sup>The mean and standard deviation of students’ grade point averages in the first period are 2.464 and .856 respectively.

<sup>32</sup>p-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.

<sup>33</sup>The 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

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.<sup>34</sup>

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.<sup>35</sup> 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. Table 5 shows a cumulative grade point average regression for all individuals who

---

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 (1998b).

<sup>34</sup>Individuals 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.

<sup>35</sup>Another 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 Stinebrickner (1999b), 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.

remained in school until their seventh semester, and, therefore, are very likely to graduate. Income is a significant predictor of cumulative grade point averages at the beginning of the seventh semester for these students (with coefficient of .028 and a t-statistic of 2.9) even though this group should be relatively unaffected by the type of grade endogeneity that was discussed 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.<sup>36</sup> Thus, the grade data does seem to provide 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 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.”<sup>37</sup> 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

---

<sup>36</sup>The 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.

<sup>37</sup>See *The Briefing Packet of the Kentucky School and District Accountability Results for Accountability Cycle 3*, Kentucky Department of Education, Wilmer S. Cody, Commissioner, 1998.

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. The pooled regression involving all semester grades in all years for the KY subsample is shown in Table 6. The coefficient and t-statistic associated with the school quality variable are .017 and 3.034 respectively which shows that higher school ratings are related to significantly higher college grades. Nonetheless, the estimated effect of family income on college grades remains large. Although the results are not shown, when the school quality measure is not included, the point estimate and t-statistic for the KY subsample are .043 and .3666 respectively. Table 6 shows that the point estimate and t-statistic for the KY subsample are .038 and 3.160 respectively when the school quality measure is included.

Although this would seem to suggest that the effect of family income on attrition will remain 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 estimate the duration model from Table 2 for the Kentucky subsample and include the school quality measure. Because the estimated effects of the majority of the variables are very similar to those in 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. 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. Certainly it is possible that a non-trivial amount of measurement error exists in the school ratings.<sup>38</sup> Nonetheless, the results seem to suggest that families have a very strong direct 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 in the presence of a full tuition subsidy. 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, it is possible to some extent to examine whether negative income shocks influence the attendance decision when a full tuition subsidy is in place, and to what extent the effect of income shocks differ by income. The results of the proportional hazard model with the additional "change in family income" variable,  $\Delta$  income, entered as a time varying covariate are shown in Table 7 and indicate that negative income shocks have an insignificant effect on the timing of exits.

Table 7 suggests that differences between income groups are not being caused by differential responses to income shocks. However, 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

---

<sup>38</sup>For example, it is likely that the ratings were not designed to exclusively measure college preparation.

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 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, this timing assumption will lead to a downward bias for the estimates of the effect of income shocks. The extent of this bias depends on the extent to which income shocks tend to cause students to exit very quickly.<sup>39</sup>

## **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 tuition explanation (assuming that selection bias has been accounted for or is not important). The previous part of the paper suggests that the family environment explanation is important. However, because tuition is zero for everyone at Berea, the Berea data cannot provide direct evidence about the importance of the tuition 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

---

<sup>39</sup>Under 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.



relative to the tuition 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 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 mimic the Berea sample to as large a degree as possible, we remove the 684 individuals in the data who have a family income that is above \$85,8000 in 1997 dollars.<sup>40</sup> 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.<sup>41</sup> The descriptive statistics for these individuals is shown in Table 8.

The analogs to column 1 and column 2 of Table 2 are shown in Table 9.<sup>42</sup> In both cases, the

---

<sup>40</sup>Although this is higher than the income threshold in the Berea data, because the income data is categorical, the next option would be to draw the cutoff at \$57,200. Although it would probably be desirable to focus on students from schools of similar size as Berea and schools in similar geographic regions, this was not possible given the size of the sample.

<sup>41</sup>The 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.

<sup>42</sup>The family income variable in the NELS-88 is categorical with classes defined as in Table 8. 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 income classes below and including the (11440,17160] class, we define the high income variable to include the income classes (40040,57200] and (57200,85800], and we define the middle income variable to include all other income classes.

