

January 2005

Multivariate Analysis
of Student Loan
Defaulters at
Texas A&M University

Conducted by TG Research
and Analytical Services

Matt Steiner
Natali Teszler

TG™

Multivariate Analysis of Student Loan Defaulters at Texas A&M University

Conducted by
TG Research and Analytical Services
Matt Steiner
Natali Teszler

Table of Contents

Executive Summary and Highlights	3
Introduction	5
Prior Research on the Factors Relating to Student Loan Default	5
Prior Research on the Factors Related to Student Loan Default at Texas A&M University (TAMU)	8
Methodology for Multivariate Analysis of Defaulters at TAMU	9
Variable Selection Process	9
Results of the Multivariate Analysis	9
Grade Point Average	13
Exit Counseling (In-person)	13
Graduation Indicator	14
College/School Last Attended	14
Age of Borrower Entering Repayment	14
Number of Hours Failed	14
Race/Ethnicity	14
Highest Educational Level of Student's Mother	15
Expected Family Contribution	15
Number of Hours Transferred	16
Adjusted Gross Income of Student	16
Gender	16
Model Performance	17
Distribution of Probabilities	18
Kolmogorov-Smirnov (K-S) Statistic	19
C Statistic	20
Receiver Operating Characteristic (ROC) Curve	20
Classification Matrix and Misclassification Rate	21
Uses of the Findings and the Model	22
Appendix	24
Bibliography	29

Multivariate Analysis of Student Loan Defaulters at Texas A&M University

Conducted by
TG Research and Analytical Services
Matt Steiner
Natali Teszler

Executive Summary and Highlights

As a follow-up to the 2002 bivariate analysis of 12,776 undergraduate borrowers who attended Texas A&M University in College Station, the present study provides a more in-depth examination of the same group of borrowers. Both studies looked at borrowers who entered repayment on TG-guaranteed Federal Family Education Loan Program (FFELP) loans during federal fiscal years 1997, 1998, and 1999. However, the present study employs an analytical technique that more precisely isolates the independent relationship of each variable to default after accounting for the relationships of other relevant variables. It both reinforces some of the findings of the earlier study and highlights some variables that were largely dismissed by the original work.

Like the earlier study, the present study finds that the borrower's success in college – as measured by college grade point average, college or school last attended at the University, number of hours failed and graduation status – has an important relationship to whether the borrower defaults on student loans. In general, the more academically successful the student borrower is, the less likely the borrower is to default. This is good news for the federal student loan programs, which are predicated upon the assumption that students who finance their educations will be making worthwhile investments that they will be able to repay, particularly when they succeed in their programs of study.

The first study concluded that background factors, as measured by financial aid, socioeconomic and demographic variables, were relatively unimportant. The present study shows that the background variables are generally not as strongly related to default as the college success variables. However, it also demonstrates that a number of background variables, including the borrower's age, gender, and race/ethnicity, have significant relationships to default even after accounting for the influence of the college success variables. Several of the variables, such as Highest Educational Level Attained by the Student's Mother and Gender of Borrower, are significant in the new study despite the assessment of the prior analysis that they have negligible relationships to default behavior.

- College grade point average is strongly related to whether or not borrowers default after leaving college. Borrowers who exit Texas A&M in College Station with a GPA of 2.0 or less have a likelihood of default that is at least 10 percentage points higher than borrowers who leave school with GPAs of more than 3.0.
- A borrower who graduates has a probability of default that is three percentage points lower than a borrower who does not graduate.

- Borrowers who attend the Liberal Arts College at Texas A&M University have a higher chance of defaulting on student loans than borrowers in other Colleges. Borrowers in the Engineering School have the lowest probability of default.
- Receipt of in-person exit counseling appears to be important, even after controlling for the influence of other variables, such as whether the borrower graduated. Borrowers who do **not** receive in-person exit counseling have a probability of default that is eight percentage points higher than borrowers who do receive it.
- Texas A&M borrowers who are at least 34 years of age at the time of entering repayment on student loans have a probability of default that is 14 percentage points higher than borrowers who enter repayment when they are 21 or 22 years old.
- The highest academic level attained by a student's mother is related to whether or not a borrower defaults on student loans. Borrowers whose mothers reached the middle school level have a likelihood of default that is five percentage points higher than borrowers whose mothers attained to the college level. However, there is no significant difference in the probability of default for borrowers whose mothers went to high school or attended college.
- Being Black is associated with a higher likelihood of student loan default, even among borrowers who are alike in many other respects (such as graduation status, grade point average and adjusted gross income). This result is consistent with the finding of the original Texas A&M study. However, it is still unclear what set of socioeconomic and cultural factors "being Black" might be a proxy for.

Multivariate Analysis of Student Loan Defaulters at Texas A&M University

Introduction

In an effort to better understand student loan default behavior at Texas A&M University (TAMU), the research staff at TG, at the request of TAMU, conducted a study of the relationship between loan default, on the one hand, and many student and borrower characteristics, on the other hand. The study examines the default behavior of 12,776 undergraduate borrowers who attended Texas A&M in College Station and who entered repayment on TG-guaranteed Federal Family Education Loan Program (FFELP) loans during federal fiscal years 1997, 1998, and 1999. The study regards a borrower as being in default if the borrower defaulted within the fiscal year the borrower entered repayment or within the following fiscal year. Texas A&M staff supplied information describing high school coursework, SAT scores, college grade point average (GPA), length of attendance at TAMU, graduation status, amounts of financial aid received, financial need assessment, gender, marital status and many other aspects of students' backgrounds and college experiences. The study organizes these student and borrower characteristics into the following categories: Preparedness, Attendance Pattern, College Success, Financial Aid, Loan Briefing (Counseling), Demographics and Loan-Related Factors.

Whereas prior analysis of this sample of borrowers (see Bibliography) assessed each characteristic's separate relationship to default by determining whether different groups of borrowers have default rates that are significantly different from each other, the present study applies a more dynamic statistical procedure. This statistical technique is called multiple logistic regression. The advantage of logistic regression over the prior two-variable analysis is that it can reveal the separate relationship of each variable to default after accounting for the independent relationships to default of all the other relevant and available variables.

This analysis will attempt to increase our understanding of the factors that are related to default at Texas A&M University. Hopefully, TAMU, TG and others will be able to use this better understanding of default behavior to identify at-risk groups of borrowers, craft intervention or default mitigation strategies, and allocate scarce default aversion resources to the situations in which they will have the greatest chances of making a difference.

Prior Research on the Factors Relating to Student Loan Default

The genesis of early default studies was the need to comment on the policy of holding schools responsible for borrower defaults. Therefore, many prior studies have concerned themselves with evaluating the relative importance of borrower and institutional characteristics. Several have found that institutional characteristics have little or no association to loan repayment behavior and that borrower variables are much more important predictors of default (Knapp & Seaks, 1990; Volkwein & Szelest, 1995; Volkwein et. al., 1995; Wilms, Moore & Bolus, 1987). Since the present analysis of Texas A&M concerns the default behavior of students at one institution, prior work on the influence of institutional characteristics is of little relevance.

Nevertheless, in their endeavor to find the factors related to default, researchers have evaluated many borrower characteristics that are relevant to the present study. These factors include demographic descriptors (such as ethnicity or race, gender, age and income), financial aid-related variables (like financial need and expected family contribution) and some high school-related variables (like ACT scores and whether the borrower has a high school diploma).

The most consistent finding of past studies is that borrowers who graduate (or who earn a degree or who do not withdraw) have a much lower probability of defaulting on their loans, as compared to borrowers who do not graduate (Dynarksi, 1994; Knapp & Seaks, 1990; Meyer, 1998; Podgursky et. al., 2000; Volkwein & Szelest, 1995; Volkwein et. al., 1995; Wilms, Moore & Bolus, 1987; Woo, 2002). For many of these studies, graduation status was the single most important variable.

