Managing Student Loan Default Risk: Evidence From a Privately-Guaranteed Portfolio

IHELG Monograph 97-9

Revised Summer 1998

Kirk Monteverde, Ph.D.
Research Director
The Access Group
P. O. Box 7460
Wilmington, DE 19803
302-477-4112
kmonteverde@accessgroup.org

© Kirk Monteverde, Ph.D., 1997
$5.00
University of Houston Law Center/Institute for Higher Education Law and Governance (IHELG)

The University of Houston Institute for Higher Education Law and Governance (IHELG) provides a unique service to colleges and universities worldwide. It has as its primary aim providing information and publications to colleges and universities related to the field of higher education law, and also has a broader mission to be a focal point for discussion and thoughtful analysis of higher education legal issues. IHELG provides information, research, and analysis for those involved in managing the higher education enterprise internationally through publications, conferences, and the maintenance of a database of individuals and institutions. IHELG is especially concerned with creating dialogue and cooperation among academic institutions in the United States, and also has interests in higher education in industrialized nations and those in the developing countries of the Third World.

The UHLC/IHELG works in a series of concentric circles. At the core of the enterprise is the analytic study of postsecondary institutions—with special emphasis on the legal issues that affect colleges and universities. The next ring of the circle is made up of affiliated scholars whose research is in law and higher education as a field of study. Many scholars from all over the world have either spent time in residence, or have participated in Institute activities. Finally, many others from governmental agencies and legislative staff concerned with higher education participate in the activities of the Center. All IHELG monographs are available to a wide audience, at low cost.

Programs and Resources

IHELG has as its purpose the stimulation of an international consciousness among higher education institutions concerning issues of higher education law and the provision of documentation and analysis relating to higher education development. The following activities form the core of the Institute’s activities:

Higher Education Law Library

Houston Roundtable on Higher Education Law

Houston Roundtable on Higher Education Finance

Publication series

Study opportunities

Conferences

Bibliographical and document service

Networking and commentary

Research projects funded internally or externally
MANAGING STUDENT LOAN DEFAULT RISK: EVIDENCE FROM A PRIVATELY-GUARANTEED PORTFOLIO

May 1998
(rev D)

Kirk Monteverde
Research Director
Access Group, Inc.
Managing Student Loan Default Risk:
Evidence from a Privately-Guaranteed Portfolio

Student loans, borrowed directly from or guaranteed indirectly by, the federal government have become the lifeblood of American higher education. The U.S. Department of Education has estimated that among 1995/1996 graduates over 50% of students earning Bachelor's or Master's degrees and over 70% of Professional degree earners had had their education at least partially financed through federal student loans.(1) And during the 1990's the sheer volume of federally-sponsored loans has mushroomed. In fiscal 1997 (the 12 months ending September 30, 1997) an estimated $31 Billion in federally-sponsored student loans were originated, more than two-and-a-half times the amount extended in fiscal 1990.(2) A college freshman today can borrow up to $2,625 and a senior $5,500, even if the student is classified as "dependent" (independent students can borrow more). Most significantly, in 1992 an "unsubsidized" category of federally-guaranteed loans was created that made such loans available to all students regardless of need. Furthermore, parents of dependent students were henceforth allowed to borrow an unlimited sum in government-guaranteed money to finance their offspring's college education.(3) Few colleges could maintain anywhere near their current levels of expenditures without the liquidity provided their customers through the various loan programs administered primarily through the U.S. Department of Education.

The reliance upon federally-sponsored loans is even more pronounced for post-baccalaureate programs. A professional or graduate school student, under his or her signature alone, can borrow up to $18,500 per year in government-insured loans, up to a cumulative total of $138,500.(4)

Yet despite this seemingly generous $18,500 per year allowance, for certain high cost graduate professional programs, such as law, business, medicine and dentistry, federally-guaranteed loans are insufficient to fund a student's tuition and living costs at many private and even some public schools. To overcome these federal limits, a number of privately-guaranteed loan programs have evolved.(5) One such program is managed by the non-profit, membership organization the Access Group, Inc. This organization, whose Board of Directors is elected by the deans of the country's 179 American Bar Association-accredited, non-profit member law schools, was spun off in 1993 from the Law School Admissions Council (LSAC), the non-profit, law school membership group responsible for administering the standardized LSAT test generally required for law school admission.
The original mandate of the Access Group was provision of educational, research, and financial assistance to students attending or seeking admission to graduate law programs. In fulfillment of its charter to help students secure education financing, what is now the Access Group has originated nearly $5 Billion in student loans since its modest start in Academic Year 1983/1984, then, of course, as part of LSAC. Although the majority of the loans have been of the more traditional "Stafford"(6) variety (guaranteed ultimately, albeit indirectly, by the federal government) over $1 Billion of these originations have been sourced wholly within the private sector---funded, insured, serviced, and securitized without federal government assurances. As is true of most privately-guaranteed student loan products, the principal private offerings of the Access Group have loan features which generally emulate the federal model. For example, the Access Group's Law Access Loan (LAL) is a long-term (15-years for the loans analyzed here), unsecured, variable rate consumer installment loan which allows for capitalization of interest during an "in-school" period.

It is this LAL private loan product which is the focus of the current paper. Specifically, this research documents the early default experience observed among a sample of over 13,000 law school private loan recipients and reports on a set of variables found predictive of default. The study, it is believed, provides the first publicly-available statistics on default for a privately-guaranteed student loan portfolio. Furthermore it is the first to focus narrowly on default determinants among a set of post-baccalaureate students. Finally, it introduces two statistical techniques, survival analysis and credit bureau scoring, which have not been used together to analyze student loan defaults in any prior published literature.

I. Federally-sponsored Student Lending and Public Policy

Because of their long history, size and importance to higher education, the federally-sponsored student loan programs have been the subject of virtually all previous academic literature on default. Furthermore, because the credit losses inherent in these loan programs are ultimately born by the American taxpayer, most of the studies on student loan default have addressed public policy issues, i.e., how should government manage credit risk while maintaining the policy purpose behind the programs? Many of the studies have taken as a point of departure the most visible mechanism the federal government has employed to manage that credit risk, the refusal to fund students attending institutions where a very high proportion of former students have failed to repay their loans. This policy has been interpreted in casual discussion as "blaming" the institutions for the actions of their students. To borrow from the title of an early student loan study, "Whose fault is default?" (Wilms, Moore, and Bolus, 1987) has been the manner in which the policy question has been posed.

To understand the context in which this question is addressed requires some background on the "idiosyncracies" (as they would seem to a commercial bank credit officer) of the federally-sponsored loan programs. First, the programs, under the mandate for broad access to higher education, require that no student attending an approved institution be denied credit, with the sole exception that an applicant cannot have previously defaulted on a student loan (other indicators
of poor credit are generally to be ignored). Secondly, the loans have a long 10-year repayment period, yet no principal or interest payments are required while a student continues in school. Interest does accrue during this in-school period and is capitalized onto the principal when the loan finally enters repayment. Third, the definition of an approved institution of higher (post-secondary) education, one whose students are therefore eligible for federally sponsored loans, is very widely drawn. Ivy League graduate students are lumped together with those attending the Professional Dealers School of America, a school for blackjack dealing, for purposes of determining federal loan eligibility (or at least they were until the latter finally went out of business having reported a default rate of 73% in 1990; Koselka and Oliver Forbes, 1995).

So called "proprietary" schools, generally for-profit training centers (e.g., hairdressing, truck-driving, commercial art) most of whom are fortunately not as dubious as the Professional Dealers School, nevertheless fall outside of what is normally understood to comprise "higher education" (which is perhaps why the term "post-secondary education" is a more appropriate term to describe what federal loan programs fund). These proprietary schools are eligible to have their students (who need not have finished high school) receive federally-sponsored loans under the same eligibility rules as are applied to Harvard students. With such anecdotal evidence as the Professional Dealers case featured in the popular press, it is easy to form the impression that what was at one point an embarrassingly high overall federal loan default rate (as high as 22.4% in FY90), had been "the fault" of certain schools, in particular a handful of perhaps unscrupulous firms, concentrated among the proprietary class of post-secondary institutions. A 1996 Department of Education press release appeared to support this view, perhaps inadvertently, in proudly featuring the elimination of schools from federal student loan eligibility as a major reason behind an apparently dramatic improvement in reported default rates (to 10.7% for FY 94 and 10.4% for FY95; Department of Education Press Release, November 12, 1997).

(Department of Education Secretary) Riley also attributed the declining default rate to the department's statutory and oversight responsibility to remove schools with high default rates from participating in federal student loan programs. Since 1993, some 600 schools have become ineligible as a result (Department of Education Press Release, January 22, 1996).

In short, and to caricature the popular view a bit, default is said to be the fault of a small group of offending schools whose elimination from student loan programs should eliminate any default concerns. But as virtually all multivariate statistical studies concur, the data simply do not support this view.
II. Previous Statistical Evidence on Student Loan Default

The first statistical study of student loan default patterns with wide academic readership was the article whose title, "Whose Fault is Default," has come to summarize the research agenda for much of the literature in the field (Wilms, Moore and Bolus, 1987). The study was initiated by the California State legislature which was specifically interested in investigating characteristics and practices of those post-secondary institutions in the state whose default rates were "excessively high" (defined as 15% or greater). This focus effectively narrowed the study to a sample of students who had attended one of 93 California community colleges or 140 accredited proprietary schools. A single academic year's students, 1982/83, were chosen for analysis; ultimately 3155 individuals with complete data were examined using hierarchical stepwise discriminant analysis. The dependent variable was the dichotomous measure of "default or not" with the independent, predictor variables being of two general classes: 1) characteristics of the borrower, and 2) characteristics and practices of the school of attendance. The "hierarchical stepwise" approach meant that the independent variables fixed first in time (the borrower characteristics) were entered first into the model, followed by the institutional factors. Using this approach it was found that, after taking into account the characteristics a student brought with him or her to post-secondary study, very little predictiveness was added to the model by also taking into account the characteristics and practices of the school the student attended. With but one exception, adding any of the institutional predictors failed to significantly increase the overall predictiveness of the model.(7) Summarizing their findings the authors emphasize the "predetermined"/non-institutional predictors of default:

In short, by knowing a borrower's ethnicity, family income, citizenship, and program---as well as whether the borrower had graduated from high school and completed his or her program---one can accurately predict default in almost two out of three cases (Wilms, Moore, and Bolus, pp.48-49).

