May 2005

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

Conducted by TG Research
and Analytical Services

Sandra Barone
Matt Steiner
Natali Teszler

TG™

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

TG Research and Analytical Services
Sandra Barone
Matt Steiner
Natali Teszler

Table of Contents

Executive Summary and Highlights 3
Introduction 5
Prior Research on the Factors Relating to Student Loan Default 6
Methodology for Multivariate Analysis of Defaulters at Texas A&M - Kingsville 8
Variable Selection Process 8
Results of the Multivariate Analysis 9
TABLE 1: Results of Multivariate Analysis Texas A&M - Kingsville 11
Grade Point Average (GPA) 12
Exit Counseling 12
Graduation Indicator 13
Highest Grade at Which Borrower Received a Loan 13
Race/Ethnicity of Borrower 13
Expected Family Contribution (EFC) 14
High School Class Rank Percentile 15
Gender 15
Age of Borrower at Time of Entering Repayment (in years) 15
Dependency Status 15
Model Performance 15
Distribution of probabilities 16
K-S Statistic 16
C Statistic 17
ROC Curve 18
Classification Matrix and Misclassification Rate 18
Uses of the Findings and the Model 20
Appendix A: Sample Definition and Variable Descriptions 22
Table A1: Characteristics of Undergraduate Borrowers at TAMUK By Default Status 26
Appendix B: Standard Errors and Confidence Intervals 32
Bibliography 33

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

TG Research and Analytical Services
Sandra Barone
Matt Steiner
Natali Teszler

Executive Summary and Highlights

This study examines the default behavior of 5,177 undergraduate student borrowers who attended Texas A&M University – Kingsville (TAMUK) and entered repayment of their TG-guaranteed Federal Family Education Loan Program (FFELP) loans between October 1, 1998 and September 30, 2002 (fiscal years 1999 – 2002). Using the Department of Education's official cohort default rate formula, 569 borrowers, or 11 percent, were in default. The study uses a statistical technique called multiple logistic regression to analyze the effects of individual student and family characteristics on the probability of default, while controlling for the effects of the other variables in the analysis.

For students at TAMUK, persisting in college to graduation, performing well while at the university, and receiving exit counseling are all extremely important in decreasing a student's likelihood of default. These results suggest that expanding campus-wide efforts to increase retention rates and improve academic performance may be the most effective default aversion strategy for TAMUK.

The study finds that a student's background characteristics, such as race, gender and family income also have significant relationships to a student's probability of default, even after accounting for the college success variables. Although these relationships are not as strong as the relationships between success in college and default, they still persist. This finding highlights the fact that many students attending TAMUK are first-generation college students and have little or no financial and family support. These students may require additional assistance in order to successfully transition to college life and remain in school when faced with many conflicting priorities.

More specifically, the key findings of this study are:

- College grade point average (GPA) is strongly related to whether or not a borrower defaults on his or her student loan after leaving college. Borrowers who leave TAMUK with a GPA of 3.0 or higher have a likelihood of default which is at least 10 percentage points lower than those who exit with a GPA of 2.5 or less, holding all other borrower characteristics constant.

- Borrowers who received exit counseling have a probability of default which is nine percentage points lower than those who did not receive exit counseling.

- Graduating from TAMUK reduces a borrower's likelihood of defaulting by six percentage points, as compared to borrowers who did not graduate.

- Borrowers who persist in school, as measured by the grade level at which they obtained their last loan, have a lower probability of default than borrowers who leave school after their freshman year.

- Students at TAMUK who are African-American or Hispanic have a higher likelihood of student loan default than Caucasian students, even after controlling for the student's performance in college and family income.

- Students who graduated in the top 20 percent of their high school class have a probability of default which is five percentage points lower than those who graduated in the middle of their high school class.

- Female students at TAMUK are less likely to default on their student loans than male students.

Multivariate Analysis of Student Loan Defaulters at Texas A&M University - Kingsville[1]

TG Research and Analytical Services
Sandra Barone
Matt Steiner
Natali Teszler

Introduction

The Federal Family Education Loan Program (FFELP) makes it possible for millions of students to obtain post-secondary education each year who would otherwise be unable to afford to attend. In fiscal year 2000 alone, students borrowed over $25 billion through this program. Historically, however, default rates have been high under this program. In fiscal year 2000, defaulted student loans cost taxpayers over $2 billion. These defaulted student loans also hurt students' credit ratings. Because of the high costs of student loan defaults, both to the student borrower and to the taxpayers, the Department of Education sanctions schools with exceptionally high percentages of defaulters. In an effort to better understand which students are likely to default, and ultimately to design programs to reduce the number of borrowers who do default, TG and Texas A&M University - Kingsville (TAMUK) have agreed to work together to perform an analysis of student loan defaulters at TAMUK. Information obtained from this study can be used by both parties, as well as lenders and servicers, to target at-risk borrowers and lower default rates.

This study examines the behavior of 5,177 undergraduate students who attended TAMUK and entered repayment of their student loans between October 1, 1998 and September 30, 2002 (fiscal years 1999 – 2002). The Department of Education's official cohort default rate formula was used to determine if a borrower was in default. According to this formula, a borrower is considered to be in default if he or she defaults during the fiscal year that the borrower entered repayment or within the following fiscal year. Using this definition, 569 borrowers, or 11 percent, defaulted.

TAMUK provided detailed data on this sample, including information on the students' high school performance, financial need, college coursework and performance, and demographic information. This study closely follows the methodology used in a similar study performed by TG in conjunction with Texas A&M University at College Station. These studies use a statistical technique called multiple logistic regression to analyze the effects of individual student and family characteristics on the probability of default, while controlling for the effects of the other variables in the analysis.

[1] The authors would like to acknowledge the financial support of the U.S. Department of Education. We would like to thank Jeff Webster and Carmen Tym for support throughout. We would also like to thank Clarissa Peereboom and Joe Braxton for their excellent consultation regarding default aversion strategies and tools. Any remaining errors are ours alone.

Prior Research on the Factors Relating to Student Loan Default[2]

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 borrowers at TAMUK 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; Steiner & Teszler, 2005; 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 African-American greatly increases the probability of default. In three of the studies (Volkwein & Szelest, 1995; Volkwein et. al., 1995 and Woo, 2002), being African-American had the largest effect of all variables, and in the remainder of the cited studies, being African-American was the second most influential factor.

Prior studies have tested only a few variables that measure the borrower's performance in college.[3] 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

[2] For a more comprehensive review of student loan default research, see TG's Student Loan Default Literature Review, McMillion (2004) available at http://www.tgslc.org/schools/index.cfm under Default Aversion.