The second most prominent finding of multivariate default studies has been that ethnicity/race is strongly related to default (Dynarksi, 1994; Knapp & Seaks, 1990; Podgursky et. al., 2000; Volkwein & Szelest, 1995; Volkwein et. al., 1995; Wilms, Moore & Bolus, 1987; Woo, 2002). In particular, being Black greatly increases the probability of default. In three of the studies (Volkwein & Szelest, 1995; Volkwein et. al., 1995 and Woo, 2002), being Black had the largest effect of all variables, and in the remainder of the cited studies, being Black was the second most influential factor.

Prior studies have tested only a few variables that measure the borrower's performance in college. Volkwein et. al. (1995) found that the borrower's GPA in college and whether the borrower was a science or technology major produced significant but relatively small decreases in the probability of default. They also determined that a variable signifying that the borrower was a transfer student did not have a significant relationship to default. A related study by Volkwein and Szelest (1995) uncovered similar results with respect to college GPA, majoring in science or technology, and transfer status. Woo (2002) found that attainment of a graduate or professional degree greatly reduces the chances of default. She further established that borrowers who attended more than one school were also less likely to default. (Woo noted that this variable partially reflects the fact that borrowers who go to graduate school frequently have attended more than one school.) Whether or not a borrower studied a business or computer curriculum did not have a significant association to default in Woo's study. Meyer found that as the academic level attained by a borrower increases, the probability of default decreases.

Previous research has determined that demographic characteristics other than ethnicity have significant, though mostly smaller, associations to default. After ethnicity, parental income appears to be the most commonly-tested demographic variable, and studies have found higher income levels to be associated with decreases in the probability of default (Dynarksi, 1994; Knapp & Seaks; Volkwein et. al., 1995; Wilms, Moore & Bolus, 1987; Woo, 2002). Gender is also routinely analyzed, and researchers usually conclude that being female is related to a substantial reduction in the likelihood of defaulting (Podgursky et. al., 2000; Volkwein et. al., 1995; Woo, 2002). Podgursky et. al., Woo and Meyer examined the age of the borrower and determined it to have a significant but small effect on default behavior, with increases in age related to higher probabilities of defaulting. In contrast, Knapp & Seaks could not detect a statistically significant relationship for either the gender or age of the borrower. Volkwein and Szelest (1995) also failed to uncover an association between gender and default behavior.

Among the other demographic variables that researchers have found to have significant relationships to default are the marital status of parents, (Knapp & Seaks, 1990), U.S. citizenship (Wilms, Moore & Bolus, 1987), the parents' educational level (Volkwein et. al., 1995), being Hispanic (Dynarksi, 1994; Woo, 2002), having dependents (Dynarksi, 1994; Volkwein & Szelest, 1995; Volkwein et. al., 1995; Woo, 2002), the marital status of the borrower (Dynarksi, 1994; Volkwein & Szelest, 1995; Volkwein et. al., 1995), the borrower's income (Dynarksi, 1994; Volkwein & Szelest, 1995; Volkwein et. al., 1995; Woo, 2002) and several others.

To a very limited extent, researchers have evaluated characteristics reflecting the borrower's experience before college. Several studies have found that graduation from high school reduces the likelihood of default (Dynarksi, 1994; Volkwein et. al., 1995; Wilms, Moore & Bolus, 1987 and Woo, 2002). However, Volkwein and Szelest did not detect a significant relationship between having a high school diploma and default behavior. Podgursky et. al. also examined ACT scores and identified a small negative effect on default.

Studies have generally paid scant attention to financial aid-related variables. Nevertheless, it is important to test whether financial assistance mitigates the probability of default in ways that are independent of income. Among the studies reviewed here, only a couple reviewed variables other than family income and family assets. Volkwein et. al. tested several financial aid-related variables – such as the receipt of scholarships/grants, whether the borrower participated in work study and whether the borrower had other employment – but found none of them to be significant. Meyer, however, determined that the probability of default declined with increases in the cost of attendance, controlling for type of institution. He further discovered that the likelihood of default increased substantially for borrowers who received more than \$1,000 from non-loan aid sources. He noted a small decrease in the chances of defaulting as the expected family contribution of borrowers increased.

Several of the studies have also included loan-related variables. Four of the analyses determined that there was not a statistically significant relationship between the amount of loans borrowed and default behavior (Knapp & Seaks, 1990; Volkwein & Szelest, 1995; Volkwein et. al., 1995; Woo, 2002). Meyer, however, found that each \$1,000 of total debt increases the probability of default by about one percentage point. And Dynarski determined that the probability of default rose with increases in the size of borrowers' monthly loan payments. Furthermore, Woo detected a small increase in the likelihood of default associated with an increase in the number of loans a borrower has. Meyer also examined the types of federal loans that borrowers received and showed that borrowers with only subsidized Stafford loans had the highest probability of default. In his study, he further demonstrated that borrowers who utilized deferments had a somewhat smaller chance of defaulting.

Compared to past studies, the present study will evaluate a far greater number and variety of variables. It will more thoroughly examine how variables that describe the performance of borrowers in college relate to the probability of default. It will look at a large number of performance characteristics, ranging from variables that measure success in college, like the number of course hours passed, to factors that describe the pattern of the borrower's attendance at TAMU, such as how many semesters the borrower attended and how many times the borrower withdrew from school. Furthermore, the study will expand the number and type of variables that depict the borrower's experience in high school, such as the number of course hours the student took in various subject areas. In addition, this analysis will consider a much larger number of

variables that portray the financial aid process, including financial need, expected family contribution and aid amounts.

Prior Research on the Factors Related to Student Loan Default at Texas A&M

In January 2003, TG published a study that analyzed the factors related to default at Texas A&M. Unlike the present multivariate study, which simultaneously examines the relationship of many variables to default, the prior bivariate study evaluated the association of each variable to default without taking into account how the other variables related to default. Though the prior TAMU study involved a less sophisticated analytical technique, its results provided guidance in the selection and construction of variables in the present analysis.

The prior and current TAMU studies share a common foundation. Both studies examine the default behavior of 12,776 undergraduate borrowers who attended Texas A&M in College Station and who entered repayment on TG-guaranteed Federal Family Education Loan Program loans during federal fiscal years 1997, 1998, and 1999. Both studies consider the same large set of student and borrower characteristics, which were organized into the following categories: Preparedness, Attendance Pattern, College Success, Financial Aid, Loan Briefing (Counseling), Demographics and Loan-Related Factors. Appendix A contains a complete list of the variables examined in the two studies.

The prior TAMU research found that background variables, which generally describe factors that the borrower cannot change once the borrower gets to college, were much less important to default behavior at Texas A&M University than performance variables, which describe what the borrower actually does while in college. The study found that background variables in the College Preparedness section (which describes borrowers' experiences in high school), the Demographic section and the Financial Aid section have weak relationships to default. In contrast, the performance variables in the College Success section and Attendance Pattern section represent most of the borrower characteristics that have strong associations to default. *In particular, college grade point average has the strongest association to default of any variable in the study*, and ethnicity has a relatively weak relationship to default. The study concluded that borrowers are succeeding in the student loan programs largely without respect to their ethnicity, their parents' educations or their family income. Moreover, the study suggested that default behavior hinges more strongly on factors that are at least partially under the borrower's control: whether the borrower graduates, how long the borrower spends in college and how well the borrower does in his or her college coursework. For the full report, go to http://www.tgslc.org/pdf/TAMU_Default_Study.pdf.

The report findings prompted the financial aid office at Texas A&M to enhance some loan-related procedures. The finding that lower grade point averages are associated with higher likelihoods of default prompted aid administrators to strengthen the enforcement of their policy on Satisfactory Academic Progress (SAP). Borrowers who fail to complete enough of their coursework to meet the SAP standard must come to the financial aid office to appeal for eligibility for future financial aid. This in-person requirement creates an opportunity for financial aid workers to "counsel" the borrower. TAMU aid officers also shored up their procedures

related to students who fail all of their courses in a semester. TAMU decided to phone each of the borrowers on its “All-Fail” list in order to offer help.