Nevertheless, the paper did not report the distribution of defaulters by individual school. School characteristics and practices, not individual school identification were used to predict default. It would have been interesting to know if particular schools or particular "vocations taught" at a group of schools were related to extraordinary default experience (the Professional Dealers anecdote again comes to mind). This is the question of whether the model was correctly specified, that is, whether institutional factors might, after all, be important predictors of default if one measured them differently.

Obvious caution is advisable in extrapolating the Wilms et al., results to try and explain loan repayment behavior among different student types, as, for example, among graduate law students, the population of concern to this current paper. Indeed, it is questionable whether the default determinants found among proprietary and community college students should be expected to bear any relationship at all to default predictors for professional school students. In the context of public policy, as Harrison (1995) and Volkwein and Szelen (1995) have stressed, it would seem misguided to design government financial aid programs under the assumption that all borrowers
sharing the broad label "post-secondary student" can be classed together. Nevertheless, studies which have followed Wilms et al. but employed broader samples of student borrowers have generally confirmed their one fundamental finding—that borrower rather than institutional characteristics best predict defaults.

Knapp and Seaks (1992), drawing methodological guidance from earlier work by Greene (1989), studied nearly 2000 traditional 2-and 4-year college students who had attended one of 26 Pennsylvania institutions, had borrowed federally-guaranteed Stafford loans, and had graduated (or dropped below half-time enrollment) during the academic year 1984-1985. Using the maximum-likelihood technique probit (Greene, 1989, had used a related technique, Tobit), the authors found each of the following borrower characteristics to have a statistically significant association with a lower probability of default: 1) parent's income, 2) presence of two parents at home, 3) the student's having graduated (vs. dropping out) and 4) the student's race (whites defaulting less than blacks). The most substantively important variables were found to be graduation and race. For example, a white student with a degree, coming from a two-parent family with a $10,000 family income was reported to have a 2.4% default probability, while the same student, were he/she not to have graduated, was found to be 15.1% default risk. A graduating black student from a two-parent, $10,000 income family, had a 14.2% default risk. That same black student, were he/she not to graduate, carried a 44.1% default risk. Importantly, the authors tested for any school-specific default risk by explicitly entering 25 dummy variables into their models (to account for the 26 different institutions from which data were drawn). They found that the 25 dummies, taken as a group, did not significantly improve the predictiveness of their models; nor did any single dummy indicate that any of the schools had a default probability significantly different than all others, after taking into account the individual characteristics of student borrowers. They conclude, emphatically, that "at least for this sample, there is nothing related to individual collegiate institutions that has any impact upon default rates" (Knapp and Seaks, p. 407; italicized in original). But, as with the earlier Wilms et al. study, the particular sample of students investigated by Knapp and Seaks, drawn exclusively from among 2-and 4-year colleges in Pennsylvania, potentially limits the generalizability of their findings.

Volkwein and Szelest (1995) were first to combine into a single statistical model the loan repayment experiences of students attending a broad spectrum of post-secondary institutions, from proprietary schools to graduate and professional schools. Volkwein and Szelest examined a sample of over 4000 borrowers, drawn from throughout the United States, who had graduated or left school by 1994 and upon whom relatively complete transcript and survey data were available from the 1987 National Post-secondary Student Aid Study (NPSAS). The NPSAS data document borrowers' personal, demographic and family characteristics, include financial and occupational information, as well as academic records from college transcripts. In the first of two models, these borrower characteristics were examined for statistical association with default along with a set of variables used to capture any association of "institution-type" with default. The test of the latter was a simple one with "institution-type" captured by categorizing a student's school of attendance as either a proprietary school, 2-year college, 4-year college, or doctoral university. In one of the paper's principal findings, the authors demonstrate that, when entered into a multivariate logistic regression (a technique akin to both probit and Tobit), the set of dummy variables developed to capture the effect of institution type failed to significantly add to the
model's ability to predict default after individual borrower characteristics were taken into account. In contrast, several borrower characteristics, for all types of students, were found significantly predictive of default risk (race, marital status, dependent children, degree completion, college GPA, parental/family support, and current income). In a second model, which because of data limitations excluded proprietary schools, a more exhaustive set of institutional descriptors was tested for association with default risk and, although two were shown to be statistically significant (proportion of minority students, and "auxiliary" expenditures per student), the authors conclude that "(w)e find little evidence that institutional characteristics have an impact on loan default" (p. 59). However, unlike the Knapp and Seaks study, neither of the Volkwein and Szelest models incorporate institution-specific dummy variables. This leaves open the possibility that individual institutions in their study may have had significantly worse default records than the norm, even after taking into account the characteristics of the student borrowers they attract.

Summarizing the evidence reviewed above, it would be fair to conclude that the consensus among investigators who have looked into the issue of student loan defaults is that the determinants of default are primarily borrower-based rather than linked, in any causal manner, to the borrowers' school of attendance. "Offending" schools do not, in general, cause default. "High default" schools are those who attract students with a high likelihood to default. This default proclivity is a pre-existing condition.

However, all of these previous studies have drawn their data from federally-guaranteed loan programs. In contrast, the data analyzed in the present paper were drawn wholly from a privately guaranteed loan portfolio. Unlike the federally-guaranteed loan program, this private program imposes certain credit criteria upon potential borrowers.(8) Furthermore, the data analyzed here represent loans to a very narrow subset of student borrowers, law school students at American Bar Association-accredited schools. It is not clear that the relationships identified in the previous empirical literature, overwhelmingly associating default risk with borrower characteristics, will necessarily hold, or hold as strongly, among law students borrowing privately-guaranteed funds. As part of our statistical analysis, therefore, the issue of "borrower proclivity" vs "school-of-attendance" as predictors of default risk will be re-addressed.

The remainder of this paper is divided into three major sections. First, we briefly detail some of the institutional arrangements comprising the private Law Access Loan program, the source for the paper's data set. Secondly, the statistical results of the investigation are reported. Finally, the paper ends with a discussion of those results viewed in the context of the previous empirical studies just reviewed. The second (statistical results) section is itself sub-divided into three parts. First is presented a baseline model for default development with time the only "predictor" variable. Here survival analysis techniques are introduced. Second, the paper’s primary measure of borrower-based default proclivity, the borrower’s credit bureau score, is introduced and its predictive power examined when added to the baseline survival analysis model (i.e., time and credit score as predictors of default risk). Finally, a third set of potential predictors—school-based, geographic and economic variables—are added incrementally to the model and evaluated for their statistical association with default risk.
III. Private Sector Student Loans

Because of limits placed on the amount the government will guarantee for any single borrower, a substantial market has developed for student loans whose credit risk is privately borne. These loan programs are especially important in funding certain courses-of-study whose tuition cost is particularly high. Such high cost curriculums include medical school, law school, dental school, and some business and graduate studies. The need for private loans is greatest for such programs offered at private (non-state-funded) schools and for curriculums requiring full-time school attendance. Among all the disciplines cataloged above, the market for “signature-only” (non-co-signed) private loans is estimated to be largest for law school borrowers. The Access Group, through its Law Access Loan program, originates a majority of these private law student loans. The Access Group’s private loan origination volume has averaged approximately $185 Million annually during the 1990’s.

Until recently (Academic Year 1998/1999), the Access Group acted as a true loan originator with no ownership rights to the loans (nor calls on those rights) and with no direct credit risk exposure. The loans were funded through a contractual program partner, which from 1990-1998, was KeyBank (or it predecessors, Ameritrust and Society) of Cleveland, Ohio. From the Bank’s perspective the Access Group functioned as an "affinity group." Although the Bank retained some credit risk, much of that risk was shifted to a second contractual program participant, a non-profit special-purpose student loan guarantor, The Education Resources Institute (TERI) in Boston. The bank collected a variable interest rate fixed at a given spread above the 91-day Treasury Bill auction rate, an approach which emulated the federal loan program. The guarantor, which insured the timely repayment of principal and interest, collected an up-front fee (similar to points in the mortgage industry) and, if a default claim was paid, obtained ownership rights to the loan in order to pursue post-default collection efforts. A third contractual partner, a transactions processing specialist, was responsible for the actual “servicing” of all loans (e.g., producing bills, sending any needed delinquency notices, and if need be, forwarding a defaulted loan to the guarantor for collection). From 1989-1998 the primary servicing partner for the Access Group’s private loan programs was the Pennsylvania Higher Education Assistance Agency (PHEAA), a state agency whose original and still primary purpose is the guaranteeing of federally-reinsured student loans. For the Access Group’s programs there has been an additional entity in this very dis-integrated industry, the final security holder. KeyBank (or its predecessors) chose to sell or securitize, rather than hold on its own balance sheet, the student loans it funded after these loans entered repayment (again emulating the federal programs, students are not required to make any interest or principal payments on Access Group private loans until some months after they cease being enrolled full-time, although the interest accrues). The Access Group loans entering repayment in the early 1990’s were sold to SLSC, Inc. of Cincinnati, Ohio, an operating unit within Student Loan Funding, a state-endorsed non-profit organization set up specifically to purchase student loans in the secondary market. More recently, KeyBank has chosen to repackgage and sell its student loans as low-interest rate asset-backed securities (offering interest rates only a shade over benchmark LIBOR or Treasury bill rates), thus removing the loans from their own books and recording a gain on the sale of each portfolio.
IV. Data Analysis

The core dataset analyzed in the following sections was originally constructed to facilitate an actuarial assessment of the capital adequacy of program partner and loan guarantor, TERI. The raw data were sourced from TERI, but all database construction and validation was performed at the Access Group. The initial dataset contained information on over 60,000 students, all borrowers with at least one private LAL loan guaranteed by TERI and graduating (or otherwise leaving school) in the years 1991-1996. Generally, this dataset contained only information directly pertinent to each borrower’s loan obligations (e.g., number of loans, dollar amount of loans, first disbursement dates, date of student’s graduation, date and amount of any default claims paid). The only additional data contained in this initial dataset were borrowers’ date of birth and “school of attendance” (i.e., school of graduation or last school attended). No other demographic data were available. It was this dataset upon which the actuarial study used to evaluate TERI’s capital adequacy was performed (results of which are not reported here for proprietary reasons). It is also the dataset described in the section below entitled “Survival Analysis.”