[3] TG's recent study of borrowers at Texas A&M College Station (Steiner & Teszler, 2005) is an exception. This study is discussed in detail below.

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 (1998) 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 parental income levels to be associated with decreases in the probability of default (Dynarksi, 1994; Knapp & Seaks, 1992; 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. Other demographic variables that researchers have found to significantly increase a borrowers probability of default are not having two parents at home, (Knapp & Seaks, 1992), having parents who did not attend college (Volkwein et. al., 1995), being Hispanic (Dynarksi, 1994; Woo, 2002), having dependents (Dynarksi, 1994; Volkwein & Szelest, 1995; Volkwein et. al., 1995; Woo, 2002), being an unmarried borrower (Dynarksi, 1994; Volkwein & Szelest, 1995; Volkwein et. al., 1995), and having low post-college borrower income (Dynarksi, 1994; Volkwein & Szelest, 1995; Volkwein et. al., 1995; Woo, 2002).

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, 1992; 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. 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.

The present study follows closely the methodology used in TG's recent multivariate analysis of borrowers at Texas A&M – College Station (TAMU). That study utilized data provided by TAMU to evaluate more variables than the research cited above, including variables which measured a borrower's success in both high school and college, a borrower's attendance pattern at TAMU, demographic variables, and a large number of financial aid variables. That study found that a borrower's probability of default was lower if the borrower: had a higher GPA at TAMU, graduated, received exit counseling, had personal income greater than $12,000, attended a college other than the liberal arts college while at TAMU, had an expected family contribution greater than $10,000, was female, did not fail any courses, or had transfer hours. Borrowers who were 35 or older when they entered repayment, were African-American, or whose mother had less than a high school education had a higher probability of default than borrowers who did not have these characteristics.[4]

Methodology for Multivariate Analysis of Defaulters at TAMUK

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.

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. (For a complete list of the variables examined in this study, refer to Appendix A.) In general, variables that did not have a statistically significant relationship to default were not included in the final model.

[4] Multivariate Analysis of Student Loan Defaulters at Texas A&M University is available at http://www.tgslc.org/schools/index.cfm. TG will release two additional reports in 2005: one looking at University of South Florida and the other at Prairie View A&M University.

Some groups of variables tend to provide similar explanations of default behavior and are therefore redundant; in many such cases it is possible to select one variable to represent the other variables. This is particularly apparent among the population of borrowers at TAMUK. As evidenced by the frequencies in Table A1, there are many variables that appear to be strongly related to the likelihood of default and that are not included in the final model. However, at TAMUK, many of these variables are so highly correlated with graduation status and/or GPA that they have no relation to default once they are included in the logistic regression model.

Sometimes a variable is so important from a theoretical or practical standpoint that the modeler must include it, even if it is not found to be significant. The borrower's age is an example of this in the current study. 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 Table 1. The table lists each variable, its reference group, the coefficient, and the change in probability of default, each of which will be explained below.

The impacts of the variables in the model are all measured in relation to a reference group for the variable. The multivariate estimation process produces a coefficient for each variable. This coefficient measures the impact of the variable on a student's likelihood of default, as compared to the reference group. 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.

The presence of an asterisk 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.

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 last column of the table represents the percentage point change in the probability of default given the presence of a characteristic. This change is only reported for variables that are statistically significant.[5]

For example, the variable Grade Point Average (GPA) = 2.51-3.00 has a coefficient equal to -.531. This means that a student with a GPA in this range has a lower likelihood of defaulting on his or her student loan within the two year cohort period than a student with a GPA between 2.01 and 2.50, the reference group. The presence of two asterisks next to the coefficient indicates that there is a 99 percent degree of confidence that students with the higher GPA have a lower

[5] For those who would like to calculate additional measures of significance, Appendix B includes a table containing the standard errors of the coefficients and confidence intervals for the change in probability.

likelihood of defaulting than students with an average GPA. In other words, there is only a one-percent chance that the difference in coefficients results from a particular characteristic of this sample, rather than representing a true relationship.

Continuing to look at the variable GPA, it is evident that GPA = 3.01-4.00 has a coefficient = -1.627, which is also statistically significant. However, looking only at these raw coefficients, it is difficult to translate the difference in magnitude between -1.627 and -.531 into a difference in default rates. The last column of the table assists in interpreting these coefficients. A baseline probability of default was calculated, based on the model when all variables are valued at their reference group. This baseline probability of default is 13 percent. The change in probability column gives the change in probability of default from this baseline due to moving a variable from its reference group to the indicated value. For example, a student with all variables measured at their reference group has a 13 percent probability of default. However, if this student has a GPA in the range 2.51-3.00, as opposed to 2.01-2.50, this probability of default drops by five percentage points to eight percent. If the student has an even higher GPA, in the 3.01-4.00 range, the probability of default drops even further, to only three percent. A discussion of the impact of each group of variables follows the table.

When interpreting these results it is important to remember that there is always uncertainty in any statistical model. The results of a statistical analysis tend to better describe the sample from which they were produced than any other sample or group. Therefore care must be taken when generalizing the results of any particular study. Despite these general limitations, there is a great deal of information to be learned from this specific study. The results of this multivariate study of students at TAMUK are very robust,[6] providing a great deal of confidence in their applicability to current and future students. The value of the model is not in predicting that students with GPAs above 3.0 have a probability of default that is exactly 10 percentage points less than that of students with a GPA between 2.0 and 2.5. Rather, the value of the model is that it provides a high level of confidence that a student's GPA provides a significant amount of information about that student's probability of default, even after controlling for the student's other characteristics.

[6] A robust statistical model is one in which the results are not highly dependent on the variables included in the model. The current study is very robust. The authors ran several versions of the model and found that the results were highly consistent across the various models.

TABLE 1
Results of Multivariate Analysis
Texas A&M – University Kingsville Undergraduates