Methodology for Multivariate Analysis of Defaulters at TAMU

TG uses logistic regression for conducting multivariate analyses of behaviors, such as repayment behavior, in which outcomes can assume one of two classes, like defaulting or not defaulting. The statistical analysis proceeds by determining the relationships between borrower characteristics and default behavior within a past population of borrowers. The known outcomes (i.e., default behaviors) of this population serve as the basis for statistical estimation. The result of the analysis is a set of coefficients or weights. The logistic regression method chooses the set of weights that would produce predictions of default that match as closely as possible to the known outcomes of default. The sign (plus or minus) of a coefficient indicates whether the presence of the characteristic increases or decreases the likelihood of default, and the size of a coefficient generally reflects the strength of the relationship between the characteristic and the occurrence of default. This approach makes it possible to make statements about the analyzed behavior that possess an intuitive appeal, such as “borrowers who land a job within their field of study within a month after graduating have a probability of defaulting that is 15 percentage points lower than other borrowers.” (Note: this is not a real finding and is provided for demonstration purposes only.)

Variable Selection Process

One goal of this analysis is to find, among all possible relevant variables, the subset of variables that best explains default behavior. This subset of variables is likely to be much smaller in number than the total number of variables that were gathered for the study – about 75 variables in all. (For a complete list of the variables examined in the two studies, refer to Appendix A.) In fact, statistical analysis in both the earlier bivariate analysis and in the current multivariate analysis showed that many variables explain very little about the likelihood of borrowers to default. Therefore, variables that have no statistically significant relationship to default were therefore dropped from inclusion in a final default model. In addition, some groups of variables tend to provide similar explanations of default behavior and are, therefore, redundant with each other; in many such cases it is possible to select one variable to represent the other variables. However, sometimes a variable is so important from a theoretical or practical standpoint that the modeler must include it, even if it overlaps with other variables. Incorporating all of these considerations, the final default model is the combined result of statistical relevance, theoretical importance, organizational requirements and human judgment.

Results of the Multivariate Analysis

The multivariate analysis produced a default model containing the variables listed in the table below. The table lists each variable, its reference group, the coefficient, and the delta-p (change in probability), each of which will be explained below. Variables are listed in alphabetical order.

The multivariate estimation process produces a coefficient for each variable. The sign (positive or negative) of a coefficient indicates whether the presence of the variable increases or decreases the likelihood of default, and the size of a coefficient generally reflects the strength of the relationship between the variable and the occurrence of default. For example, any grade point average (GPA) below 3.01 is associated with an increase in a borrower's chances of defaulting (since the coefficients are all positive). Moreover, as GPA decreases, the probability of default increases (since the coefficients are larger for lower GPA categories). In contrast, failing six or fewer hours is associated with a decrease in a borrower's chances of defaulting (since the coefficients for these categories are all negative).

The presence of asterisks next to a coefficient indicates that the variable has a statistically significant relationship to default behavior. Statistical significance means that there is a relatively high confidence that a relationship really exists – that the size of the coefficient did not result from the peculiarities of the sample that we analyzed. The more asterisks there are, the higher the level of confidence that a true relationship exists between a variable and default behavior. Thus, whereas all the variables based upon college GPA have a statistically significant relationship to default, the variable indicating that a borrower failed “4 to 6” course hours does not have a statistically significant relationship.

Unfortunately, the coefficients are difficult to interpret in their raw form. In order to more easily understand their meaning it is necessary to convert them to another form. The delta-p represents the percentage point change in the probability of default given the presence of a characteristic. For example, a borrower who graduated has a probability of default that is two percentage points lower than a borrower who did not graduate. This means that if all other variables in the model besides the graduation variable indicated that a borrower had a 4.7 percent probability of defaulting, the fact that the borrower graduated would lower the likelihood of default to 1.7 percent (4.7 percent minus 3.0 percent). The default rate of 4.7 percent is the overall average for the study sample, and the delta-p statistics in the following table always reflect a change relative to this default rate. Delta-p statistics are only provided for variables that are statistically significant.

A reference group is required for interpreting the variables used in this analysis. Consider the group of age variables. Borrowers who entered repayment at 34 or more years of age have a likelihood of default that is 13 percentage points higher than borrowers between the ages of 21 and 22 years, who belong to the reference group. Similarly, borrowers who entered repayment between the ages of 27 and 30 years have a likelihood of default that is three percentage points higher than borrowers between the ages of 21 and 22 years (the reference group). In most cases, the analysis would produce an equivalent result regardless of which category served as the reference group. However, in some cases, the desire to test a prior hypothesis has suggested a useful reference category. For example, the belief that students of traditional college-graduating age (21 to 22 years) differ in repayment behavior from other students might lead a researcher to select the 21 to 22 year category as the reference group.

Results of Multivariate Analysis Texas A&M - College Station

Variable Group	Variable	Reference Group	Coefficient	Delta-p (Change in probability)
Intercept			-3.57 ***	
Adjusted Gross Income (AGI) of Student	Zero or less	\$1 - \$2,000	0.27	
	Zero	\$1 - \$2,000	0.25	
	\$2,001 - \$4,000	\$1 - \$2,000	-0.17	
	\$4,001 - \$6,000	\$1 - \$2,000	0.00	
	\$6,001 - \$12,000	\$1 - \$2,000	-0.19	
	\$12,001 or more	\$1 - \$2,000	-0.64 **	-2%
Age of Borrower at Time of Entering Repayment (in years)	17 - 20	21 - 22	-0.11	
	23 - 26	21 - 22	0.06	
	27 - 30	21 - 22	0.54 **	3%
	31 - 34	21 - 22	0.51	
	35 or more	21 - 22	1.51 ***	14%
College or School Last Attended	Engineering	Liberal Arts	-0.83 ***	-3%
	Other	Liberal Arts	-0.40 ***	-2%
Expected Family Contribution (EFC)	Zero	\$2,001 - \$3,000	0.46 **	3%
	\$1 - \$500	\$2,001 - \$3,000	0.43 *	2%
	\$501 - \$1,000	\$2,001 - \$3,000	-0.11	
	\$1,001 - \$2,000	\$2,001 - \$3,000	0.39 *	2%
	\$3,001 - \$5,000	\$2,001 - \$3,000	0.09	
	\$5,001 - \$7,000	\$2,001 - \$3,000	-0.07	
	\$7,001 - \$10,000	\$2,001 - \$3,000	-0.19	
	\$10,001 or more	\$2,001 - \$3,000	-0.48 *	-2%
Exit Counseling (In-Person)	Did not receive	Did receive	1.12 ***	8%
Gender of Borrower	Female	Male	-0.38 ***	-1%
Grade Point Average (GPA)	0	3.01 - 4.00	1.72 ***	17%
	0.01 - 1.00	3.01 - 4.00	1.54 ***	14%
	1.01 - 2.00	3.01 - 4.00	1.22 ***	10%
	2.01 - 2.50	3.01 - 4.00	1.11 ***	8%
	2.51 - 3.00	3.01 - 4.00	0.80 ***	5%
Graduation Indicator	Graduated	Did not graduate	-0.82 ***	-3%
Highest Educational Level Attained by Student's Mother	High school	College or beyond	0.00	
	Middle school	College or beyond	0.74 ***	5%
	Missing	College or beyond	0.00	
	Unknown	College or beyond	0.25	
Number of Hours Failed	Zero	7 to 9	-0.73 ***	-2%
	1 to 3	7 to 9	-0.33	
	4 to 6	7 to 9	-0.09	
	10_or_more	7 to 9	0.35 *	2%

Results of Multivariate Analysis Texas A&M - College Station

Variable Group	Variable	Reference Group	Coefficient	Delta-p (Change in probability)
Number of Hours Transferred	0	1 to 11	0.03	
	12 to 23	1 to 11	-0.34 *	-1%
	24 to 39	1 to 11	-0.25	
	40 to 57	1 to 11	-0.34 *	-1%
	58 to 93	1 to 11	-0.18	
	94 or more	1 to 11	-0.98 *	-3%
Race and Ethnicity	Black	White	0.60 ***	4%
	Hispanic	White	0.04	
	Other	White	0.26	

Sample size: 12,776

Defaulters: 600 (4.7 percent)

Chi-square (Likelihood Ratio) 1197.56 with 46 degrees of freedom (Pr > Chisq < .0001)

C statistic: 86.9 percent

*Statistically significant at the 0.05 level

**Statistically significant at the 0.01 level

***Statistically significant at the 0.001 level

No single statistic – whether the coefficient, the level of significance or the delta-p (change in probability) – provides an unambiguous way of ranking variables in terms of how adequately they explain default behavior. Each statistic in the table above has its drawbacks in depicting the strength of the relationship between these variables and whether or not borrowers default. As a consequence, the subsequent discussion of the variables relies upon a composite picture of the various statistics. In general, groups of variables with large coefficients, whether positive or negative, and high levels of significance (more asterisks) have stronger relationships to default behavior; groups of variables with smaller coefficients and low levels of significance have weaker associations to whether or not borrowers default. The variable groups are discussed in roughly the order of their strength of association to default.