A subset of the aforementioned 60,000 record dataset was then formed to investigate in more detail the possible predictors of default. This subset was comprised of all borrowers of a Law Access (private) Loan during Academic Year 1991/1992, which represented a sampling of law student borrowers generally scheduled to graduate in the years 1992, 1993, and 1994, depending upon their law school grade level in 1991/1992. Credit scores for each of these borrowers at the time of loan underwriting were sought from archival sources and, when available, appended to the borrowers loan repayment record. A dataset was thus formed containing over 13,000 1991/1992 private borrowers with loan performance information coupled to each borrower’s historic credit score. This dataset serves as the basis of the analysis described in the section below entitled “Borrower Credit Risk Class as a Default Predictor.”

Finally, school-of-attendance descriptors (e.g., school identity dummies or reputational measures) along with geographic and economic data (e.g., the unemployment rate among new law graduates in the region of each student’s graduation the year they graduated) were added as data elements to the credit scored dataset to investigate various alternative explanatory models of default behavior. These models are examined in the paper’s third and last data analysis section.

1. Survival Analysis

As its name implies, survival analysis comprises a set of statistical techniques originally designed to model mortality patterns, death, over time among an organic, often human, population. One branch of these techniques developed from the common life expectancy tables, critical to the insurance industry, which trace as far back, at least, as Edmund Halley’s (of Halley’s Comet fame) study of 17th century Breslaw (now Wroclaw in Poland). Survival analysis is particularly intuitive when an expected "event" pattern (death being one type of "event") within a homogeneous population can be expressed as a simple mathematical function. The classic functional form, especially in the natural sciences, is exponential. For example, let \( S(t) \) represent the probability of any individual in the population under study surviving beyond time \( t \) and specify this "survival function" as \( S(t) = e^{-\lambda t} \) where \( t \) is continuous time and \( \lambda \) is some constant.
Associated with each survival function is a "hazard function" which captures the important intuitive notion of risk; that is, associated with S(t) is a hazard function h(t) which may be viewed as giving the instantaneous probability at time t of the event occurring (dying, for example).\(10\) The beauty of the exponential survival function is that its associated hazard function is a constant, \(h(t) = \lambda\). If your survival function is exponential, your risk (instantaneous probability) of dying remains constant over time. This does not mean, of course, that for a population of individuals with identical survival functions the expected number of deaths is constant each period, but only that, of those individuals alive at the beginning of each period (some having died earlier), the same proportion is expected to die.\(11\)

Loan defaults can be thought of as analogous to deaths, "borrower mortality" perhaps. We suspect that the hazard function (default risk as a function of time) for a population of borrowers, initially here all considered identical, can be modeled simply, albeit perhaps not quite as simply as \(h(t) = \lambda\). Indeed, actuarial studies assessing the capital adequacy of loan guarantors (insurers against default risk) use such modeling techniques to guide their extrapolation of a portfolio's observed early default experience to estimate an ultimate life-of-the-portfolio default rate. Figure 1 displays data from just such an actuarial study, one performed to estimate the long-run default rate for TERI's entire portfolio of Law Access Loans. Displayed are the actual and projected conditional quarterly rates of borrower default for all privately-guaranteed law student loans originated by the Access Group since Academic Year 1990/1991 through Summer, 1996 for borrowers leaving school in the period 1991-1996. Actual conditional default rates are displayed as red and blue bars and projected rates are the green bars. Over $1 Billion dollars in original principal and more than 60,000 borrowers are represented. The x-axis of the graph is the time dimension measured in quarters since borrowers' graduation (or otherwise withdrawal) from law school. Note that the x-axis does not represent fixed calendar dates. Each borrower is examined for his or her own elapsed time since their graduation and categorized together with other borrowers with identical elapsed times, independent of the actual calendar dates of their graduations.\(12\) These data form a super-set of the 13,000+ borrowers examined in greater detail in the remainder of this paper. And, although nothing beyond time trend (survival) analysis was performed on this larger dataset, two fundamental findings about privately-guaranteed law student loan default patterns may be derived from the conditional default graph in Figure 1.

First, for institutional reasons detailed in Section 2b, there appears to be no simple hazard function to describe the default pattern for loans defaulting in or prior to the 9th quarter after graduation. This period, shown in red on Figure 1 we will refer to as the period of loan "seasoning." In subsequent analysis we will capture the effect of this early period (using a dummy variable), but will not attempt to specify any functional form for the default hazard during this period of seasoning. In fact, all periods prior to and including the 9th quarter will be treated as if they were one period.\(13\)

The second important observation to be made from Figure 1 concerns the conditional default pattern for those periods following loan seasoning. Beginning in Quarter 10, something akin to a mortality curve, does appear. The actuarial study performed under the sponsorship of the Access Group used data from Quarters 10-20, the blue bars, to extrapolate future conditional defaults, the green bars. Enabling this extrapolation was the discovery of a simple functional form
Figure 1
Quarterly Conditional Default Rates

Quarters since graduation
which fit reasonably well the limited experience of the portfolio. The function was not, however, as simple as the exponential survival function (constant hazard function) discussed earlier; that would have lead to a projection of flat conditional default rates. Instead, the fitted hazard rate was of the form \( h(t) = \lambda t^{-\kappa} \), (a Weibull distribution) with \( \lambda \) some constant and, in this case, \( \kappa < 0 \); the hazard rate itself declines with the passage of time.\(^{(14)}\) The Weibull hazard function implies a linear relationship between the log of time and log of hazard,

\[
\log h(t) = \mu + \kappa \log t, \text{ where } \mu = \log \lambda.
\]

We will use this relationship when we model the influence of explanatory variables on the risk of default. Specifically, we model:

\[
\log h_i(t) = \mu + \kappa \log t_i + \beta_1 x_{i1} + \beta_2 x_{i2} + \ldots + \beta_k x_{ik},
\]

where \( x_{i1}, x_{i2}, \ldots, x_{ik} \) are values of a set of predictor variables for individual \( i \), and \( t \) is measured in discrete quarters.

Although the time trend provided by this initial survival analysis is an important prerequisite for modeling default, our real interest is in finding a set of predictor variables which, after taking into account the general default pattern over time, significantly adds incremental explanatory power to the model. We seek predictor variables which differentiate between borrowers, something missing from simple actuarial time trending. The first such variable to be considered, and the one used here to operationalize the key construct of borrower-based default proclivity, are borrowers' "credit scores" at the time of their loan origination.

2. Borrower Credit Risk Class as a Default Predictor

2a) Credit Scoring: Background

Generally, two types of credit scoring schemes have evolved in consumer lending. The first type of score, often called a generic (or "bureau") score, is typically available in the U.S. from each of the three major credit bureaus (TransUnion, Experian, and Equifax) who gather and maintain the elementary, detailed information on each individual upon which the score itself is based. These elementary data include such things such as an individual's number of credit cards, the number of 60-day delinquencies in the last 3 years, the number of derogatory notations, notices of bankruptcy, etc. The second type, often labeled proprietary score card development, typically uses credit bureau information plus other data about potential borrowers such as their length of residency, length of employment, and income. These items are normally gathered from a loan application. The score card is called "proprietary" because each credit grantor assigns its own "weights," measures of relative importance, to the borrower attributes (whether sourced from a credit bureau or a loan application) that are then used to construct a summary "credit score." These weights are determined by previous statistical modeling of that credit grantor's own specific applicant pool. Both types of scores have the same purpose, to provide a single summary measure validated to statistically predict default risk.\(^{(15)}\)
The credit scores used in this study are strictly of the first type, generic scores. Accordingly, individual characteristics such as, for example, personal or family income, assets owned, race and marital status are not a part of the study because they are not direct credit history elements and, therefore, are not inputs to any generic credit bureau scoring algorithm. The specific credit score series used was TransUnion's, brand named Empirica; this series returns values ranging from 395 to 848.

Although a set of credit hurdles had always been employed to qualify the Access Group private borrowers, the generic credit score number itself had not been a part of the underwriting decision process. Nor had past values for borrowers been routinely retained by the Access Group or its program participants. However, historical credit reports are archived by the credit bureaus themselves. Therefore, with the aid of TransUnion, credit scores were recalculated for Law Access Loan (LAL) borrowers who had taken out this private loan product during Academic Year 1991/1992 (the oldest group of originations for which we had confidence in the subsequent tracking of repayment experience). Archived credit reports were rescored for 13,508 unique borrowers; these constitute the database upon which the remainder of this paper's models are based. The recalculated credit scores were "as of" the approximate time that the loans were underwritten for the upcoming 91/92 Academic Year (June, 1991).

In the analysis which follows, each borrower's credit score was used as the single summary measure of their relative credit-worthiness. Although use of this borrower-based measure has not been previously reported in the student loan default literature, it is a variable which quite directly captures the "borrower-risk proclivity" construct that this literature has generally found most predictive of default. But rather than use the raw scores themselves in this study, each borrower was classified into one of four risk classes based on their score. The categories were determined by first finding ranges which divided the 13,508 borrowers into roughly equal thirds. The two best risk class borrowers were designated "low risk" and "moderate risk," respectively. The bottom third of borrowers, which comprise roughly what the consumer credit industry would normally rate "sub-prime," was then subdivided again into "high risk" and "very high risk" borrowers. The cut-off between these two latter categories was set at the score below which a co-signer was required, in a recent tightening of the Law Access Program credit requirements, for the 1997/1998 Academic Year. For the 1991/1992 retroactive sample, 9% of borrowers had scores which would have placed them in the "very high risk" class.

2b) Credit Class Analysis

To test the predictiveness of the credit score classifications, two models were initially built, then later combined. First, a simple test of credit classification as a predictor of default prior to the end of the 9th quarter after graduation, the seasoning period, was performed. Second, a model was constructed of credit classification as a default predictor after the seasoning period. This latter model takes into account the time pattern of default by incorporating the specific functional form discussed earlier for loan mortality among borrowers surviving into the "post-seasoning" period.
Model 1

Institutional reasons largely explain why no simple conditional default pattern is detectable for loans within the first 9 quarters after borrowers’ graduation. First, all borrowers are generally allowed 2 quarters after graduation in which, by contract, no repayment of principal or interest is due ("grace period").(19) This is in addition to an "in-school" period, generally 3 years for law students, in which, likewise, no payment is required (interest does accrue during grace and in-school periods and is capitalized onto principal once when the loan enters repayment). If a borrower is unable to make regular principal and interest payments, the lender may accept interest-only payments, or the borrower can defer all payments with a capitalized interest forbearance. An initial forbearance may be granted for up to six months, with an extension to twelve months often available. The lender has typically been generous in granting forbearance. Note that the quarter with the single highest conditional default rate is the 6th after graduation. Although this peak could, in theory, be accounted for by students who default after a short forbearance, it is more likely, because of the very long procedural delays involved in having a guaranteed claim actually paid, to represent students who made no efforts to repay nor even to apply for a forbearance. Instead, a secondary peak occurring one year later, 10 quarters after graduation, is likely to represent largely those borrowers who default after a full twelve months of forbearance (again taking into account the procedural delays involved in having a claim actually paid).