Variable	Value	Reference Group	Coefficient		Change in Probability
	Intercept		-1.902	***	
College Success Variables					
Graduation Indicator	Graduated	Did not graduate	-0.708	***	-6%
Grade Point Average	0-1.00	2.01-2.50	0.466	**	6%
	1.01-2.00	2.01-2.50	0.178		
	2.51-3.00	2.01-2.50	-0.531	**	-5%
	3.01-4.00	2.01-2.50	-1.627	***	-10%
College Preparedness					
High School Class Rank Percentile	0% - 25%	51% - 65%	0.242		
	26% - 50%	51% - 65%	-0.031		
	66% - 80%	51% - 65%	-0.070		
	81% - 100%	51% - 65%	-0.590	*	-5%
	Missing	51% - 65%	0.103		
Demographics					
Age of Borrower At Time of Entering Repayment	17-20	21-22	-0.066		
	23-26	21-22	-0.063		
	27-30	21-22	-0.206		
	31-34	21-22	-0.253		
	35 or older	21-22	0.352		
Gender of Borrower	Female	Male	-0.438	***	-4%
Ethnicity of Borrower	African-American	Caucasian and Asian/ Pacific Islander	0.918	***	14%
	Hispanic		0.469	**	6%
Financial Aid Variables					
Expected Family Contribution (EFC)	Zero	$1 - 5,000	0.268	*	3%
	$5,001 and higher	$1 - 5,000	-0.684	**	-6%
	Missing	$1 - 5,000	-0.181		
Dependency Status	Dependent	Independent and Missing	-0.133		
Exit Counseling	Did not receive	Did receive	0.613	***	9%
Loan-Related Variables					
Highest Grade At Which Borrower Received a Loan	Sophomore	Freshmen	-0.515	**	-5%
	Junior	Freshmen	-0.337		
	Senior	Freshmen	-0.744	***	-6%
	Graduate	Freshmen	-0.802	*	-7%

Sample Size: 5,177 Defaulters: 569 (11.0 percent)

-2 log likelihood: Intercept and covariates: 2,891

Chi-Square: 694.33 with 27 degrees of freedom (Pr > ChiSq = <.0001)

C Statistic: 81.6 percent

Baseline probability of default (intercept only): 13.0 percent

* Statistically significant at the 0.05 level

** Statistically significant at the 0.01 level

*** Statistically significant at the 0.001 level

Note: The value “African-American” includes 342 borrowers whose ethnicity was reported as “Black,” 18 borrowers of ethnicity “Native Indian/Alaskan” and 15 borrowers of ethnicity “Other.”

No single statistic – whether the coefficient, the level of significance or the 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 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 higher a student's grade point average is, the less likely the student is to default on his or her student loan, after controlling for the other variables in the model. For borrowers at TAMUK, this relationship is strong and statistically significant. Borrowers with a GPA between 3.01 and 4.00 have a probability of default which is 10 percentage points lower than borrowers with a GPA between 2.01 and 2.50. Borrowers with a GPA between 2.51 and 3.00 have a probability of default which is five percentage points lower than borrowers with an average GPA. However, borrowers with a low GPA of 0 to 1.00 have a probability of default which is 6 percentage points higher than borrowers with a GPA between 2.01 and 2.50. This result is especially useful for a school — such as TAMUK — that is interested in lowering its default rate. It is relatively simple for a financial aid office to obtain information about a student's GPA. This result suggests that by monitoring students with low GPAs and providing additional financial aid counseling to these students, TAMUK may be able to lower its cohort default rate.

These results of the multivariate analysis confirm the relationship between GPA and default as noted in the bivariate table in Appendix A. That table shows that borrowers with a GPA of 3.01 to 4.00 have a default rate of only 1.2 percent, whereas borrowers with a GPA of 0 to 1.00 have a default rate of 28.8 percent. The multivariate analysis reveals that even after controlling for factors such as graduation, exit counseling and demographics, there is a strong relationship between a borrower's GPA and his or her probability of default.

There are many reasons why a higher GPA may lead to a lower probability of default. It is likely that this variable measures personal characteristics such as conscientiousness, persistence, motivation, intelligence and discipline which lead to success both in college and in loan repayment after college. Students who perform well in college are also more likely to complete their degrees and may earn more after college, making it easier to repay their student loans.

Exit Counseling

The borrower who does not receive exit counseling has a chance of defaulting that is nine percentage points higher than the borrower who does receive 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 exit counseling helps prevent student loan default.

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 exit counseling requirements. Since enforcing 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. However, even given this caveat, the exit counseling variable has considerable practical implications, especially in combination with the information provided by GPA. These results suggest that increased efforts to provide counseling to at-risk students may have a measurable impact on TAMUK's default rate.

<u>Graduation Indicator</u>

The graduation indicator is an especially important variable at TAMUK. Looking at the frequency table in Appendix A, we see that only 43.1 percent of the sample of borrowers graduated from TAMUK. However, of those borrowers who did graduate, only 2.4 percent of them defaulted on their student loans. Borrowers who did not graduate had a default rate of 17.5 percent. The results of the multivariate analysis show that this highly significant relationship between graduation status and default holds true, even after controlling for other student characteristics. A borrower who graduates has a chance of defaulting that is six percentage points lower than a borrower who does not graduate, all other things being equal.
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) toward whom they can direct default aversion efforts.

<u>Highest Grade at Which Borrower Received a Loan</u>

This variable is a proxy for persistence in school. As expected, the model predicts that students who remain in school longer have a lower probability of defaulting on their student loans than those who leave school after their freshman year. Students who take out their last loan during their senior year have a probability of defaulting that is six percentage points lower than those who take out their last loan during their freshman year. As noted in Appendix A, this difference is even greater at the bivariate level. Those who left TAMUK after their freshman year had a default rate of 23.0 percent, whereas those who stayed until their senior year had a default rate of only 4.0 percent. Once again, the data show that persistence in school is directly related to student loan default behavior.

<u>Race/Ethnicity of Borrower</u>

Consistent with other research on default behavior, this study finds that race/ethnicity is strongly related to default.[7] Holding all else equal, African-American students have a likelihood of default which is 14 percentage points higher than Caucasian students, and Hispanic students have a probability of default which is six percentage points higher.

---

[7] The authors would like to make it clear that they are not making a recommendation that any default aversion policies be based upon race. Race and gender are included in the model as control variables in order to correctly evaluate the effects of other variables such as GPA, graduation status, and exit counseling, upon which default aversion policies can be based.

Race or ethnicity is unlikely to have any direct relationship to default behavior. This study, just like prior research, is unable to explain the strong relationship between race and default behavior. It is most likely that the racial category is a proxy or label for a set of socioeconomic conditions, which are not otherwise measured, that make it more or less likely that one group will default relative to another group.

This model shows that African-American students are more likely to default on their student loans, at every given level of college performance. However, this difference is lower for students who perform better in college. For example, a Caucasian male who received exit counseling, graduated from TAMUK with a 2.7 GPA, had no expected family contribution, graduated in the 60^{th} percentile of his high school class, and entered repayment at age 22 has a 2.6 percent probability of defaulting during the cohort window. An African-American male with the same characteristics has a 6.3 percent probability of defaulting. While this difference of 3.7 percentage points is still pronounced, it is much less than the 14 percentage point difference for students who did not graduate and left school after their freshman year.

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, Caucasian and African-American 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 African-American and Caucasian borrowers with the same credentials, then the higher likelihood of African-American borrowers to default might merely reflect their diminished capacity to repay as a result of this discrepancy in 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, which are not controlled for in the model. While the presence of these persistent differences is discouraging, it is encouraging to realize that, even for the highest risk group (African-American students), success in college translates into a significantly lower probability of default on student loans.