Grade Point Average (GPA)

The lower the borrower's grade point average is, the more likely the borrower is to default. For borrowers with grade point averages of 2.00 or less, the probability of default is at least 10 percentage points higher than it is for borrowers whose GPAs exceed 3.00 (i.e., for the reference group). Each of the categories of grade point below 3.01 has a likelihood of default that is greater than the reference group. The coefficients for the variables in this group are relatively large and each variable has a highly significant relationship to default behavior at Texas A&M. This group of variables has significant practical importance because financial aid workers could easily monitor grade point average on an ongoing basis.

This finding reinforces the finding of the previous, bivariate study of TAMU defaulters. Borrowers who are successful students carry that success over into the repayment of their loans.

This analysis does not reveal the cause of the association between grade point average and default. It is possible that qualities of character – like conscientiousness, persistence, motivation, intelligence and discipline – contribute both to the borrower's success in college and to the borrower's propensity to repay his or her loans after college. Alternatively, borrowers with high GPAs might fetch higher-paying jobs in the labor market, thereby making it relatively easier for them to repay their student loans.

Exit Counseling (In-person)

The borrower who does not receive in-person exit counseling has a chance of defaulting that is eight percentage points higher than the borrower who does receive in-person exit counseling, assuming the two borrowers share the same characteristics based upon all the other variables. This finding suggests, but does not prove, that provision of in-person exit counseling might help prevent student loan default. However, it is at least theoretically possible that a factor that was not included in the analysis (an external factor) causes some borrowers both to default and to evade their in-person exit counseling requirements. Since enforcing in-person exit counseling upon these borrowers would not necessarily eliminate this external factor, these borrowers might have a higher likelihood of default even if they receive the counseling. Like the GPA variable, the exit counseling variable might have considerable practical implications, since whether or not borrowers receive counseling might be partially within the influence of the financial aid office.

Graduation Indicator

Whether or not a borrower graduates has a highly significant relationship to whether or not a borrower defaults. A borrower who graduates has a chance of defaulting that is three percentage points lower than a borrower who does not graduate, all other things being equal. While this might seem like a modest decrease in the probability of defaulting, consider that the average default rate for the study is already fairly low at 4.7 percent. Thus, for a borrower who otherwise has an average likelihood of defaulting, graduating reduces his or her probability of default by more than 63 percent (3 percent divided by 4.7 percent). Though financial aid officers might have little direct impact on whether or not borrowers graduate, this variable might assist them in identifying at-risk borrowers (i.e., the ones who do not graduate) to whom they can direct default aversion efforts.

College or School Last Attended

Borrowers in the Liberal Arts College have a higher probability of default than borrowers in other colleges, holding all other factors constant. Engineering students have a likelihood of default that is three percentage points lower than borrowers in the Liberal Arts College. All other borrowers in colleges outside of Liberal Arts have a probability of default that is two percentage points lower than borrowers in the Liberal Arts College. The Other category combines all other colleges and schools at Texas A&M University.

Age of Borrower at Time of Entering Repayment (in years)

In general, the results for this group of variables suggests that the older the borrower is at the time of entering repayment, the more likely the borrower is to default. In particular, borrowers who are 34 years of age or older have a probability of default that is 14 percentage points higher than borrowers who are 21 or 22 years old, all other factors being equal. These non-traditional students might have families and jobs that result in a weaker integration to the campus, and therefore a weaker sense of obligation to repay student loans. They also might have established, non-educational financial commitments, like home mortgages, that make it relatively difficult for them to manage their student loan debts.

Number of Hours Failed

It appears that the more hours a borrower fails, the higher the borrower's probability of default. (Note that one of the variables was not statistically significant.) In particular, a borrower who has not failed any course hours has a lower likelihood of default than borrowers who have failed some coursework but who otherwise share the same characteristics. Like the GPA variable and the Graduation Indicator, the Number of Hours Failed measures the success or failure of borrowers in college and suggests that educational achievement is related to repayment success.

Race/Ethnicity of Borrower

At Texas A&M, Black borrowers have a chance of defaulting that is four percentage points higher than White borrowers who otherwise possess the same characteristics (as indicated by the variables in the model). This result is statistically significant. None of the other ethnicity variables has a significant relationship to default behavior. However, while this result supports the general finding of past studies that being Black is associated with a higher likelihood of default, it indicates nowhere near the strength of relationship that prior studies revealed.

Race or ethnicity are unlikely to have any direct relationship to default behavior. Rather, it is possible that a racial category is a proxy or label for a set of socioeconomic conditions that make it more or less likely that one group will default relative to another group. This multivariate analysis shows that among borrowers who graduate, avoid failing any coursework, obtain GPAs above 3.00, share the same majors and receive in-person exit counseling, the Black borrowers still have a chance of defaulting that is four percentage points higher than White borrowers. Based upon their success in college, labor markets should value students from both racial groups similarly by bestowing approximately the same salaries upon them and by fostering their careers in similar ways. Based upon similar salaries and income growth, White and Black borrowers would have approximately the same ability to repay their student loans, all other things being equal. If labor markets do not assign the same value to Black and White borrowers with the same credentials, then the higher likelihood of Black borrowers to default might merely reflect their diminished capacity to repay as a result of the prejudicial treatment they receive within the employment market. If labor markets are fair, then differentials in repayment risk might still be explained by other socioeconomic differences related to race, such as by differences in wealth.

Highest Educational Level Attained by Student's Mother

Borrowers whose mothers reached no further than the middle school level are more likely to default – by five percentage points – than borrowers whose mothers attained to the college level, (assuming that the borrowers are alike in other respects). However, borrowers whose mothers attended school no further than high school were no more likely to default on student loans than borrowers whose mothers attended college.

Researchers include this variable – and the comparable variable for the father's education attainment – in analyses of college attendance, retention and completion because they believe that it reflects the student's "inherited" valuation of higher education. The theory is that students whose parents did not attend college are less likely to attend or complete college themselves because higher education is not a traditional value of their families. The present analysis might extend that logic: borrowers whose mothers do not attain to very high levels of education will be less likely to graduate and will therefore be less likely (or able) to repay their student loans. However, this analysis already accounts for the separate effect of graduation. Borrowers who do not graduate are indeed less likely to repay their student loans; but this is the case regardless of the educational attainment of their mothers. The mother's educational attainment therefore has an effect that extends beyond the borrower's valuation of education, as indicated by the borrower's graduation. One possibility is that borrowers whose mothers attain only to the middle school level have fewer parental resources to rely upon when they encounter repayment difficulties. (However, the statistical model also accounts for Expected Family Contribution, which would be a proxy for the borrower's parental resources.) An interesting point to keep in mind is that nearly three-fourths of the borrowers whose mothers attained only to the middle school level are Hispanic or Asian, allowing for the possibility that this variable is somehow related to the immigrant status of the borrowers' parents.