When a student is in forbearance, none of the procedural steps required to declare a loan in default and have a claim paid are pursued. These procedural steps themselves involve considerable time delay. At about 75 days delinquent an account is referred to TERI, the guarantor, for "pre-claims assistance." At between 120 and 180 days delinquent a default claim is formally filed. Of course, any payments received on delinquent accounts reduce a borrower’s "days-delinquent" measure and forestall, at least temporarily, a claim being filed. Ultimately, an account may be as delinquent as 240 days before a cash claim is actually paid. Claim payment is what is represented in our data as default. In essence, then, the period we’ve called seasoning is one in which avoidance of default is negotiated on a loan-by-loan basis.

The first modeling question is simply whether a borrower’s credit category (low risk, moderate risk, high risk, or very high risk) predicts default during the period prior to the end of the 9th quarter after graduation (or withdrawal) from law school. Although several techniques are available for modeling this case of a dichotomous dependent variable and a single four level predictor variable, the logit technique is used here because it is an approach employed in our later, more involved models and has the advantage of a straightforward interpretation to its coefficients.(20) The logit model expresses the risk of default as "odds" rather than as simple probability, but the relationship between the two is simple, probability = odds/(1+odds), and for small probability values, they are nearly identical. The logit model says that the log of the odds of default can be expressed as a linear function of a set of predictor variables:
\[
\log \text{odds}_i = \log \frac{P_i}{1-P_i} = \alpha + \beta_1 x_{1i} + \beta_2 x_{2i} + \ldots + \beta_k x_{ki}.
\]

In our first logit run the predictor variables are 3 dummies representing 3 of the 4 levels that credit category can take (the 4th level serving as a reference point).

Table 1 summarizes the results of this first run. The table displays estimated coefficients for each of three levels of credit risk, their associated standard errors and chi-squared statistics. Also shown is the chi-squared statistic for all three levels of credit category taken together and the log likelihood statistic, a measure of overall model goodness-of-fit. All test statistics show significance at a .0001 (very high) level of confidence.

**TABLE 1**

**Model 1: Predicting Default Odds During Seasoning Period From Borrower Credit Risk Category**

<table>
<thead>
<tr>
<th>Variable</th>
<th>Coefficient Estimate</th>
<th>Standard Error</th>
<th>Chi-Squared</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>-2.165</td>
<td>.095</td>
<td>517.2****</td>
</tr>
<tr>
<td>Credit Class</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Low Risk</td>
<td>-1.765</td>
<td>.145</td>
<td>148.7****</td>
</tr>
<tr>
<td>Moderate</td>
<td>-1.103</td>
<td>.123</td>
<td>79.8****</td>
</tr>
<tr>
<td>High Risk</td>
<td>-0.484</td>
<td>.118</td>
<td>16.7****</td>
</tr>
</tbody>
</table>

Together: 183.2****

Log Likelihood: -2336.77

****Significant at the .0001 confidence level

The intercept represents an estimate for the log odds of default for the reference level of the risk category, that is for borrowers categorized "very high risk." Taking the antilog of the intercept, exp(-2.165), gives the odds for default for very high risk borrowers during the period of seasoning at .115 or a bit higher than 1 default to every 9 non-defaults (approx. probability = 10%). To assess the odds of default for borrowers at any other risk class, the coefficient associated with that risk class is summed with the intercept value before the antilog is taken. For example, the odds of a low risk borrower defaulting during the period of seasoning is exp(-2.165-1.765) = .020 or
1 to 50 (approx. probability = 2%). Perhaps more important than the statistical significance demonstrated in Table 1 is the striking substantive interpretation that can be given the model's coefficients. At least during the period of seasoning, the probability of a very high risk borrower defaulting is roughly 5 times the probability of default by a low risk borrower.

Model 2 (A and B)
That credit score category should predict early default is not really surprising; generic credit scores were, in fact, developed from data that tracked borrowers' changed credit history profile over a fairly short 2-year time horizon (short, at least, when compared with the 15 year term of the student loans being examined here). The key question for this section's second model is whether credit score at time of underwriting is predictive of default risk beyond two years. All seasoned loans in our sample, simply by virtue of the definition of the seasoning period, were beyond this two year window. Does credit score category retain its statistical significance (and substantive predictiveness) for "post-seasoned" default risk?

Because the earlier survival analysis suggested a general mortality pattern for post-seasoned loans, any predictive power attributable to credit category must be incremental to the effect of the simple passage of time on default risk. Recall that a specific functional form for the hazard function for post-seasoned loans was identified, the Weibull distribution. To capture this Weibull mortality pattern precisely we need to model the hazard function directly as the dependent variable rather than model odds (as was done with the logit). The widely used proportional hazards model of Cox (1972) can be applied and, rather than "log odds" estimated as a linear function of predictor variables, a transformation known as the complementary log-log is estimated. Specifically, for individual i, we estimate:

$$\log [-\log (1 - P_i)] = \alpha + \beta_1 x_{1i} + \beta_2 x_{2i} + \ldots + \beta_k x_{ki} + \log t$$

The data set was restricted to borrowers who were "alive" (had not defaulted) as of the beginning of their 10th quarter after graduation. As in Model 1, we employ 3 dummy variables to capture the effect of credit risk categorization, but now also include as a separate predictor the log of time, measured in quarters. This specification estimates a Weibull model for the effect of time on default risk.(21) Table 2A gives the results of the complementary log-log run for post-seasoned data. As can be seen, all credit class variables remain significant even in the presence of the log time variable. Log time, as expected, is also significant.
TABLE 2A

Model 2A: Predicting Default Hazard in the Post-Seasoning Period From Borrower Credit Risk Category (complementary log-log formulation)

<table>
<thead>
<tr>
<th>Variable</th>
<th>Coefficient Estimate</th>
<th>Standard Error</th>
<th>Chi-Squared</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>-3.315</td>
<td>.122</td>
<td>734.8****</td>
</tr>
<tr>
<td>Credit Class</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Low Risk</td>
<td>-1.504</td>
<td>.162</td>
<td>86.2****</td>
</tr>
<tr>
<td>Moderate</td>
<td>-1.090</td>
<td>.143</td>
<td>58.5****</td>
</tr>
<tr>
<td>High Risk</td>
<td>-0.522</td>
<td>.135</td>
<td>14.9****</td>
</tr>
<tr>
<td>Together</td>
<td></td>
<td></td>
<td>108.6****</td>
</tr>
<tr>
<td>Log-of-Time-Period</td>
<td>-0.687</td>
<td>.070</td>
<td>95.1****</td>
</tr>
</tbody>
</table>

Log Likelihood:  -2387.2

****Significant at the .0001 confidence level

TABLE 2B

Model 2B: Predicting Default Odds in the Post-Seasoning Period From Borrower Credit Risk Category (logistic regression formulation)

<table>
<thead>
<tr>
<th>Variable</th>
<th>Coefficient Estimate</th>
<th>Standard Error</th>
<th>Chi-Squared</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>-3.302</td>
<td>.124</td>
<td>711.6****</td>
</tr>
<tr>
<td>Credit Class</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Low Risk</td>
<td>-1.512</td>
<td>.163</td>
<td>85.9****</td>
</tr>
<tr>
<td>Moderate</td>
<td>-1.096</td>
<td>.144</td>
<td>58.2****</td>
</tr>
<tr>
<td>High Risk</td>
<td>-0.525</td>
<td>.136</td>
<td>14.8****</td>
</tr>
<tr>
<td>Together</td>
<td></td>
<td></td>
<td>108.2****</td>
</tr>
<tr>
<td>Log-of-Time-Period</td>
<td>-0.690</td>
<td>.071</td>
<td>94.8****</td>
</tr>
</tbody>
</table>

Log Likelihood:  -2387.3

****Significant at the .0001 confidence level
Although the complementary log-log model is the only method to precisely estimate the Weibull distribution for the effect of time on dichotomous default risk, fortunately the logit model often returns estimates and test statistics very close to those of the complementary log-log transformation. Accordingly, logistic regression (logit) results are presented in Table 2B, which upon inspection do, indeed, approximate those of Table 2A. The model which generated the Table 2B results is similar to that constructed as Model 1 above except that the log of time is entered as an additional predictor and now the data set is comprised of only those borrowers with non-defaulted loans entering the post-season period.

A more direct comparison can now be made between the estimates derived in logit Model 1 for loans in their period of seasoning and the logit results displayed in Table 2B for loans in the post-seasoned period. First, it is worth noting the substantive effect of time on default odds as calculated in Table 2B. The coefficient for log time is -.69. This means that for every doubling of the time measure (e.g., going from the first period of the post-season, Quarter 10, to the second post-season period, Quarter 11; or from Quarter 11 to Quarter 13, the fourth quarter of the post-season) the odds of default drop by 38%.

Now consider the credit class variables. Model 2B implies that the odds of default for high risk borrowers is .037, or a 3.5% default probability, in Quarter 10 alone (this compares to an approximate 10% probability of default implied for high risk borrowers during the entire seasoning period). The odds of default for low risk borrowers in Quarter 10 is the antilog of the intercept term aggregated with the coefficient for “low risk class borrower,” or \( \exp(-3.302-1.512) = .008 \), equivalent to a .8% quarterly conditional default probability. The ratio of 3.5% to .8% (=4.4X) demonstrates the persistent predictive power of credit score, indeed "stale" scores, in assessing even later stage default risk. As modeled, this differential in default risk holds not only in Quarter 10, but for all post-seasoning quarters. Borrowers whose credit scores at the time of loan underwriting indicated that they were very high risk credits were still roughly 4 1/2 times more likely to default than low risk borrowers even among those individuals who had not yet defaulted by the end of the 9th quarter after graduation. Recall that the ratio of the probabilities for these same two risk classes was a similar 5-times during the period of seasoning (Model 1).