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 average borrower at TAMUK has an EFC between $1 and $5,000. Borrowers with an EFC of zero have a probability of default that is three percentage points higher than this reference group, and borrowers with an EFC greater than $5,000 have a probability of default that is six percentage points lower than the reference group. In general, theory suggests that higher amounts of expected family contribution are associated with higher family incomes. For independent students, the EFC represents resources that are directly available to repay student loans. For dependent students, the EFC represents the income of a student's parents. While this income 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.

High School Class Rank Percentile

This variable is a measure of a student's preparedness for college. Theory predicts that students who are better prepared for college will have an easier time adjusting to college life and will be more successful in college. Since we know that college success is a strong factor in default aversion, it follows that high school success will help to predict default behavior. The average student at TAMUK graduated in the 60^{th} percentile of his or her high school class. Therefore, the group containing this percentile was used as the reference group. Students who graduated in the top 20 percent of their high school class have a probability of default that is five percentage points lower than those graduating in the 51^{st} to 65^{th} percentile. No other group has a significantly different likelihood of default after controlling for the other variables in the model.

Gender

Female borrowers at TAMUK have a probability of default which is four percentage points lower than male borrowers. Though the reasons for this finding are unclear, it is consistent with prior research in this area.

Age of Borrower at Time of Entering Repayment

The age of the borrower at the time of entering repayment does not have a significant effect on a borrower's probability of default at TAMUK. This variable was included in the model because other studies, including a similar analysis of borrowers at Texas A&M – College Station performed by TG, have found that a borrower's age is related to his or her probability of default. Although the bivariate results indicate that younger repayers have a higher probability of default, once the model has controlled for the other, more significant factors such as graduation and GPA, age no longer has a significant effect in the model.

Dependency Status

Dependency status — another variable that was found to be significant in some other research — does not have a significant impact on the probability of default at TAMUK. It is important that this variable remain in the model as a control variable. The formula for calculating a student's EFC is affected by the student's dependency status. Therefore, by including dependency status in the model, we obtain a more precise estimate of the effect of the student's EFC.

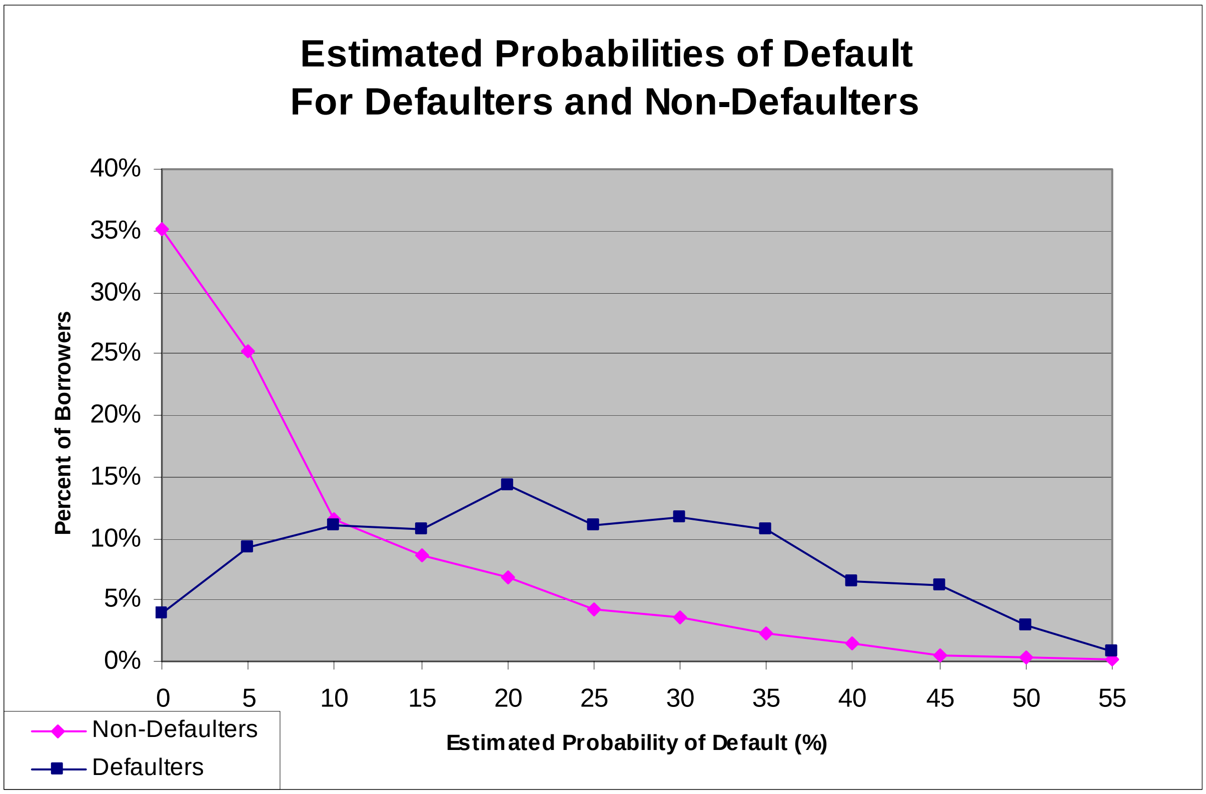
Model Performance

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.