Expected Family Contribution (EFC)

In general, the higher the Expected Family Contribution of the borrower's family is, the lower the likelihood that the borrower will default. The borrowers who have zero EFCs have a probability of default that is three percentage points higher than borrowers with EFCs of \$2,001

to \$3,000. Similarly, borrowers with EFCs between \$1 and \$500 have a likelihood of default that is two percentage points higher than the reference group. At the other end of the spectrum, borrowers who have EFCs of more than \$10,000 have a likelihood of default that is two percentage points lower than the reference category. In general, theory suggests that higher amounts of expected family contribution are associated with higher family incomes. Though the income of a student's parents is not necessarily accessible to the student as a source for repaying student loans, students whose families have higher EFCs might have more financial resources available to them in times of repayment difficulties.

Number of Hours Transferred

In general, the more hours that a borrower transfers to Texas A&M, the lower the likelihood that the borrower will default, all other factors being equal. (Note, however, that some of the variables in this group are not statistically significant.) Borrowers who transfer at least 94 course hours have a probability of default that is three percentage points lower than borrowers who transfer between one and eleven hours. There are several possible explanations for this effect. "Hours Transferred" might merely be a partial proxy for grade level: the more hours transferred, the farther the student has progressed in college and the less likely the borrower is to default. Alternatively, students who transfer to TAMU have demonstrated some success in college; this group has been subjected to a process of education that has weeded out some of the students who started college at the same time but who failed to progress and transfer. In contrast, some of the students who enter TAMU as freshmen (with no transfer hours) will borrow, fail to progress academically, and leave school with low academic achievement. Therefore, there might be a tendency for the freshman admissions to contain a larger proportion of higher-risk borrowers than the transfer students.

Adjusted Gross Income (AGI) of Student (in dollars)

Borrowers with relatively high adjusted gross incomes – more than \$12,000 – have a likelihood of default that is two percentage points lower than borrowers who have AGIs between one dollar and \$2,000. None of the other variables in this group have a statistically significant relationship to default behavior. This is not especially surprising. The income of a borrower during college is not necessarily correlated with the income that the borrower will receive after leaving college. If college makes the difference it is supposed to make, students who enter college with different AGIs and who achieve the same academic success during college should garner approximately the same income after college and should therefore have similar capacities to manage their debt. The fact that a very high AGI lowers the probability of default could mean several things. For example, borrowers with high AGIs might be more likely to be able to save money during college that could help them repay loans later. Or such borrowers might be more likely to have sources of income -- such as rental income or investment earnings -- that persist after college and increase their ability to repay student loans.

Gender

Female borrowers are slightly less likely to default than male borrowers. Though the reasons for this finding are unclear, the small difference should cause little concern. It is worth noting that this finding agrees with prior research in this subject area (see the Prior Research section).

Model Performance

Based upon the characteristics of a borrower, it is possible to sum the coefficients for the variables in the prior section and to convert that sum to a probability that the borrower will default. The estimated probability can then be compared to the known outcome for the borrower. This comparison can be made for all borrowers in the study in order to gauge the performance of the multivariate model. In general, the performance measures in this section assess how well the statistical model correctly classifies defaulters and non-defaulters.

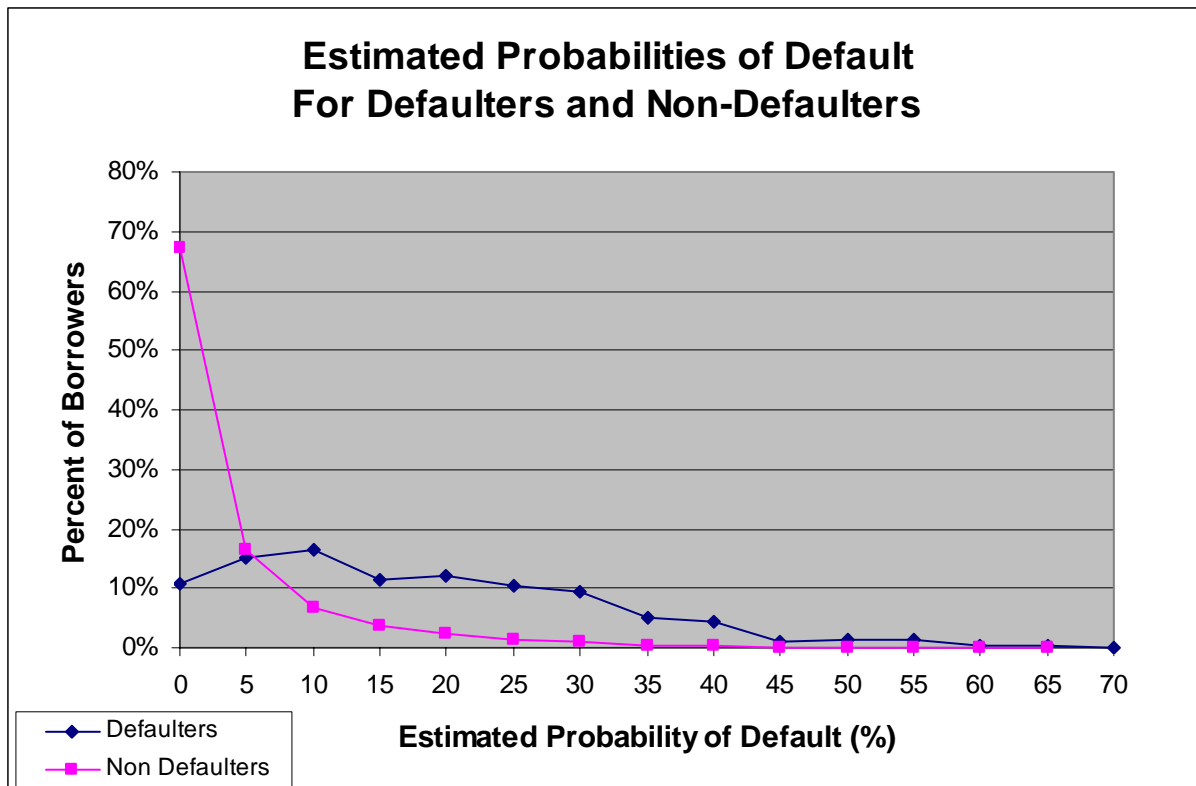
The performance measures indicate that this statistical model performs very well. It does an excellent job in assigning high probabilities of default to borrowers who actually defaulted and low likelihoods of default to borrowers who did not actually default.

The following pages outline the performance of the model using various methods of assessment, including:

Distribution of Probabilities	page 17
Kolmogorov-Smirnov (K-S) Statistic	page 18
C Statistic	page 19
Receiver Operating Characteristic (ROC) Curve	page 19
Classification Matrix and Misclassification Rate	page 20