**Combined Model**

Having uncovered a persistent effect of credit risk category as a predictor of default even among post-seasoned loans, we now combine into a single model the data which were separately analyzed above as Models 1 and 2B. Recall (footnote 21) that the logistic regression of Model 2B employed a pooled borrower cross-section/time series database which contained, for each borrower surviving into "post-seasoning," one observation for each quarter up to and including the quarter in which the borrower either defaulted or in which observations on that borrower were "censored" (that is, last observed as a non-defaulter in the last observation period). In the combined model, "post-seasoning" time periods were re-indexed to the beginning of each borrower’s post-season; for example, observations on borrowers made in the 10th quarter after their graduation from law school were re-indexed to time Period 1, the first quarter of their post-season. To build the combined model, the Model 2 database was appended with one observation
each for all borrowers (whether or not they survived into a post-season), each borrower coded as having defaulted or not sometime during the multi-quarter seasoning period; for consistency this time period, although longer than subsequent periods, was indexed as Period 0. The logistic regression of Model 2B was then re-run with a new dummy variable labeled “Seasoned?” which was set to NO if the observation was for the borrower during the period of seasoning (Period 0) and YES otherwise. In addition, for this re-run the log time variable was treated analogously to an interactive variable, only entering the model when the dummy “Seasoned?” was YES, that is, only for observations during a borrower’s post-seasoning period. Results of the combined model are given in Table 3 below.

**TABLE 3**

**Combined Model: Predicting Default Odds From Borrower Credit Risk Category**

<table>
<thead>
<tr>
<th>Variable</th>
<th>Coefficient Estimate</th>
<th>Standard Error</th>
<th>Chi-Squared</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>-2.176</td>
<td>.077</td>
<td>795.4****</td>
</tr>
<tr>
<td>Credit Class</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Low Risk</td>
<td>-1.658</td>
<td>.108</td>
<td>234.5****</td>
</tr>
<tr>
<td>Moderate</td>
<td>-1.100</td>
<td>.094</td>
<td>138.0****</td>
</tr>
<tr>
<td>High Risk</td>
<td>-0.502</td>
<td>.089</td>
<td>31.6****</td>
</tr>
<tr>
<td>Together</td>
<td></td>
<td></td>
<td>290.4****</td>
</tr>
<tr>
<td>Log-of-Time-Period</td>
<td>-0.690</td>
<td>.071</td>
<td>94.8****</td>
</tr>
<tr>
<td>Seasoned?</td>
<td>-1.110</td>
<td>.087</td>
<td>164.3****</td>
</tr>
</tbody>
</table>

Log Likelihood -4725.25

****Significant at the .0001 confidence level

The intercept of the combined model represents the log odds of default for very high risk borrowers (the reference credit class) during the seasoning period (the reference value for “Seasoned?” = NO) where log-of-time-period variable does not come into the model, as it does not for borrowers still in their seasoning period. The odds of default for very high risk borrowers during the seasoning period, therefore, is exp(-2.176) = .113, or a default probability of 10% which is the same as was derived in Model 1 for very high risk borrowers during their seasoning period. For low risk borrowers the default odds are calculated, as in Model 1, by simply aggregating the coefficient associated with the “low risk” credit class with the intercept before taking the antilog; that is, the default odds are exp(-2.176-1.658) = .022, or a default probability of 2% for low risk borrowers during their period of loan seasoning, again roughly equivalent to the figure obtained earlier from Model 1. Furthermore, the approximate 5X ratio of default likelihood between very high risk and low risk borrowers obtains as it did in Model 1.
3. School-of-Attendance, Geographic and Economic Variables as Default Predictors

3a) Statistical Association of School-of-Attendance with Student Loan Repayment Behavior

All previous literature on student lending has de-emphasized the predictive value of knowing a student’s school-of-attendance, after having accounted for individual borrower attributes, when assessing default risk. However, only one of the studies reviewed here, Knapp and Seaks (1992), had explicitly tested for a “school effect” at the individual institution level. Most studies had instead first grouped schools by type (e.g., proprietary vs 2-year college vs 4-year college vs doctoral university) then sought a relationship between type-of-school and default; or, alternatively the studies sought a statistical relationship between a set of school characteristics (e.g., proportion of minority students, expenditures per student) and the likelihood of borrower default among graduates. We follow Knapp and Seaks here, entering into the "combined" model described above a set of 126 dummy variables representing the possible schools-of-attendance (i.e., law school graduated or last attended) with which each borrower can be uniquely associated, with a 127th school acting as a “reference school.”(25) Only data from students attending 127 schools were useable here (vs data from students at 168 schools used to construct the “combined” model of Table 3) because some low volume schools had no defaulters (or, as in one case, the single borrower from a school defaulted). The number of useable borrower records was thus reduced from 13,508 to 13,003. Results of the logistic regression incorporating both borrower characteristics (i.e., credit category) and school-of-attendance are given in Table 4. Statistics for only the combined effect of all school-of-attendance dummy variables, taken together, is shown here. Notably, this variable set does significantly add to the predictiveness of the model even after accounting for borrower credit score categories. Indeed, the set of variables is highly significant, judging from the associated chi-squared statistic. Furthermore, the log likelihood statistic of the new model increases (gets less negative) by 171.8 when compared to the log likelihood statistic of the "combined" model (Table 3), an improvement which is again highly significant with 126 degrees of freedom (equal to the number of dummies entered into the new model). This finding is in stark contrast to the results reported by Knapp and Seaks whose set of 25 school-specific dummies failed to significantly add to the predictiveness of their student default model. In further contrast to Knapp and Seaks, who reported that none of their individual school dummies were significant, results here indicate that students attending 11 specific schools had significantly worse (at a .05 confidence level) default odds than students attending the reference school; and students at 2 schools had significantly better (at the .05 confidence level) default odds than reference school students, even after accounting for each borrower’s individual credit risk.

To better assess the substantive (as opposed to the statistical) significance of school-of-attendance as a predictor of default, the original set of school dummies was collapsed into just two. The first of the two new dummies represented attendance (or not) at one of the 11 schools whose students had had a significantly worse default history than would have been expected, even after taking individuals’ credit scores into account; the second dummy represented attendance (or not) at one of the 2 schools whose students were indicated to be significantly better default risks. All students attending the other 114 schools (including the previous reference school) were formed into a new reference group. Results from a logistic regression on the newly-coded data are summarized in Table 5.
TABLE 4 (Model 4): Predicting Default Odds From Borrower Credit Risk Category and Identity of School-of-Attendance

<table>
<thead>
<tr>
<th>Variable</th>
<th>Coefficient Estimate</th>
<th>Standard Error</th>
<th>Chi-Squared</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>-2.220</td>
<td>.223</td>
<td>98.8****</td>
</tr>
<tr>
<td>Credit Class</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Low Risk</td>
<td>-1.630</td>
<td>.110</td>
<td>219.0****</td>
</tr>
<tr>
<td>Moderate</td>
<td>-1.080</td>
<td>.096</td>
<td>127.1****</td>
</tr>
<tr>
<td>High Risk</td>
<td>-0.507</td>
<td>.091</td>
<td>30.7****</td>
</tr>
<tr>
<td>Together</td>
<td></td>
<td></td>
<td>268.0****</td>
</tr>
<tr>
<td>Log-of-Time-Period</td>
<td>-0.673</td>
<td>.071</td>
<td>89.2****</td>
</tr>
<tr>
<td>Seasoned</td>
<td>-1.110</td>
<td>.087</td>
<td>161.9****</td>
</tr>
<tr>
<td>School Identity</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>126 dummies together</td>
<td></td>
<td></td>
<td>236.6****</td>
</tr>
<tr>
<td>Log Likelihood</td>
<td>-4553.44</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

****Significant at the .0001 confidence level

TABLE 5 (Model 5): Predicting Default Odds from Borrower Credit Risk Category and School of Attendance Categorization

<table>
<thead>
<tr>
<th>Variable</th>
<th>Coefficient Estimate</th>
<th>Standard Error</th>
<th>Chi-Squared</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>-2.204</td>
<td>.079</td>
<td>776.3****</td>
</tr>
<tr>
<td>Credit Class</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Low Risk</td>
<td>-1.649</td>
<td>.109</td>
<td>230.0****</td>
</tr>
<tr>
<td>Moderate</td>
<td>-1.110</td>
<td>.094</td>
<td>138.6****</td>
</tr>
<tr>
<td>High Risk</td>
<td>-0.530</td>
<td>.090</td>
<td>34.7****</td>
</tr>
<tr>
<td>Together</td>
<td></td>
<td></td>
<td>282.4****</td>
</tr>
<tr>
<td>Log-of-Time-Period</td>
<td>-0.680</td>
<td>.071</td>
<td>91.6****</td>
</tr>
<tr>
<td>Seasoned</td>
<td>-1.119</td>
<td>.087</td>
<td>165.4****</td>
</tr>
<tr>
<td>School Category</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Positive effect</td>
<td>-0.831</td>
<td>.223</td>
<td>13.9***</td>
</tr>
<tr>
<td>Negative effect</td>
<td>0.957</td>
<td>.089</td>
<td>115.0****</td>
</tr>
<tr>
<td>Together</td>
<td></td>
<td></td>
<td>134.2****</td>
</tr>
<tr>
<td>Log Likelihood</td>
<td>-4616.85</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

****Significant at the .001 confidence level

***Significant at the .001 confidence level
The coefficients for both the “positive” and “negative” school effects are statistically significant (although the “positive” school effect at a lesser level of confidence, albeit still quite high by traditional standards). The substantive importance of the apparent “school effect” can be gauged by examining the antilog of the coefficients for both “positive” and “negative,” exp(-.831) = .44 and exp(-.957) = 2.60, respectively. Compared to students graduating from a reference group school, students attending the 2 “positive” schools had less than half (44%) the odds of defaulting even after taking into account all other predictors. On the other hand, students attending the 11 “negative” schools had a greater than 2 1/2 times (2.6 X) higher odds of defaulting than reference school students, all other things being equal. Comparing the relative odds of default for students at “positive” vs “negative” schools leads to the interpretation that, even after controlling for individual borrower credit characteristics, “negative” school graduates have about a 6 times (2.6/.44) greater odds of defaulting than “positive” school graduates. Note that the credit rank coefficients remain relatively stable in the new model. The coefficient for “low risk” borrowers is -.65, meaning that the odds of default for these borrowers, all other things (including school) being equal, are only 19%, exp(-.65), of those of the reference credit class, “very high risk” borrowers. Said another way, even after controlling for any effect of school-of-attendance, very high risk borrowers remain, as in earlier models, roughly 5 times more likely to default than do borrowers classed as low risk.(26)

3b) School Reputation, Economic and Geographic Considerations

Although evidence for a statistically significant association of school-of-attendance with default likelihood has been uncovered, this does not necessarily imply that a school’s specific administration or faculty have any influence per se, for better or worse, on their graduates’ loan repayment behavior. Our findings are completely consistent with the observation made by Knapp and Seaks, based on their own very different results, that any two schools “could exchange the entire staffs of their financial aid offices and continue to observe essentially the same default rates” (p. 411). It is possible, for example, that a school’s historic reputation may color employers’ views concerning the quality of the school’s law graduates and that it is through that relative perception that graduates get higher or lower paying jobs (or jobs at all). Employment opportunities, in turn, may be the more direct influence on default likelihood.