Distribution of probabilities

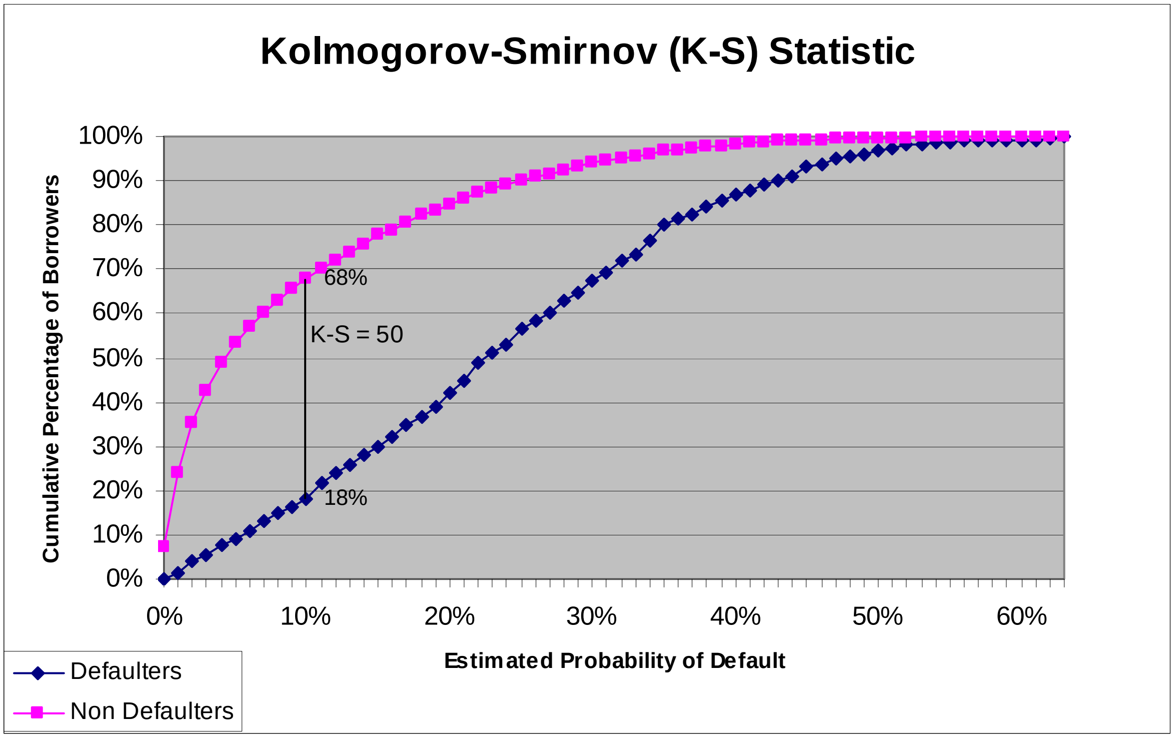
The following chart shows the default probabilities assigned by the multivariate model to borrowers in the study. The chart provides a separate distribution of probabilities for actual defaulters and actual non-defaulters. (Each borrower's estimated probability of default was rounded to the nearest five 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 35 percent of actual non-defaulters, it assigned a zero percent (rounded) probability of default to only four 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.



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 68 percent of actual non-defaulters have an estimated probability of default that is less than or equal to ten percent and that only 18 percent of actual defaulters have an estimated probability of default in that range. As it turns out, at 10 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 50 percent (68%-18%). Models with large K-S statistics are said to have done a good job of distinguishing between defaulters and non-defaulters. Fifty percent (50%) is a 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 an outcome of default is predicted for borrowers to whom the model assigned a default probability greater than 10 percent. The K-S of 50 percent indicates that using 10 percent as the prediction cutoff means that this model will predict default 50 percent more frequently for defaulters than for non-defaulters. At 10 percent, the model would predict 82 percent of actual defaulters to default (that is, one minus the 18 percent with probabilities less than or equal to ten percent), but it would only predict 32 percent of actual non-defaulters to default (one minus the 68 percent with probabilities less than or equal to 10 percent).

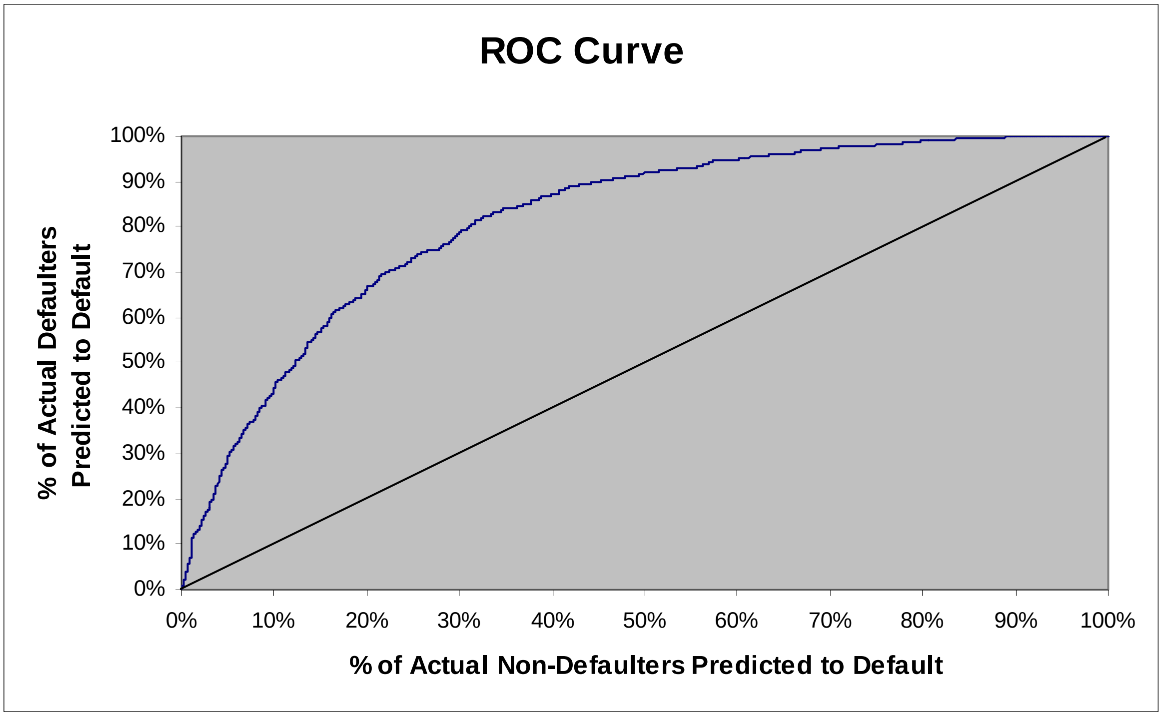


C Statistic

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 2,621,952 pairings (569 defaulters multiplied by 4,608 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 81.6 percent – a very 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: 81.6 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 10 percent or more, the borrower is classified as a defaulter; a borrower with less than a 10 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=5,177	Default	Non Default
Actual	Default	470	99
Outcome	Non Default	1,539	3,069

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 (1,638) is the sum of the defaulters who the model predicted to be non-defaulters (99) and the non-defaulters who the model predicted to be defaulters (1,539). The misclassification rate is 32 percent (=1,638/5,177).

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 32 percent is very good. Treating all borrowers as potential defaulters will misclassify all 4,608 non-defaulters and result in a misclassification rate of 89 percent. In this comparison, using the model results in a 64 percent 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 53 percent, since 76 defaulters in the study have GPAs greater than 2.5, and 2,667 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.

This misclassification rate is comprised of two measurements. A low misclassification rate indicates that a method of prediction is successful at predicting both defaulters and non-defaulters. By successfully predicting both, a school can most effectively target its resources to the predicted defaulters. Using the model, with a cutoff of 10 percent to predict defaulters, TAMUK would correctly identify 83 percent of the defaulters and 67 percent of the non-defaulters. As can be seen in the above chart, this means that they would needlessly counsel 1,539 non-defaulters and fail to counsel 99 defaulters.