Distribution of probabilities

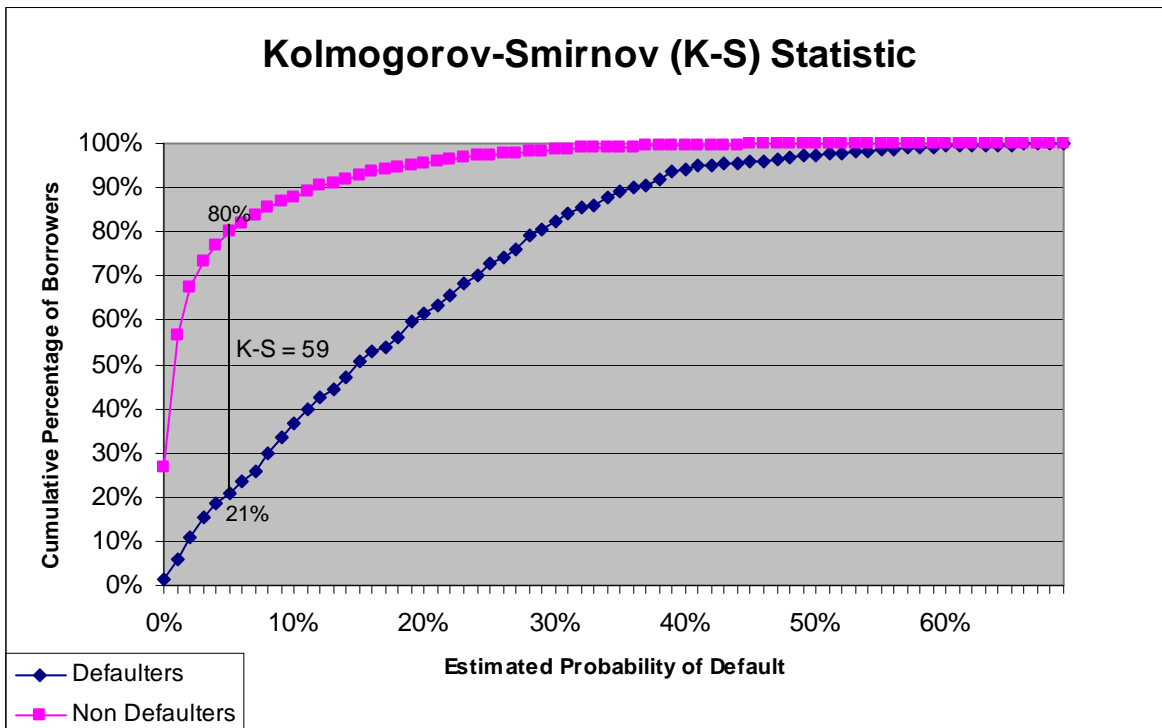
The following chart shows the default probabilities assigned by the multivariate model to borrowers in the study. The chart provides a separate distribution of probabilities for actual defaulters and actual non-defaulters. (Each borrower's estimated probability of default was rounded to the nearest five percent.) The vertical axis shows the percentage of borrowers who were assigned each probability. Thus, whereas the model assigned estimates of a zero percent (rounded) probability of default to 67 percent of actual non-defaulters, it assigned a zero percent (rounded) probability of default to only 11 percent of actual defaulters. In general, if the model is performing well, the curve for the non-defaulters should be higher than the curve for the defaulters on the left side of the chart. Similarly, the curve for the defaulters should be higher than the curve for the non-defaulters on the right side of the chart. The visual impression of this chart is that the model appears to have performed well.



Kolmogorov-Smirnov (K-S) Statistic

The previous distributions can be transformed into a set of cumulative distributions. Cumulative distributions give the percentage of borrowers who have an estimated probability that is equal to, or less than, a given point along the horizontal axis. For example, the chart below shows that 80 percent of actual non-defaulters have an estimated probability of default that is less than or equal to five percent and that only 21 percent of actual defaulters have an estimated probability of default in that range. As it turns out, at five percent (along the horizontal axis), the curves for defaulters and non-defaulters are separated by the greatest distance. This distance is known as the Kolmogorov-Smirnov (K-S) statistic. For the present model, the K-S statistic is 59 percent (80 percent minus 21 percent). Models with large K-S statistics are said to have done a good job of distinguishing between defaulters and non-defaulters. Fifty-nine percent is a very high K-S statistic and indicates that the model does well in separating defaulters and non-defaulters.

A high K-S means that a model will predict default outcomes for a much higher percentage of actual defaulters than non-defaulters. Suppose we predicted default for borrowers who the model assigned a default probability greater than five percent. The K-S of 59 percent indicates that using five percent as the prediction cutoff means that we will predict default 59 percent more frequently for defaulters than for non-defaulters. At five percent, the model would predict 79 percent of actual defaulters to default (that is, one minus the 21 percent with probabilities less than or equal to five percent), but it would only predict 20 percent of actual non-defaulters to default (one minus the 80 percent with probabilities less than or equal to five percent).

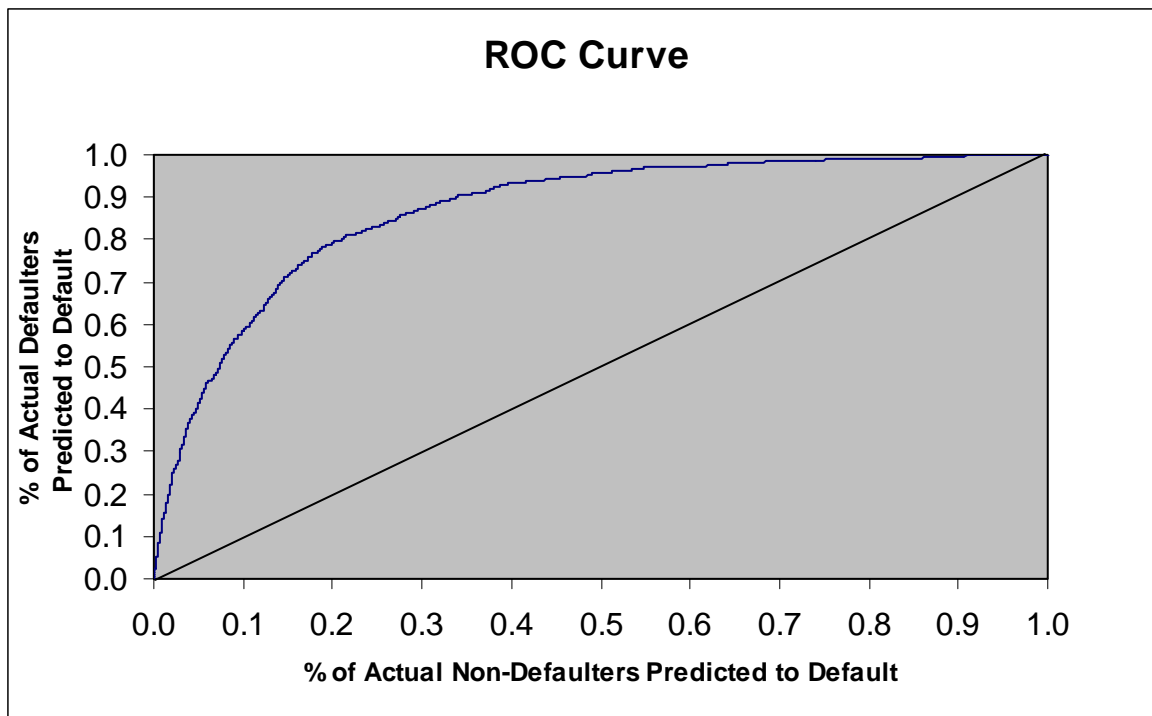


C Statistic

The c statistic measures how consistently a model assigns higher probabilities to actual defaulters than it does to actual non-defaulters. It compares each defaulter with each non-defaulter. In the present analysis, there are therefore 7,305,600 pairings (600 defaulters multiplied by 12,176 non-defaulters). The c statistic indicates the proportion of these cases for which the model assigns a higher probability of defaulting to the defaulter than it assigns to the non-defaulter. For the present model, the c statistic is 87 percent – an extremely high value for this statistic.

Receiver Operating Characteristic (ROC) Curve

The c statistic is represented graphically in the chart below. The area under the curve – called a Receiver Operating Characteristic (ROC) curve – is the c statistic: 87 percent of the chart is below the curve. A statistical model that assigned the same probabilities to defaulters and non-defaulters – a model that does no better than chance – would have an ROC curve that formed a diagonal running from the lower left corner of the chart to the upper right corner. To the extent that an ROC curve bows above the diagonal, the performance of the model increases. A model that perfectly separates defaulters and non-defaulters would have an ROC curve that hugged the left-hand side and top of the chart. The ROC curve for this model ranges well above a diagonal and indicates a high level of performance.



Classification Matrix and Misclassification Rate

Constructing a classification matrix provides an easy way to assess how well the statistical model classifies defaulters and non-defaulters. In the following example, the matrix employs a classification rule: if the model assigns a probability of default of five percent or more, the borrower is classified as a defaulter; a borrower with less than a five percent probability of default is predicted to be a non-defaulter. The matrix shows the numbers of actual defaulters that the classification rule predicts to be defaulters and non-defaulters. It also provides the same information for actual non-defaulters.