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 from Table 8 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

estimated effects of family income are very similar to those in Table 2, despite the fact that the students in the NELS-88 are attending tuition charging institutions.<sup>43</sup> 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 tuition costs under current tuition/financial aid programs is influenced by both the possibility that low income students find paying for tuition burdensome for reasons such as liquidity constraints and the reality that students from low income families are likely to face lower tuition 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 tuition on college outcomes is close to zero, this does not directly imply that liquidity constraints are unimportant for low income students.

## **V. 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 less than \$1,000 for room, board, and college fees), suggest that college exits often occur for reasons that are unrelated to the costs of college. As discussed in detail in the paper, it important to note that in general it will always be difficult

---

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.

<sup>43</sup>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 .71 and .64 respectively.

to rule out with certainty the possibility that the observed relationship between income and college attendance duration is influenced by selection bias. Nonetheless, given the implications from the theoretical model described in the paper, the statistically significant and quantitatively large relationship that is found between income and performance at Berea, 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 tuition was 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 tuition 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.

## Appendix

If individual  $i$ 's attendance spell is censored after the start of the  $t+1^{\text{st}}$  semester, the appropriate likelihood contribution is the survivor function evaluated at  $t$  (the person completed more than  $t$  full semesters). In a model without unobserved heterogeneity, this can be written as

$$L_i = \exp\left\{-\sum_{s=1}^{t+1} h_i(s)\right\}$$

If the individual leaves school sometime during semester  $t+1$  (the spell is uncensored), the exit probability is the difference between the survivor function evaluated at time  $t$  and the survivor function evaluated at time  $t+1$  (the student completed more than  $t$  full semesters semester but did not complete more than  $t+1$  full semesters). In a model without unobserved heterogeneity, this can be written as

$$L_i = \exp\left\{-\sum_{s=1}^t h_i(s)\right\} [1 - \exp\{-h_i(t+1)\}].$$

The log likelihood function,  $L$ , for the entire sample is the sum of the log likelihood contributions of each person in the sample,  $L = \sum_i \log(L_i)$ . When unobserved heterogeneity is included, the likelihood contribution involves integrating the above likelihood contributions over the distributions of the heterogeneity component.

## 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
- 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, 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," 1998a
- Stinebrickner, Todd R. "Estimation of a Duration Model in the Presence of Missing Data," 1998b, forthcoming *The Review of Economics and Statistics*.
- 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 sample and 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	.454	.446	.439	.476
black	.100	.131	.087	.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	.396	.388	.385	.415
number in family	4.050 1.516	3.728 1.515	3.827 1.311	4.595 1.560



Table 2 Proportional hazard model of retention

	n=2821	n=2821	n=2821	n=2442
variable	estimate SE	estimate SE	estimate SE	estimate SE
male	.242* (.067)	.240* (.067)	.243 (.067)	0.068 (.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 <sup>nd</sup> 1/6			.329* (.112)	
income in 3 <sup>rd</sup> 1/6			.291* (.112)	
income in 4 <sup>th</sup> 1/6			.222* (.112)	
income in 5 <sup>th</sup> 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	-3484.610	-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  
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 retention including cumulative college grade point average  
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 Regression of cumulative GPA  
for all students who completed six semesters n=963

	GPA start of seventh semester
constant	1.599* (.092)
male	-.089* (.028)
black	-.180* (.047)
verbal act	.023* (.003)
math act	.034* (.003)
income/10000	.028* (.010)
	R <sup>2</sup> =.240

Table 6  
Pooled Regression of College Grades  
for Kentucky subsample, 1117 individuals, 4569 total observations

	GPA first semester
constant	.465 (.247)
male	-.177 (.036)
black	-.249* (.065)
verbal act	.034* (.005)
math act	.034* (.005)
income/10000	.038* (.012)
school quality	.017* (.006)
	R <sup>2</sup> =.156

regression estimated using Stata including cluster and robust options.