However, if a school can counsel additional borrowers at a very low cost, it may choose to use a more aggressive method in order to capture a greater percentage of defaulters. For example, if TAMUK used a student's GPA as the predictor of default and provided additional counseling to all students with a GPA of 2.5 or lower, they would correctly identify 87 percent of the defaulters (493 of 569) but only 54 percent of the non-defaulters (2,510 of 4,608). In this scenario, TAMUK would provide unneeded counseling to 2,354 non-defaulters and fail to counsel 76 defaulters.

Which method a school uses to identify potential defaulters will ultimately depend on the costs of implementing the prediction versus the costs of needlessly counseling borrowers who would not otherwise default.

Uses of the Findings and the Model

There are many ways in which the results of this study can be used to assist TAMUK in preventing defaults by their student borrowers. These possibilities range from simple solutions requiring only minor changes to existing policies and procedures, to slightly more involved solutions involving the coordination of financial aid goals with the efforts of other campus functions such as academic advising, career counseling, and instruction. There are also more sophisticated ways in which TAMUK could use this statistical model to identify at-risk borrowers, although these may not be the most cost-effective methods for this campus.[8]

The findings of this study are largely consistent with the findings of other studies of default behavior. However, this study does highlight the strong significance of a relatively few number of variables at TAMUK. The importance of the completion, college success, and exit counseling variables suggest that targeting increased intervention to certain at-risk students could have a significant impact on TAMUK's default rate. Other research has suggested that by involving the entire campus community in assisting the financial aid office with default aversion policies, an institution can not only lower its default rate, but have the added benefits of increasing its completion rate and promoting student success on campus.[9] The results of this study suggest that this approach may be particularly rewarding at TAMUK.

- A student's grade point average is a very powerful predictor of student loan default. The model found that a student with a GPA of 3.01-4.00 has a probability of default which is 10 percentage points lower than a student with a GPA of 2.01-2.5, even after controlling for demographics, background characteristics, and graduation status. On a bivariate level, the relationship is even stronger. Students who left TAMUK with a GPA of 2.5 or less had a default rate of 18.4%, whereas those who left TAMUK with a GPA above 2.5 had a default rate of only 3.0%. By requiring any student whose GPA drops below 2.5 to receive both academic and financial counseling, TAMUK may reach students before they decide to leave school.
- The model illustrates the effects of persistence and completion in reducing a borrower's likelihood of default. After controlling for other characteristics, students who graduate have a probability of default which is six percentage points lower than students who do not graduate. This is the same percentage decrease seen when comparing students who leave after their freshman year to those who stay in school until their senior year. Once again, this effect is particularly dramatic on a bivariate level. Students who received their last loan as freshmen have a default rate of 23 percent, whereas those who received their

[8] The conclusions drawn in this section represent the informed opinions of the authors. They are in no way intended to be exhaustive or exclusive of other conclusions. These ideas are shared in the spirit of starting discussions, not ending them.
[9] See "A Clear and Present Danger to Institutional and Student Success." TG and the Council for the Management of Educational Finance, http://www.tgslc.org/pdf/default_model.pdf for more details.

last loan as seniors have a default rate of only four percent. Students who graduate from TAMUK have a default rate of only 2.4 percent. This result highlights the importance of working with the entire campus community to increase retention and graduation rates. Increases in these rates are good for the students, the campus as a whole, and as a bi-product, lower default rates.

- One very encouraging result provided by the model is the effectiveness of exit counseling. Students who receive exit counseling have an expected default rate which is nine percentage points lower than those who do not. This suggests that the payoffs to increased counseling will be high. Perhaps all freshmen should receive in-person counseling before their loan checks are processed. This is an excellent opportunity to work with financial aid partners, such as lenders. If they have not already done so, TAMUK may want to ask their primary lenders to come in the first week of classes and give 30- to 60-minute presentations to incoming freshmen, raising their awareness about the responsibilities of taking out student loans and giving some basic financial management tips. TG also offers free consulting services to assist campuses in designing successful default aversion strategies.

This model highlights the fact that successful default aversion strategies must be a part of a larger, campus-wide effort to retain students and improve their academic performance. Achieving these results is good for both students and the university, and lowers default rates. Borrowers who stay in school, maintain a B average, and successfully graduate have a much lower default rate than those who do not. Improving on these performance measures is especially difficult on a campus such as TAMUK, where many students are first-generation college students and have little or no financial and family support. However, the results of the model suggest that even slight improvements in these measures can have a significant impact on the cohort default rate, as well as be extremely beneficial to the students themselves.

Appendix A
Sample Definition and Variable Descriptions

Sample Definition

Using TG's loan data we identified students who attended TAMUK and who entered repayment on their TG-guaranteed Federal Family Education Loan Program (FFELP) loans between October 1, 1998 and September 30, 2002 (fiscal years 1999 – 2002). This resulted in a sample of 5,509 borrowers. Staff at TAMUK then retrieved detailed data for each of these borrowers from the school's internal databases. As described in the text, only a subset of these variables was used in the final analysis. However, we include a description of each of the variables in this section. Table A1, following the variable descriptions, contains basic descriptive statistics for each variable. Although data was collected for all 5,509 borrowers, this report analyzes the data for only the 5,177 borrowers who were undergraduate students at TAMUK. Therefore, the descriptive statistics are provided for only those undergraduate students.[10]

Variable Descriptions

Default or Not

This study uses the Department of Education's official cohort default rate formula to determine if a borrower is in default. Under this definition, a borrower is considered to be in default if a default claim is paid on his or her behalf during the fiscal year in which the borrower enters repayment, or the following fiscal year.[11] Out of the 5,177 borrowers in the sample of undergraduates, 569 (or 11 percent) defaulted under this definition. This variable is the focus of this study.

The subsequent variables all describe the characteristics of the borrowers. These variables will be used to determine if default behavior among borrowers at TAMUK varies among groups of borrowers with different characteristics.

College Success Variables

These variables were provided by TAMUK and measure the borrowers' performance while in attendance at TAMUK. These variables measure both the quantity of education received at TAMUK and the quality of the students' performance in college. It is expected that students who stay in school longer, obtain a degree, and receive higher grades are more likely to repay their student loans. Students who remain in school and receive high grades are more likely to obtain jobs and have higher earnings after college, enabling them to repay their student loans. Variables such as "Undergraduate GPA" and "Graduation Indicator" are also correlated with personal qualities, such as persistence and discipline, that make a borrower more likely to repay his or her loans.