To test for the possible influence of perception of school reputation on default, each borrower record was appended with the value for median 1995 matriculate LSAT score for the borrower’s school-of-attendance. The model of Table 5, with this new predictor variable added, was then rerun. Any measure of school reputation is inherently controversial. Median LSAT has the advantage, at least, of being a standardized measure. And, controversial or not, average matriculate LSAT of a graduate’s school of attendance is a variable we would be negligent not to include in a model of loan repayment, especially if potential employers may be suspected of using it to select recruiting venues.(27)

When entered into the model, median LSAT score for school attended is a significant predictor of default odds (at the .0001 level of significance). Yet all other previous predictor variables also remain significant at their previously high confidence level (except the "positive" school effect
which, while significant, drops to a .05 level). Coefficients on all but the school class predictor variables remain very close to their Table 5 values. However, with the addition of the median LSAT variable, both school rank dummies indicate less of a differentiation of "positive-" and "negative-" classed school students from the set of reference school students in predicted default risk and less separation between the two extremes themselves, “positive” and “negative” school students. After taking into account all other predictors, including now also median school LSAT, a student attending a school classed as “negative” has a 3.5 times higher odds of default than a student at a “positive”-classed school. This compares to a 6 times differential before the median LSAT variable was considered.

Before examining this paper’s concluding model, consider two additional constructs. First, if as was suggested above, economic factors are truly those which are most directly influential on loan repayment behavior, then some measure of the economic climate facing each law school student upon graduation might prove quite useful in predicting default likelihood. The National Association for Law Placement (NALP), an association of law school placement officers, publishes a statistical series which traces the percentage of law school students from each graduating class unemployed (and seeking employment) six months after graduation. Furthermore, since at least 1992, NALP has published this series by region based upon reports received from schools in each of nine geographic divisions of the country concerning the employment success of their graduates.(28) Nine regional statistics were available, therefore, for each of the three graduating law school classes (1992, 1993, and 1994) investigated in this research. A variable was formed from these NALP data and appended to each borrower record in the dataset reflecting the unemployment rate among all law students graduating in a borrower’s same graduation year from schools in the same region as the borrower’s school. The measure, therefore, may be thought of as reflecting any supply/demand imbalance for new lawyers graduating from schools in the region of the borrower’s own school during the same period the borrower would have been looking for his or her first job as an attorney.

Next, a need to account for specific regional differences is suggested by an examination of the location of the schools categorized as having a "negative" school effect. Ten of the 11 negatively-classed schools were located in just 3 of the 9 geographic regions of the country, the Pacific (West Coast states plus Hawaii), the West South Central Region (Texas, Oklahoma, Arkansas, and Louisiana) and the South Atlantic states (the East Coast from Florida up to Delaware, including West Virginia and Washington, D.C.). The 11th school was located in a state contiguous to the West South Central Region. Accordingly, a new dummy variable (call it, imprecisely, the "Sunbelt" effect variable) was formed and was set to 1 for all borrowers graduating from schools in one of the 3 above-mentioned regions and 0 otherwise.

Employing the three new constructs introduced in this section, 1)perceived school reputation, measured by the median matriculate (1995) LSAT score for the borrower’s school-of-attendance, 2)regional labor market supply/demand imbalances, measured by the Regional Law Graduate Unemployment Rate for the borrower’s graduation year, and 3)specific geographic differences, measured by the dummy variable "Sunbelt," differentiating law school graduates from the Pacific, West South Central, and South Atlantic Regions from law school graduates elsewhere, plus certain interaction effects between these three new constructs, the paper’s final model may be constructed.
This model's statistics are displayed in Table 6. The model's goal was the prediction of default odds using variables that substitute for the explicit use of school-identity dummies or the use of school-effect aggregates. Consequently the "positive" and "negative' school effect variables used in Model 5 are removed in this final model, Model 6. The premise is that the apparent "school effect" may actually be masking underlying economic determinants of default likelihood.

When only the new main-effect variables, median LSAT, Regional Law Graduate Unemployment Rate and the "Sunbelt" dummy, were entered into a preliminary model (not shown here) substantial colinearity between the "Sunbelt" and Unemployment variables resulted, requiring that the latter be dropped as a main-effect predictor in the final model, entering instead as an interactive variable. In fact, two interactive variables were found to incrementally add to the model's predictiveness: 1) "Sunbelt" crossed (interacting with) with median LSAT, defined as the median matriculate LSAT score of school-of-attendance for Pacific, West South Central and South Atlantic Region graduates only (otherwise 0), and 2) "Sunbelt" crossed (interacting with) with Regional Law Graduate Unemployment Rate, defined, for "Sunbelt" graduates only, as the unemployment rate in their region for the year of their graduation (otherwise 0).

<table>
<thead>
<tr>
<th>Variable</th>
<th>Coefficient Estimate</th>
<th>Standard Error</th>
<th>Chi-Squared</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>4.438</td>
<td>1.308</td>
<td>11.5 ***</td>
</tr>
<tr>
<td>Credit Class</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Low Risk</td>
<td>-1.614</td>
<td>.109</td>
<td>219.8****</td>
</tr>
<tr>
<td>Moderate</td>
<td>-1.078</td>
<td>.094</td>
<td>130.5****</td>
</tr>
<tr>
<td>High Risk</td>
<td>-0.496</td>
<td>.090</td>
<td>30.3****</td>
</tr>
<tr>
<td>Together</td>
<td></td>
<td></td>
<td>271.8****</td>
</tr>
<tr>
<td>Log-of-Time-Period</td>
<td>-0.684</td>
<td>.071</td>
<td>92.7****</td>
</tr>
<tr>
<td>Seasoned</td>
<td>-1.124</td>
<td>.087</td>
<td>167.4****</td>
</tr>
<tr>
<td>Median LSAT</td>
<td>-0.043</td>
<td>.008</td>
<td>25.9****</td>
</tr>
<tr>
<td>&quot;Sunbelt&quot; Region</td>
<td>4.079</td>
<td>1.956</td>
<td>4.3 **</td>
</tr>
<tr>
<td>&quot;Sunbelt&quot; X LSAT</td>
<td>-0.029</td>
<td>.013</td>
<td>5.5 **</td>
</tr>
<tr>
<td>&quot;Sunbelt&quot; X Unemployed</td>
<td>4.639</td>
<td>1.341</td>
<td>12.0 ***</td>
</tr>
</tbody>
</table>

Log Likelihood -4623.6

****Significant at the .0001 confidence level
***  Significant at the .001 confidence level
**   Significant at the .05 confidence level
Table 7 below outlines the substantive significance of each of the predictor variables found to be statistically significant in Table 6. Note that the set of “Sunbelt” variables effectively divide the model in two, one for graduates from the Pacific, West South Central and South Atlantic regions and the second for graduates from schools in all other regions. The substantive importance of certain of the predictors, therefore, can be most easily illustrated by examining their influence on default odds for each of these two geographically-defined student populations considered separately. This approach is taken in the bottom section of Table 7 for predictors 4 and 5. For the set of predictors numbered 1-3 (the original variables of the Combined Model of Table 3, Section 2b), their influence on default odds continues to be modeled as independent of school reputation, economics, and geography; the estimated coefficients for these predictors and therefore their substantive importance are the same for all borrowers in all regions. These predictors, whose influence on default odds remains quite consistent with earlier models, are discussed in the top section of Table 7.

**TABLE 7**

**Independent Influence of Predictor Variables on Default Odds as Derived from Model 6**

1. **Credit Class:** A borrower whose credit score at loan origination indicated that he or she was a "low credit risk" has default odds 1/5 those of a "very high risk" borrower, all other things (e.g. time into repayment and year and region of graduation) being equal. The analogous relative odds figures for moderate and high risk borrowers when compared with very high risk borrowers are 1/3 and 3/5, respectively.

2. **Time: Seasoned?** The odds of any particular borrower defaulting in the first quarter after their period of loan seasoning (i.e., the 10th quarter after graduation) are about 1/3 whatever that borrower’s odds had been during their entire seasoning period.

3. **Time: Log-of-Time-Period** For all borrowers who survive beyond the period of loan seasoning, quarterly default odds decrease by 1/3 for every doubling of the time period into that post-season.

**For “Sunbelt”-Regions Graduates:**

4. **LSAT:** For each 1 point drop in the median matriculate LSAT score of the borrower’s school-of-attendance, default odds increase by 7.5%.

5. **Regional Law Grad Unemployment Rate:** Each percentage point increase in regional unemployment for new attorneys is associated with a 4.7% increase in default odds.

**For non-"Sunbelt”-Regions Graduates:**

4. **LSAT:** For each 1 point drop in the median matriculate LSAT score of the borrower's school-of-attendance, default odds increase by 4.4%.