Table 7 Proportional hazard model of retention including information on family income shocks  
n=2446

male	.220* (.071)	.234* (.071)	
black	-.160 (.118)	-.146 (.118)	
verbal ACT	-.017 (.009)	-.016 (.009)	
math ACT	-.049* (.010)	-.051* (.010)	
distance from home - close	-.211* (.074)	-.201* (.074)	
income/10000	-.098* (.028)	-.089* (.114)	
$\Delta$ income/10000	-.010 (.023)	.068 (.074)	
$\Delta$ income/10000 xincome/10000		-.037 (.029)	
t=1	-1.905* (.127)	-1.966* (.127)	
t=2	-2.134* (.136)	-2.207* (.136)	
t=3	-2.186* (.141)	-2.259* (.141)	
t=4	-2.680* (.158)	-2.757* (.159)	
t=5	-2.703* (.167)	-2.781* (.168)	
t=6	-3.118* (.196)	-3.199* (.196)	
loglike value	-2694.923	-2693.970	

\* represents a t statistic greater than two.

$\Delta$ income measures deviation of income in a particular year from the family income when the student enrolled.

Table 8 - Data description- NELS-88 sample  
n=1468

	Full Sample
	mean   std dev.
male	.461
black	.102
verbal standardized test score	55.491   8.300
math standardized test score	56.728   8.184
family income categories	
[0,0]	.003
(0,1144]	.002
(1144,3432]	.001
(3432,5720]	.008
(5720,8580]	.018
(8580,11440]	.026
(11440,17160]	.040
(17160,22880]	.062
(22880,28600]	.086
(28600,40040]	.150
(40040,57200]	.268
(57200,85800]	.333

Table 9 Proportional hazard model of retention-NELS-88

	n=2821		n=2821	
variable	estimate	SE	estimate	SE
male	.153 (.010)		.152 (.101)	
black	-.273 (.158)		-.246 (.160)	
verbal standardized test	-.007 (.008)		-.007 (.008)	
math standardized test	-.047* (.008)		-.047 (.008)	
income/10000	-.069* (.025)			
indicator for income in bottom 1/3			.270 (.159)	
indicator for income in middle 1/3			.130 (.111)	
t=1	-.520* (.242)		-.873* (.245)	
t=2	.342 (.245)		-.011 (.246)	
t=3	-.379* (.258)		-.735* (.261)	
log likelihood	-1221.835		-1223.662	