		Predicted Outcome	
	N=12,776	Default	Non Default
Actual Outcome	Default	484	116
	Non Default	2,615	9,561

It is possible to derive a misclassification rate from the classification matrix. When the predicted outcome does not align with the actual outcome, the classification rule resulted in a misclassification. The total number of misclassifications (2,731) is the sum of the defaulters who the model predicted to be non-defaulters (116) and the non-defaulters who the model predicted to be defaulters (2,615). The misclassification rate is 21 percent (2,731 divided by 12,776).

Whether or not this misclassification rate is good depends upon the frame of reference. If the school's alternative to using the model is to treat all borrowers as if they are potential defaulters, then a misclassification rate of 21 percent is very good. Treating all borrowers as potential defaulters will misclassify all 12,176 non-defaulters and result in a misclassification rate of 95 percent. In this comparison, use of the model produces a four-fold reduction in the misclassification rate. If the school's alternative to using the model is to provide counseling to borrowers who have a GPA of 2.5 or lower, the misclassification rate will be about 35 percent, since 105 defaulters in the study have GPAs greater than 2.5, and 4,189 non-defaulters have GPAs of 2.5 or less. Relative to this alternative, the model still provides a significant, though more modest, reduction in the misclassification rate.

Uses of the Findings and the Model

This multivariate analysis has a multitude of potential practical implications. Some findings might lead to obvious and easily applicable solutions that in some cases can take advantage of existing policies and procedures. Other findings might highlight the value of coordinating financial aid goals with the efforts of academic advising, career counseling and other campus functions. There are also sophisticated ways that a school could apply this statistical model to identify at-risk borrowers and intervene cost-effectively. And while some of the insights provided by this study might come as no surprise to many, they nonetheless provide evidentiary support to our intuitions and hopefully allow us to proceed with our solutions more confidently.

Texas A&M University (and perhaps other postsecondary institutions) can use the findings related to individual variables to adjust policies and procedures in the financial aid office. Texas A&M already modified procedures based upon the earlier bivariate analysis, which showed that the likelihood of default is associated with factors like college GPA and number of hours failed. Financial aid staffers began calling borrowers on the “All Fail List”, which the aid office had been receiving all along. (The All Fail List identifies borrowers who failed all of their courses within a semester.) Similarly, aid administrators decided to require borrowers to come to the financial aid office if they failed to meet standards for satisfactory academic progress (SAP), which is based upon course completion and is necessary for maintaining aid eligibility. This latter requirement creates an opportunity for aid workers to counsel students in-person.

The multivariate analysis offers additional ways that TAMU can use the findings related to individual variables in order to change existing procedures, introduce new procedures, or identify potential problem areas that require more research:

- Given the strong connection of grade point average to default behavior, TAMU could target at-risk borrowers by using a GPA cutoff to trigger an intervention. For example, the financial aid office could provide supplemental loan/debt counseling for borrowers the first-time they drop below a 2.5 GPA. Though it would be desirable to also provide such counseling when borrowers leave school with similarly low GPAs, many poor performing students leave college before the institution knows that they are gone. At any rate, a policy based upon a GPA threshold represents a possible improvement over one based upon a SAP policy. Borrowers who meet the SAP standard by passing their classes might thereby opt out of needed counseling, even though they performed very poorly in their coursework and are at risk of exiting the institution soon.
- The findings suggest that TAMU could improve its cohort default rate and prevent some defaults by bolstering the exit counseling it provides. The study shows that the in-person exit counseling is effective for those who receive it. However, borrowers who do not receive it face a relatively high probability of default. Bivariate statistics, which are not part of this study, show that in-person exit counseling makes a much bigger difference for borrowers who do not graduate than it does for borrowers who graduate. The challenge for TAMU would be to create a set of incentives that could entice non-completers to attend in-person exit counseling.
- Borrowers who enter repayment at 34 years of age or older have a much higher probability of default than younger borrowers. Policies that recognize the special circumstances of older borrowers, who sometimes face financial burdens and family

and career obligations not confronted by younger borrowers, might lower the chances of default for these borrowers.

- Borrowers who attend different colleges at the University have different likelihoods of default. As this might reflect the fact that borrowers with different college majors garner different wages in the labor market and consequently vary in their abilities to repay student loans, this finding lends further support to debt counseling practices that make a connection between borrowing and future salary expectations.

TAMU could also use the model, or a variant of it, to predict a more precise risk of default than can be obtained from any individual variable. As mentioned in the previous section, the multivariate model has a lower error in classifying defaulters than some more simple, one-variable models. (An example of a one-variable model is using college GPA to provide debt counseling for borrowers with GPAs of 2.5 or less.) Thus, TAMU could calculate an overall probability of default for individual borrowers by applying the coefficients (or scores) to the borrowers' characteristics, summing the results for all characteristics and transforming the sum to a probability through the use of a simple formula. TAMU could then use a probability threshold or cutoff to separate likely defaulters from likely non-defaulters. For example, the university might choose a cutoff of five percent. TAMU could classify borrowers with more than a five percent probability of default (according to the model) as having an unacceptably high risk of default; otherwise, it would classify the borrowers as likely non-defaulters. TAMU could then intervene in some way with the borrowers who were predicted to be defaulters. Although the authors did not create the model described by this paper specifically for the purpose of prediction, a few modifications could produce a viable prediction model.

Aside from the usefulness of the model and its components in targeting default aversion efforts, the model provides more general insights. For example, the importance of college success variables – such as college GPA, graduation of the borrower and number of course hours failed – appears to link borrower repayment behavior more broadly to campus goals that involve the responsibilities of college administrators outside of the financial aid office. A borrower who the university retains, who persists to graduation, who has a high GPA and who did not fail courses along the way has a much lower chance of defaulting than an otherwise similar borrower who did not exhibit these signs of success. While such success indicators might say as much about the character and strengths of the students who embody them, surely the university makes a difference in the quality of the students who leave its institution. To the extent that this is true, efforts that increase persistence, improve the quality of education and place students on rewarding career paths, worthy goals in themselves, will incidentally lower default rates. And because so many diverse factors influence whether these goals are achieved, no one campus office can be expected to carry the burden for their achievement. Greater success in these areas requires a systematic and coordinated effort that brings together several administrative and academic offices.

Appendix

Study Variables

The study examined the variables listed in the following subsections. The variables are grouped into seven categories: College Success, Attendance Pattern, Preparedness, Demographics, Financial Aid, Loan Briefing, and Loan-related. In a joint effort, representatives from Texas A&M and TG defined these categories and selected variables that had potential value in illuminating the default behavior of Texas A&M borrowers. Some of the variables that were originally selected were not included in the study because of data problems or because they were not applicable to undergraduate borrowers. Other variables in the list were not selected by the study team but are derived from data in the original database.

I. Default or Not

Whether or not borrowers default is the focus of this study. The study regards a borrower as being in default if the borrower defaulted within the fiscal year the borrower entered repayment or within the following fiscal year. The study will determine whether default rates vary between different groups of borrowers or for different characteristics of borrowers. The rest of the variables, as described below, will define the characteristics of borrowers whose default rates are to be compared. As an example, the study will check whether default rates change as SAT equivalency scores increase. If they do and statistical tests show that the relationship is significant, then the findings might provide a valuable insight or, at least, a confirmation of conventional wisdom.

II. College Success

The “College Success” variables measure the academic performance of borrowers at TAMU. Borrowers can differ in the quantity and market value of their educations and in whether they attain their degrees. Variables such as “Hours Passed” and “Number of Degrees” measure the quantity of education that borrowers experience. “College” and “Degree” might implicitly reflect the market value of the education students receive. In general, the authors expect that labor markets will place a higher value on greater quantities and certain programs of study and that they will prefer degree completers to non-completers. If so, borrowers who succeed in college should obtain higher-paying jobs with better career tracks and they should, therefore, be better positioned to repay their student loans. Furthermore, the personal qualities, like persistence and discipline, that contribute to a borrower’s success in college might also make the borrower more likely to repay his or her loans.