N/A (not applicable in non-"Sunbelt” model)
One difference that the interactive model highlights is an apparent geographic difference in the strength of influence that perceived school reputation, as measured by median matriculate LSAT score, apparently has on default risk. Even outside the three “Sunbelt” regions, a school’s median LSAT is a statistically significant and substantive predictor of graduates’ default risk even after controlling for an individual borrower’s credit profile. The spread in median 1995 matriculate LSAT score for schools outside the “Sunbelt” regions was 24 points (from 171 to 147), implying a predicted default rate 2.8 times higher (=\([1.044]^{24}\)) for graduates from schools with the lowest "reputation" vs graduates from schools with the highest "reputations," all other things (e.g., borrower’s credit profile) being equal. But in the Pacific, West South Central, and South Atlantic regions, at least during the period investigated here, the effect of school reputation appears to have been even more dramatic. For schools within these “Sunbelt” regions, the relative default odds between graduates from schools differing in median LSAT scores by a similar 24 points is 5.6-times (=\(\exp(0.0426 + 0.0294)\))\(^{24}\). And this differential holds even after controlling, as before, for individual borrower credit risk and now also for the statistically significant effect of differing levels of new attorney unemployment during the years 1992-1994.(29)

The effect of new attorney unemployment rate on default risk is a particularly intriguing finding, especially as modeled above as an effect pronounced only in the "Sunbelt" regions.(30) As background, it should be noted that the association of a borrower’s default likelihood with the simple fact of their having graduated from a “Sunbelt” school (the “main” effect) is statistically significant and, generally, positive.(31) That is, for example, consider a graduate from a “Sunbelt” law school with an average matriculate LSAT median (156, the median of medians among all schools in the sample) seeking their first job in a year of average new graduate unemployment in the “Sunbelt” (13.4%, the median unemployment rate recorded among the three “Sunbelt” regions over the three years, 1992-1994). That graduate is modeled as having a nearly 20% higher odds-of-default than an otherwise identical graduate from a non-“Sunbelt” school with the same “average” reputation (as measured by matriculate LSAT median). On top of this is added the variability in default likelihood contributed by swings in “Sunbelt” new graduate unemployment rates themselves, adding 4.7% to a “Sunbelt” graduate’s odds-of-default for each one-percentage point unemployment rate rise. A borrower from our “average” law school who graduated under the best (albeit not good) of employment conditions existing for “Sunbelt” graduates (1993 in the West South Central Region with an unemployment rate of 10.3%) nevertheless had default likelihood only 3/5 that of an otherwise identical graduate seeking a job in the worst of recruiting years in the worst of regions (1992 in the Pacific Region with an unemployment rate of 20.7%). In summary, for a law graduate from an average reputation school, “Sunbelt” regionality is associated with both a higher expected odds-of-default and greater variability in those odds, the latter because of swings in new attorney unemployment rates.
V. Discussion and Conclusions

Perhaps the question that has set the agenda for student loan repayment research, "Whose fault is default," is simply the wrong question to ask. At least as it pertains to the narrow category of law school borrowers, student loan default is most constructively analyzed not as an issue of morality or blame, but as one of economics. Applying the commercial bankers' old platitude, loan repayment is simply a matter of the borrower's ability and willingness to repay.

In this research "willingness to repay" may be thought of as having been proxied by an individual’s credit bureau score at the point of loan origination. This metric has been long accepted among consumer lenders as a reliable measure of a loan prospect’s past credit behavior and as an statistical predictor of future such behavior. However, this is the first published research, we believe, demonstrating generic credit score to be a robust predictor of default odds for privately insured student loans. Using a survival analysis approach, it was demonstrated that the predictive power of generic credit scores persists even beyond the period of loan seasoning, a period which itself stretches more than two years after a student's law school graduation and from 3-5 years after the loan was originated and the credit score calculated.

Not only was a persistent and statistically significant association found between an individual borrower’s credit score classification and their default odds, but this relationship was found to have substantive and, above all, quite practical implications. Guided (misguided?) in part by the conventional student loan research agenda, this paper has devoted considerable space searching for default predictors underlying a statistically significant “school-effect” uncovered in the data, considerations such as school reputation, geography and macroeconomic conditions. And, although these results are certainly of academic interest and offer some mild credit management direction, they can also become a practical distraction. At the end of the day, the borrower’s individual credit profile remains of paramount importance in predicting relative default risk. Consider the predictions of Model 6, the paper’s final and most elaborate derivation. In this model Yale University, with its top reputation as measured by its high median LSAT score, is estimated to produce law graduates with the best default profile of all non-"Sunbelt" schools (and, because it is a non-"Sunbelt" school its graduates' default odds are unaffected by regional unemployment swings). Nevertheless, a very high risk individual borrower graduating from Yale is a poorer default risk, according to Model 6, than a low risk borrower who happens to have graduated from a school with even the very lowest reputation among all schools in the Pacific Region in 1992, the year when graduates from this “Sunbelt” region’s low-reputation schools were most disadvantaged. The borrower-based credit effect simply swamps any school reputation, region or macroeconomic considerations.

Although discounting its practical importance, as mentioned above, this research, does ironically uncover a statistically significant association of school-of-attendance with graduates’ default risk. This is a relationship all previous literature had failed to document, generally to the relief of the investigators analyzing the data. Why relief? Within the framework of “Whose fault is default,” evidence of a “school effect” in predicting graduates’ repayment behavior, incremental to the effect of borrower-based predictors, would naturally have lead to an uncomfortable discussion of what one school may be doing better than another school. But, again, the discussion is only
uncomfortable if “blame” (or credit) is at issue. Perhaps a school can do nothing differently to produce any appreciable improvement (or deterioration) in its graduates’ loan repayment prospects.

A statistical “school effect” may simply be one manifestation of graduates’ “ability to repay” their student loans. Indeed, our results indicate that it is another set of variables, more economically-inspired predictors based upon a school’s location, its reputation, and in some regions, the prevailing attorney labor market, which may actually lie behind the apparent school effect. It is tempting to weave new hypotheses concerning the determinants of loan default from the results of our final model. Could it be that certain regions of the country have structural imbalances (too many law schools serving essentially local labor markets)? Could this structural imbalance explain why schools in some regions produce graduates whose prospects are more sensitive to school reputation than in other regions?(33)

While interesting, such speculation is of limited practical importance. If it is ability to repay that we are seeing manifested, then anything that helps or hinders a student’s capacity to service whatever debt they may have incurred should be a predictor of repayment behavior. A student’s individual success (or failure) in gaining employment and the direct determinants of that success are examples of default predictors which might prove more powerful than such considerations as “school-of-attendance,” or regionality which, at best, are only indirectly related to employment success. If we wish to move beyond “willingness to repay” in seeking default predictors (which, as a practical matter, we may not need to do), then we should look to indicators which measure, as directly as possible, an individual student’s future “ability to repay.” In short, we will search for different predictors of student loan default, perhaps with greater success, to the extent that we jettison the older question of “Whose fault is default,” place the responsibility squarely on the borrower, and seek to understand what separates borrowers who are both willing and able to meet their student loan obligations from ones who either cannot or will not.


(6) “Stafford” loans are of two types, subsidized and unsubsidized. A subsidized Stafford loan is a low interest loan available only to students qualifying on a “needs” test. The federal government pays the interest on the loan while the student is in school and during certain grace and deferment periods. As of the time of this research (May 1998) a graduate or professional school student could borrow a maximum of $8,500 a year in subsidized Stafford loans. An unsubsidized Stafford loan is a federally-guaranteed loan offered at the same low rates as subsidized Stafford loans, but not requiring a “needs” test (although overall student borrowing cannot exceed the “cost of attendance” as certified by the student’s school). A graduate student may borrow a maximum of $18,500 per year in total subsidized and unsubsidized Stafford loans up to a cumulative maximum of $138,500, including any Stafford loans taken out as an undergraduate. The federal government does not pay the interest on unsubsidized Stafford loans, although the interest need not be paid out during the in-school, grace and deferment periods and instead can be (and normally is) capitalized onto the loan principal. Both subsidized and unsubsidized Stafford loans are variable interest rate loans with an interest rate cap of 8.25%.

(7) The examined institutional factors included three of seven "good practices" endorsed by the California Student Aid Commission which had each been found to be statistically correlated to lower default in the absence of all other possible predictors. The one exception referred to in the text was "type of school" (community college vs proprietary) which increased the model’s predictiveness statistically (its canonical correlation coefficient increased), but left the proportion of defaulters that could be correctly forecast essentially unchanged.
(8) For Academic Year 1997/98, Access Group credit criteria required that a borrower have:
 o no more than one account currently rated 60 or more days delinquent at the time of
   the credit report;
 o not more than two accounts that have been 60 or more days delinquent in the past two
   years;
 o no account that has been delinquent 90 or more days in the past five years (credit reports
   listing accounts with a status of "Not Paid as Agreed" are classified as delinquent 90
   days);
 o no record of a paid or unpaid collection or charged-off account in the past two years;
 o no record of a foreclosure, repossession, open judgement or suit, unpaid prior educational
   loan default, or other negative public credit record items in the past six years;
 o no record of a bankruptcy in the past seven years; and
 o no more than three inquiries to an authorized credit reporting agency in the past six
   months.

This last criteria was not in place when the loans being investigated in this current study were
originated.

(9) For a popular discussion of Halley’s investigation and other early developments of life
    table methodology see Bernstein, Against the Gods, esp. pp. 84-88.

(10) Allison, Survival Analysis Using the SAS System (p. 17) offers the following caution in
     interpreting the hazard function as an instantaneous probability:

     "Although it may be helpful to think of the hazard as the instantaneous probability of an
     event at time t, it’s not really a probability because the hazard can be greater than 1.0...
     With regard to numerical magnitude, the hazard is a dimensional quantity that has the
     form of number of events per interval of time, which is why the hazard is sometimes
called a rate. To interpret the value of the hazard, then, you must know the units in
which time is measured. Suppose, for example, that I somehow know that my hazard for
contracting influenza at some particular point in time is .015, with time measured in
months. This means that if my hazard stays at that value over a period of one month, I
would expect to contract influenza .015 times. Remember, this is not a probability. If
my hazard was 1.3 with time measured in years, then I would expect to contract influenza
1.3 times over the course of a year (assuming that my hazard stays constant during that
year)."

(11) Consider a hypothetical high school biology fruit fly experiment in which we are told that
     there is a constant 10% mortality rate (i.e., .1 hazard rate per period). Out of the 1000 fruit
     flies with which we begin the experiment, 100 are expected to die during the first period. Entering
     period 2 with the remaining 900, we expect 90 deaths. Entering period 3 with the 810 survivors,
     81 deaths are expected, etc. The "conditional probability" of any representative fruit fly's dying
     in any period is constant at 1 in 10; but the risk is conditional upon that fruit fly having survived
     to begin the period. A hazard is defined for a particular point in time so it is redundant to say
     "conditional hazard," yet it is a hazard function (in discrete time) that this fruit fly example is
     attempting to describe. Recall also that it is technically not a "probability" that is being
described but a rate of event occurrence (see footnote 10 above). Nevertheless, when an “event” is non-reoccurring for an individual (like death), interpretation of the hazard rate as an instantaneous probability is quite intuitively appealing.