\* t statistic greater than 2.0

figure 1

**Relative frequency of family income  
in 1997 dollars**

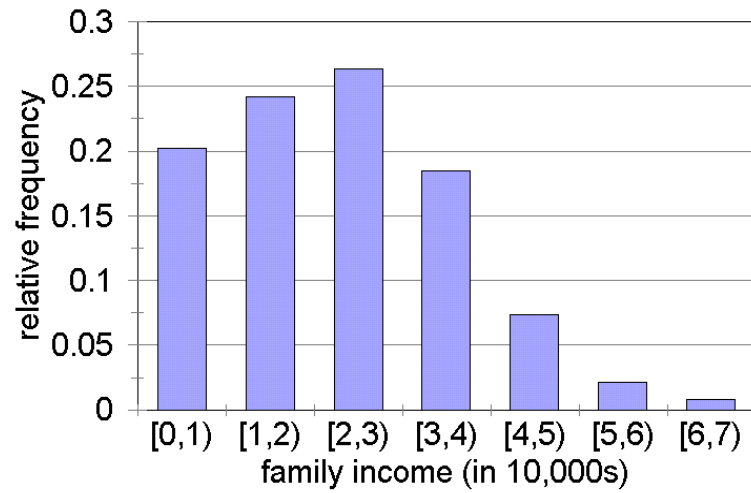


figure 2

**Kaplan-Meier survivor function  
full sample n=2821**

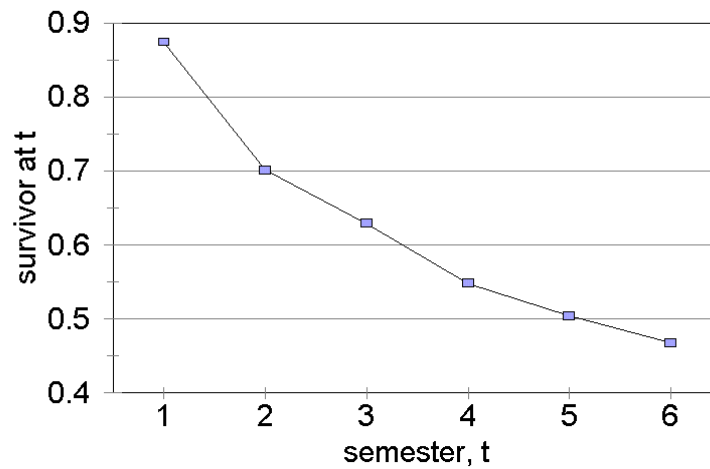




figure 3

**Kaplan-Meier survivor functions**  
for three income groups

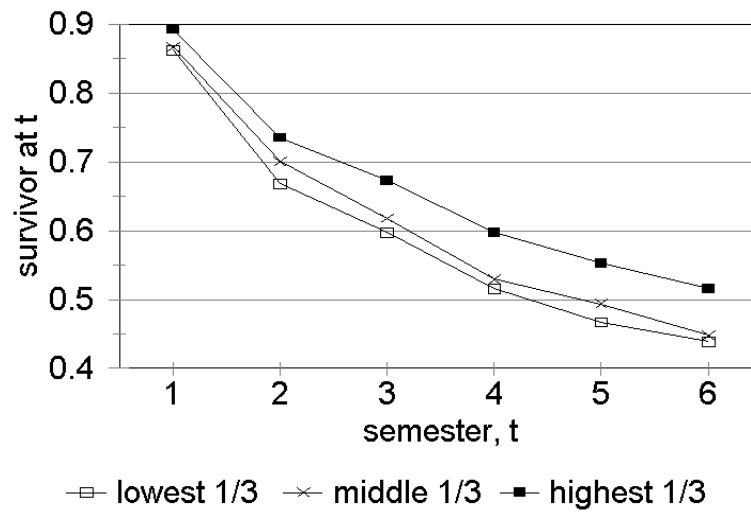


figure 4

**Predicted Survivor Functions**  
for baseline person by income

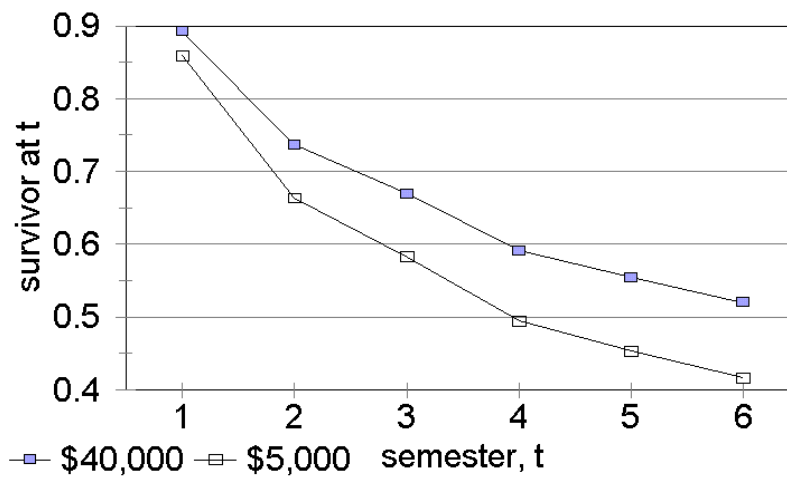


figure 5

**Predicted survivor functions**  
for three income groups