---

[10] Statistics for the full sample, including graduate students are available from TG upon request.
[11] A fiscal year is defined as October 1 – September 30. For example, fiscal year 2000 covers the period October 1, 1999 – September 30, 2000.

Graduation Date - Date graduated from TAMUK.
Undergraduate GPA
Number of Hours Passed
Type of Degree

Attendance Pattern

The “Attendance Pattern” variables describe the length and intensity of a borrower’s attendance at TAMUK. 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. It is also possible that variables such as the student’s major or college of admittance are correlated with the probability of default. Students in certain majors may earn more than those in others, making them more likely to repay their loans successfully.

Type of Admission - Indicates if a student was admitted as a freshman, a transfer student, or readmitted.
College of Admittance
Admission Major
Date of First Attendance - Date first attended TAMUK.
Total Hours
Withdrawal Indicator – Indicates whether a borrower ever withdrew from TAMUK.
Number of Hours Transferred

College Preparedness

This set of variables describes students’ achievements in high school, before entering college. Past research has found that success is college is related to lower default rates on student loans. It is expected that success in high school is a good predictor of success in college. Therefore, we expect students who were successful in high school to have lower default rates than those who were not successful.

High School Class Rank Percentile
Total SAT Score
High School Graduation Date

Demographics

A borrower’s demographic characteristics describe the cultural and family background which a student brings to college. These variables affect a student’s probability of repaying his or her student loans in several different ways. Some of the variables, such as parent’s educational attainment may reflect a borrower's previous exposure to responsibilities such as repaying a student loan. Other variables, such as marital status and family size may proxy for the amount of resources a student has available to repay his or her student loan, with married students and students from large families having more competing demands on their resources, making them more likely to default. Prior research has shown that both the gender and ethnicity of a borrower affect the probability that a borrower defaults, with female and Caucasian borrowers having the lowest probability of default. However, the reasons for this relationship are not clear.

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
Texas Residency Status of Borrower
State of Permanent Address
Highest Degree Obtained by Borrower's Father
Highest Degree Obtained by Borrower's Mother
Parent's Marital Status

Financial Aid Variables

The financial aid variables capture the financial resources available to student borrowers when they enter school. It is expected that students who come from families with lower incomes will have a higher probability of default than students from higher income families. However, it may be that these students are more committed to their education and appreciate the chance they have, making them less likely to default.

Amount of Need
Expected Family Contribution
Adjusted Gross Income of Parents
Adjusted Gross Income of Student
Total Loan Aid
Total Work Aid
Total Other Aid
Dependency Status

Received Exit Counseling

Students who receive exit counseling default at lower rates than students who do not receive this counseling.

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. The highest grade at which a borrower received a TG loan is a direct proxy for persistence in school. The longer a borrower remains in school, the less likely he or she is to default.

Number of Loans
Consolidation Loan: Whether or not the borrower has a consolidation loan.
Total TG Loan Amount
Highest Grade At Which Borrower Received a TG Loan

TABLE A1
Characteristics of Undergraduate Borrowers at Texas A&M University - Kingsville
By Default Status

			Total		Defaulters	
			N	% of cases	N	% of row
Total Sample			5,177	100.0%	569	11.0%
College Success Variables						
	Graduation Indicator					
		Graduated	2,229	43.1%	54	2.4%
		Did Not Graduate	2,948	56.9%	515	17.5%
	Undergraduate GPA					
		0-1.00	579	11.2%	167	28.8%
		1.01-2.00	967	18.7%	203	21.0%
		2.01-2.50	1,121	21.7%	123	11.0%
		2.51-3.00	1,370	26.5%	62	4.5%
		3.01-4.00	1,140	22.0%	14	1.2%
	Number of Hours Passed					
		0	148	2.9%	45	30.4%
		1-25	1,108	21.4%	258	23.3%
		26-50	646	12.5%	90	13.9%
		51-100	784	15.1%	90	11.5%
		101-150	1,394	26.9%	55	3.9%
		151 or more	1,097	21.2%	31	2.8%
	Type of Degree					
		No Degree Earned	2,948	56.9%	515	17.5%
		Bachelor of Arts	605	11.7%	22	3.6%
		Bachelors Degree	43	0.8%	2	4.7%
		Bachelor of Business Administration	256	4.9%	3	1.2%
		Bachelor of Engineering	358	6.9%	7	2.0%
		Bachelor Science	967	18.7%	20	2.1%

TABLE A1 – cont'd

			Total		Defaulters	
			N	% of cases	N	% of row
Total Sample			5,177	100.0%	569	11.0%
Attendance Pattern						
	Type of Admission					
		Freshman	1,434	27.7%	249	17.4%
		Readmit	1,271	24.6%	142	11.2%
		Transfer	854	16.5%	76	8.9%
		Missing	1,618	31.3%	102	6.3%
	College of Admittance					
		Agriculture	159	3.1%	8	5.0%
		Business	231	4.5%	21	9.1%
		Engineering	191	3.7%	13	6.8%
		Liberal Arts	2,056	39.7%	371	18.0%
		Science	625	12.1%	35	5.6%
		Other	1,915	37.0%	121	6.3%
	Date of First Attendance					
		1962 - 1979	63	1.2%	8	12.7%
		1980 - 1989	231	4.5%	16	6.9%
		1990	89	1.7%	5	5.6%
		1991	139	2.7%	13	9.4%
		1992	250	4.8%	14	5.6%
		1993	388	7.5%	27	7.0%
		1994	485	9.4%	27	5.6%
		1995	551	10.6%	27	4.9%
		1996	678	13.1%	40	5.9%
		1997	692	13.4%	91	13.2%
		1998	672	13.0%	118	17.6%
		1999	483	9.3%	93	19.3%
		2000	370	7.1%	77	20.8%
		2001	90	1.7%	13	14.4%
		2002	2	0.0%	0	0.0%
	Total Hours					
		0 - 25	695	13.4%	142	20.4%
		26-50	1,073	20.7%	208	19.4%
		51-100	1,081	20.9%	117	10.8%
		101-150	1,033	20.0%	46	4.5%
		151-250	1,199	23.2%	50	4.2%
		251 or more	96	1.9%	6	6.3%
	Withdrawal Indicator					
		No	4,624	89.3%	507	11.0%
		Yes	553	10.7%	62	11.2%
	Number of Hours Transferred					
		0	3,287	63.5%	430	13.1%
		1-23	868	16.8%	72	8.3%
		24-57	676	13.1%	42	6.2%
		58-90	294	5.7%	22	7.5%
		91 or more	52	1.0%	3	5.8%