(12) The construction of this conditional default table requires that we begin to address one of the fundamental issues generally associated with survival analysis, "censored" observations. In the case of the student loan data set interrogated here, observations on an individual borrower are censored (incomplete) when that borrower has not defaulted nor completely repaid his or her loan at the time the data set is formed (strictly defined this is right-censored data; for a fuller discussion of censored data of all varieties see Yamaguchi, Event History Analysis, pp. 3-9). Because the loans comprising this study's student loan portfolio were only about 2 to 5 years into their 15 year repayment terms when the data set was formed, observation of the vast majority of sample borrowers was censored. In applying a survival approach to such data we need to know not only whether a borrower defaulted, but also, if he or she defaulted, how long a period they were "at risk" before default occurred; if they did not (yet) default (or completely repay their loan) we need to know how long (e.g., how many quarters) we have observed them "at risk" before our observation period ended. This "aging" information (associated with each borrower) will be used more thoroughly in the statistical models discussed in this paper's next data analysis section (models incorporating borrower credit risk as a default predictor). For the purposes of constructing the conditional default pattern displayed in Figure 1, however, censoring was taken into account only in that the denominator of each conditional probability calculation for any given quarter includes only borrowers who were observed to have made payments throughout the period (i.e., were not censored that quarter or earlier nor defaulted earlier) plus any borrowers defaulting that quarter. The numerator, of course, consists of a count of the borrowers defaulting during the specific quarter in question. One implication of this approach is that the blue bars on Figure 1 are constructed from the "oldest" borrowers, those who were "at risk" the longest, generally those graduating in 1992 and, to a lesser degree, 1993. It should also be noted here that Figure 1 suggests some borrowers to have defaulted before graduating from law school (red bars in the negative "quarters since graduation" region of the graph). While this is theoretically possible (as when a law student declares bankruptcy but remains in school) most of the anomaly is explained by poor "school separation" data. These students did default, and most probably well within two years after leaving school, but they left school earlier than was originally projected (and for some small minority of loans only projected rather than actual separation dates were available).

(13) For a short discussion of this approach see Allison, p.225.

(14) The "<" term in a Weibull distribution need not necessarily be negative, but if it is, as we assume here, hazard is always decreasing with time.

(15) For a primer on credit scoring and proprietary score card development see Edward Lewis, An Introduction to Credit Scoring, 1994.
(16) Generic or bureau scores are also occasionally referred to as “FICO” scores, named after the Fair, Isaac Company which developed the algorithm for generic scoring. Indeed, although marketed independently by each of the credit bureaus, the actual algorithm used by all the bureaus is identical. Scores calculated by any of the three bureaus should, therefore, be the same. It is important to note, however, that if the detailed credit history data maintained at the bureaus differs (which is sometimes does), then the calculated scores may differ (although they need not necessarily be different if the elements that differ in the credit histories do not enter into the scoring algorithm).

(17) For Law Access Loans disbursed for the 1997/1998 academic year a borrower credit score cut-off was, for the first time, imposed below which a co-signer was required before a student was offered a loan. Furthermore, the “price” of the loan (the guarantee fee) varied depending upon the student’s generic credit score.

(18) Recall that generic bureau scores are only ordinal level predictors (rankings) of creditworthiness to begin with, so an aggregation simply reduces granularity.

(19) For loans originated beginning in academic year 1995/1996 The Access Group extended the grace period to 9 months for Law Access Loans. All loans examined in this section, however, because they were originated in 1991/1992, were unaffected by this change.


(21) In order to perform the complementary log-log analysis on the post-seasoning data (Model 2A) and for all the logistic regressions which follow in this paper (Models 2B, 3-6) we employ the “aging” information discussed above in footnote 12 to first build a pooled cross-sectional/time series data set. A series of records are created for each borrower representing one data point for each time period (normally quarter) in which the borrower was observed, up to the time in which he or she either defaulted or was censored. For a discussion of the construction of such a data set for survival analysis see Allison, especially pp. 211-219. Also of interest is Allison’s discussion of why, contrary to normal statistical intuition, creating multiple observations per individual for survival analysis does not bias standard error estimates nor inflate test statistics (p. 223).

(22) It has been shown, for example, that if the underlying event time data are truly continuous (where the complementary log-log method is most appropriate) yet are grouped together into discrete time intervals for analysis, the logit model (proportional odds model) converges to the proportional hazards model as the interval length gets smaller (Allison, p. 136). For a discussion concerning the theoretical reasons for preferring one method over the other, see Allison, pp 136 and 216-217.

(23) We are no longer estimating a Weibull model but modeling log time to be linear in the log odds of default.
(24) The logistic regression is a "proportional odds" model, meaning that in each quarter of the post-seasoning period the ratio of the odds of default (rather than the ratio of probabilities) between, for example, very high risk borrowers vs low risk borrowers (.037/.008 = 4.6) is the same in each time period. That is, for each doubling of the time since the beginning of the post-season, the odds of default both for very high risk and for low risk borrowers decreases by 38%, but the ratio of their default odds remains 4.6X.

(25) The “reference school” chosen was a high volume law school whose graduates’ historic default rate was near the median of rates among the 127 included schools when the data was segmented along the school dimension only.

(26) As constructed here the substantive importance of the “school” and “borrower” effects appear, at first impression, to be of the same order of magnitude when comparing the extremes (low risk vs very high risk borrowers and “positive” vs “negative” schools). However, these comparisons of extremes are somewhat arbitrary and potentially misleading. While over 40% of borrowers fall into the two extremes along the credit rank dimension, only 12% of the sample borrowers are in the two “school” dimension extremes, “positive” or “negative.”

(27) The median 1995 matriculate LSAT figures were drawn from the 1996 U.S. News and World Report publication America’s Best Graduate Schools, pp.18-24, “A Move to Ethics,” and pp. 87-101. Ideally we might have preferred these relative measures from earlier years because they would have been closer to the time of graduation for the students comprising our sample. However, because of questions concerning the quality of earlier published LSAT figures (see sidebar discussion, p. 20), it was thought prudent to use what were likely to be more accurate, albeit more recent, data to proxy school reputation.

(28) NALP divides the country into regions as defined by the U.S. Census Bureau.

(29) At this stage it would be mere speculation, or more properly mere hypothesis-generation, to try and explain the apparent regional differences in the strength of the influence of median matriculate LSAT score on default likelihood. Nevertheless, one hypothesis might be that certain regions contain a higher proportion of younger law schools which lack long-established alumni networks. Local employers in these regions, therefore, might be expected to rely more upon externally-established reputational measures (e.g., a school’s median LSAT score) to help them make interviewing and recruiting decisions.

(30) This does not mean that employment conditions for new attorneys is unimportant outside of the “Sunbelt” regions but only that, as statistically modeled here, the effect is only detected in these regions. Recall that colinearity between the “Sunbelt” variable and the Unemployment variable prohibited entry of both of these as “main effect” predictors; in Model 6 “Sunbelt” but not Unemployment was included. If instead we model Unemployment as the main effect variable and eliminate the three variables related to “Sunbelt” (the main effect and two interactive variables) the effect in all regions of a one percentage point increase in regional new attorney unemployment is a 3.2% increase in default odds. The colinearity between “Sunbelt” and Unemployment stems from the historic coincidence of there being
generally much higher new attorney unemployment in the regions this paper has labeled "Sunbelt" during the years of the data, 1992-1994, than outside these regions. Thus, with this limited dataset there is no way to differentiate between the possible effect of "Sunbelt" geographic location vs the effect of high unemployment alone. Ideally, we would also like to observe periods of time when "Sunbelt" new attorney unemployment rates were lower than elsewhere to be able to statistically separate the relative influence of these two candidate predictors of default likelihood. As a cause of higher default risk (rather than as a simple predictor) geographic location is hard to argue. Does sunlight cause default? Rather, "Sunbelt" regionality likely proxies for specific regional economic conditions, including perhaps structurally higher-than-average new attorney unemployment rates, a hypothesis that is at least suggested by our 1992-1994 data. But if this is the case, the question then is why such structural differences between regions exist. One possible hypothesis concerning this latter issue is discussed in footnote 33.

(31) The greater sensitivity of borrower default odds to school reputation found among Pacific, West South Central and South Atlantic schools actually cuts both ways. As would be expected, graduates of schools in these regions at the low end of the reputation spectrum have estimated default odds worse than those students graduating from non-"Sunbelt" schools with identical reputation measures; and the differential compounds the lower the median LSAT score. However, less obvious, Model 6 also implies that graduates from "Sunbelt" schools with the very highest reputations will generally have better default odds than those graduating from non-"Sunbelt" schools with identical reputation measures. Nevertheless, the number of schools for whom "Sunbelt" regionality is modeled as an advantage rather than as a disadvantage is very limited at high levels of regional law graduate unemployment. For example, at 13.4\% (the median regional unemployment rate observed during the period under investigation within the "Sunbelt") only graduates from schools with median LSAT scores at or above 163 (only 8 of the 73 schools in the "Sunbelt" regions) had predicted default odds better than identical graduates from schools having identical reputations but located outside the "Sunbelt."

(32) Although credit bureau scores may be said to measure one's past credit practices and, as such, largely predict a "willingness to repay" future obligations, it would be an exaggeration to characterize bureau scores as measuring strictly future "willingness." At least indirectly, bureau scores also measure repayment ability.

(33) Corroborative evidence that the regions this paper has called "Sunbelt" may indeed suffer from some structural imbalance, at least in the short term, between demand and supply for new attorneys may be inferred from demographic data. Analyzing projected regional changes in the "20- to 34-Year Old" demographic cohort through the year 2010, Sam Kipp (1998, pp.13-14) warns of shrinking enrollment pools for law schools in roughly all areas except those that this paper has labeled the "Sunbelt" (areas Kipp lists as favored by demographic trends include Texas, Florida, Georgia, Tennessee, and the West). Perhaps law school capacity has been disproportionally increased in areas projected to be favored by future demographic trends in anticipation of expanded enrollment pools (and higher future demand for law graduates). Ironically, this capacity built ahead of demand and the resulting short
term structural "over-production" of attorneys today could explain why default is a more likely event for current law school graduates in just those geographic regions that the demographers see as most favored economically moving into the 21st century.
REFERENCES


U.S. Department of Education; National Center for Education Statistics; *National Postsecondary Student Aid Study, 1992-93 and 1995-96*.

U.S. Department of Education; Office of Postsecondary Education; Student Financial Assistance Programs; Policy, Training, and Analysis Service; Analysis and Forecasting Division; Loans Branch; *FY 1993 Loan Programs Data Book*.


Figure 1

Quarterly Conditional Default Rates

Quarterly Conditional Default Rate

Quarters since graduation