- **College of the Student’s Most Recent Major**
- **Type of Degree**
- **Graduation Indicator**
- **Highest Degree Attained**
- **Highest Level Attained**
- **Indicator of Minor**

- **Indicator of Secondary Major**
- **Minor**
- **Number of Changes in Major**
- **Number of Degrees**
- **Number of Hours Failed**
- **Number of Hours Incomplete**
- **Number of Hours Passed**
- **Number of Hours Q Dropped:** a grade of ‘Q’ is assigned when a student drops a course without penalty on or before the 50th day of classes in a fall or spring semester, the 15th day of a five-week summer term or the 35th day of a 10-week summer term.
- **Primary Major**
- **Secondary Major**
- **Undergraduate GPA**

III. Attendance Pattern

The “Attendance Pattern” variables describe the length and intensity of a borrower’s attendance at Texas A&M. Some of the variables also indicate whether there were interruptions in the borrower’s course of study and whether the borrower was a transfer student. As a group, the variables are intended to signify the borrower’s commitment to the education he or she is pursuing. The study’s authors anticipate that borrowers who finish their programs of study, finish sooner rather than later, and finish with few interruptions will default with less frequency than other groups of borrowers.

Admission Code: whether the student was admitted automatically on the basis of class rank and SAT score (i.e., “admitted by academics”), was admitted after additional review of the student’s admissions application and high school transcript (i.e., “admitted by review”), was accepted on a provisional basis (i.e., “accepted on provisional”), in which full freshman admission in the fall is conditional upon making satisfactory progress in two summer semesters, or was accepted through some other means (i.e., “Other”).

- **Admission Major**
- **College of Admittance:** college (within TAMU) to which the student was admitted
- **Highest Number of Semester Hours**
- **Number of Hours Transferred**
- **Lowest Number of Semester Hours**
- **Number of Semesters Enrolled Before Departure**
- **Number of Semesters Enrolled Less than Full Time**
- **Number of Semesters in a Dorm**
- **Number of Summer Semesters Attended**
- **Number of Withdrawals**
- **Number of Years Between First Attendance and Most Recent Departure**
- **Previous College:** college that the borrower attended prior to attending TAMU
- **Total Hours – TAMU Hours plus Transfer Hours**
- **Total Number of Hours Taken at TAMU**

- **Type of Admission:** whether the student was admitted as a freshman, a transfer, an international student or a readmission
- **Withdrawal Indicator:** indicates whether the borrower ever withdrew from TAMU
- **Withdrawal Type:** the reason for the withdrawal

IV. Preparedness

In general, “Preparedness” variables reflect student performance before entering the university. They indicate the extent to which the student is ready for university programs of study. Since other research shows that college success/completion increases the likelihood that borrowers will repay student loans (see Prior Studies section) and since high school preparedness probably increases college success, we can expect that good performance before entering the university is related to lower default rates.

- **Number of Advanced Placement Credits**
- **Number of High School Advanced Math Credits**
- **Number of High School Algebra I Credits**
- **Number of High School Algebra II Credits**
- **Number of High School Biology Credits**
- **Number of High School Chemistry Credits**
- **High School Class Rank Percentile**
- **Number of High School Computer Science Credits**
- **Number of High School English Credits**
- **Number of High School Foreign Language Credits**
- **Number of High School Geometry Credits**
- **Number of High School Other Science Credits**
- **Number of High School Physics Credits**
- **SAT Equivalency Score:** SAT score or the conversion of an ACT score to an equivalent SAT scale.

V. Demographics

The demographic characteristics of borrowers might be related to default behavior in various ways. Perhaps degree attainment by the borrower’s mother or father is related to success in repaying student loans; this might occur as a result of the borrower internalizing the parent’s educational success as a model for responsibility and hard work and bringing those values to bear in many aspects of his or her life. Marital status and family size might be proxies for the amount of financial resources that are available to repay loans; married students or dependent students from large families might have fewer resources available to them from family income and therefore might be more likely to default.

- **Age of Borrower:** borrower’s age at the time of entering repayment
- **Gender of Borrower**
- **Ethnicity of Borrower**
- **Marital Status of Borrower**

- **Citizenship of Borrower**
- **Residency Status**
- **Country of Permanent Address**
- **State of Permanent Address**
- **Country of Local Address**
- **State of Local Address**
- **Highest Degree Father**
- **Highest Degree Mother**
- **Parental Marital Status**
- **Parental Family Size**

VI. Financial Aid

“Financial Aid” variables measure the monetary resources available to students from their own income, from parental income and from student financial aid programs. While it is likely that high-need/low-income borrowers will have higher probabilities of defaulting, there are other possibilities. The receipt of financial assistance might decrease the risk of default in groups that would otherwise be at high risk of default. It is also possible that high-income, low-need borrowers lack the commitment to their loan obligations that comes with having a high need for them, and so they default at rates comparable to other borrowers.

- **Adjusted Gross Income of Parents**
- **Adjusted Gross Income of Student**
- **Amount of Need**
- **Dependency Status**
- **Expected Family Contribution**
- **Total Family Contribution:** The Expected Family Contribution (EFC) minus certain financial aid amounts that cover the borrower’s EFC.
- **Total Loan Aid**
- **Total Other Aid**
- **Total Work Study Aid**

VII. Loan Briefing Variables

If debt counseling and exit counseling are operating effectively, they should lower the rate at which borrowers default.

- **Debt Counseling**
- **Exit Counseling (In-person)**

VIII. Loan-related Variables

The “Loan-related Variables” originate from TG’s databases. They represent basic measures and indicators of the borrower’s student loan experience that might have a relationship to default behavior. In theory, the “Number of Loans” and the “Total Loan Amount” could indicate the

repayment burden that a borrower faces: the higher the burden, the greater the likelihood of default. Alternatively, those variables could simply be a proxy for how long the borrower went to school: generally speaking, the higher the loan amount, the more education the borrower received and, therefore, the less likely the borrower is to default. Participation in the consolidation loan program could signify that the borrower has a knowledge and experience of the tools that will help keep a person out of default.

- **Number of Lenders**
- **Number of Loans**
- **Consolidation Loan:** whether or not the borrower has a consolidation loan
- **Total TG Loan Amount**

Bibliography

- Dynarski, M. Who defaults on student loans? Findings from the *National Postsecondary Student Aid Study*. *Economics of Education Review*, 13(1), 55-68.
- Knapp, L.G., and Seaks, T.G. (1992). An analysis of the probability of default on federally guaranteed student loans. *Review of Economics and Statistics*, 74(3), 404-411.
- Meyer, D. (1998). Predicting which borrowers are most likely to default. In: Webster, J., Meyer, D. & Arnold, A. *Student loan defaults in Texas: Yesterday, today, and tomorrow*. Austin, TX: Texas Guaranteed Student Loan Corporation.
- Monteverde, K. (2000). Managing student loan default risk: Evidence from a privately-guaranteed portfolio. *Research in Higher Education*, 41(3), 331-352.
- Podgursky, M., Ehlert, M., Monroe, R. and Watson, D. (2000). Student loan defaults and enrollment persistence. Columbia, MO. October.
- Steiner, M. and Teszler, N (2003). The Characteristics Associated with Student Loan Default at Texas A&M University. Austin, TX. January.
- Volkwein, J. F., Cabrera, A. F., Szelest, B. P., and Napierski, M. (1995). *Characteristics of student loan defaulters among different racial and ethnic groups*. Boston, MA: AIR 1995 Annual Forum Paper. (ERIC Document Reproduction Service No. ED 386 972).
- Volkwein, J. F., and Szelest, B. P. (1995). Individual and campus characteristics associated with student loan default. *Research in Higher Education*, 36(1), 41-72.
- Wilms, W. W., Moore, R. W., and Bolus, R. E. (1987). Whose fault is default? A study of the impact of student characteristics and institutional practices on guaranteed student loan default rates in California. *Educational Evaluation and Policy Analysis*, 9(1), 41-54.
- Woo, J. H. (2002). Factors affecting the probability of default: Student loans in California. *Journal of Student Financial Aid*, 32(2), 5-23.