TABLE A1 – cont'd

			Total		Defaulters	
			N	% of cases	N	% of row
Total Sample			5,177	100.0%	569	11.0%
College Preparedness						
	High School Class Rank Percentile					
		0% - 25%	418	8.1%	102	24.4%
		26% - 50%	701	13.5%	111	15.8%
		51% - 65%	582	11.2%	74	12.7%
		66% - 80%	700	13.5%	64	9.1%
		81% - 100%	727	14.0%	27	3.7%
		Missing	2,049	39.6%	191	9.3%
	Total SAT Score					
		0 to 800	477	9.2%	47	9.9%
		801 to 930	273	5.3%	25	9.2%
		931 to 1120	212	4.1%	20	9.4%
		1121 or more	52	1.0%	6	11.5%
		Missing	4,163	80.4%	471	11.3%
	High School Graduation Date					
		1958 - 1979	170	3.3%	16	9.4%
		1980 - 1989	490	9.5%	28	5.7%
		1990 - 1994	1,565	30.2%	105	6.7%
		1995	521	10.1%	37	7.1%
		1996	506	9.8%	50	9.9%
		1997	469	9.1%	70	14.9%
		1998	430	8.3%	95	22.1%
		1999	245	4.7%	50	20.4%
		2000 - 2001	225	4.3%	41	18.2%
		Missing	556	10.7%	77	13.8%

TABLE A1 – cont'd

			Total		Defaulters	
			N	% of cases	N	% of row
Total Sample			5,177	100.0%	569	11.0%
Demographics						
	Borrower's Age When Enter Repayment					
		17-20	446	8.6%	96	21.5%
		21-22	921	17.8%	173	18.8%
		23-26	2,167	41.9%	167	7.7%
		27-30	870	16.8%	64	7.4%
		30-34	297	5.7%	23	7.7%
		34+	476	9.2%	46	9.7%
	Borrower's Gender					
		Female	2,518	48.6%	203	8.1%
		Male	2,659	51.4%	366	13.8%
	Borrower's Ethnicity					
		Hispanic	3,741	72.3%	415	11.1%
		Caucasian	1,044	20.2%	65	6.2%
		African-American	342	6.6%	82	24.0%
		Native Indian/Alaskan	18	0.3%	5	27.8%
		Asian/Pacific Islander	17	0.3%	0	0.0%
		Other	15	0.3%	2	13.3%
	Borrower's Marital Status					
		Single	4,348	84.0%	503	11.6%
		Married	658	12.7%	42	6.4%
		Divorced and Other	171	3.3%	24	14.0%
	Citizenship of Borrower					
		Non-U.S.	31	0.6%	3	9.7%
		United States	17	0.3%	0	0.0%
		Missing	5,129	99.1%	566	11.0%
	Texas Residency Status of Borrower					
		Texas	4,235	81.8%	504	11.9%
		Non-Texas	89	1.7%	7	7.9%
		Missing	853	16.5%	58	6.8%
	State of Permanent Address					
		Texas	5061	97.8%	558	11.0%
		Non-Texas	105	2.0%	8	7.6%
		Missing	11	0.2%	3	27.3%
	Highest Degree Obtained by Borrower's Father					
		Middle School/Junior High	511	9.9%	57	11.2%
		High School	1,315	25.4%	149	11.3%
		College or Beyond	771	14.9%	57	7.4%
		Unknown	373	7.2%	52	13.9%
		Missing	2,207	42.6%	254	11.5%
	Highest Degree Obtained by Borrower's Mother					
		Middle School/Junior High	461	8.9%	49	10.6%
		High School	1,190	23.0%	114	9.6%
		College or Beyond	786	15.2%	70	8.9%
		Unknown	230	4.4%	28	12.2%
		Missing	2,510	48.5%	308	12.3%

TABLE A1 – cont'd

			Total		Defaulters	
			N	% of cases	N	% of row
Total Sample			5,177	100.0%	569	11.0%
Demographics – cont'd						
	Parent's Marital Status					
		Married/Remarried	1,168	22.6%	116	9.9%
		Separated	499	9.6%	60	12.0%
		Unmarried	101	2.0%	22	21.8%
		Widowed	91	1.8%	15	16.5%
		Missing	3,318	64.1%	356	10.7%
Financial Aid Variables						
	Amount of Need					
		Zero	233	4.5%	11	4.7%
		$1-2,500	333	6.4%	20	6.0%
		$2,501-7,500	1,756	33.9%	183	10.4%
		$7,501-10,000	2,066	39.9%	245	11.9%
		$10,001 and higher	266	5.1%	42	15.8%
		Missing	519	10.0%	68	13.1%
	Expected Family Contribution					
		Zero	1,512	29.2%	254	14.4%
		$1-500	453	8.8%	51	10.1%
		$501-1,000	278	5.4%	31	10.0%
		$1,001-2,000	432	8.3%	43	9.1%
		$2,001-3,000	287	5.5%	35	10.9%
		$3,001-5,000	358	6.9%	37	9.4%
		$5,001-7,000	190	3.7%	8	4.0%
		$7,001-10,000	174	3.4%	8	4.4%
		$10,001 and higher	249	4.8%	13	5.0%
		Missing	675	13.0%	89	11.6%
	Adjusted Gross Income of Parents					
		Zero	3,938	76.1%	438	11.1%
		$1-20,000	480	9.3%	64	13.3%
		$20,001-30,000	196	3.8%	22	11.2%
		$30,001-40,000	138	2.7%	17	12.3%
		$40,001-60,000	198	3.8%	17	8.6%
		$60,001-80,000	150	2.9%	6	4.0%
		$80,001 and higher	77	1.5%	5	6.5%
	Adjusted Gross Income of Student					
		Zero	3,574	69.0%	437	12.2%
		$1-2,000	385	7.4%	36	9.4%
		$2001-4,000	306	5.9%	26	8.5%
		$4001-6,000	291	5.6%	24	8.2%
		$6,001-12,000	378	7.3%	33	8.7%
		$12,001 and higher	243	4.7%	13	5.3%
	Total Loan Aid					
		Up to $3,000	952	18.4%	201	21.1%
		$3,001 to 5,000	447	8.6%	66	14.8%
		$5,001 to 7,000	597	11.5%	91	15.2%
		$7,001 to 9,000	311	6.0%	35	11.3%
		$9,001 to 12,000	419	8.1%	45	10.7%
		$12,001 or more	2,451	47.3%	131	5.3%

TABLE A1 – cont'd

			Total		Defaulters	
			N	% of cases	N	% of row
Total Sample			5,177	100.0%	569	11.0%
Financial Aid Variables – cont'd						
	Work Study Aid					
		No Work Study Aid	4,367	84.4%	515	11.8%
		Received Work Study Aid	810	15.6%	54	6.7%
	Total Other Aid					
		$0	1,206	23.3%	94	7.8%
		$1 to 1,000	293	5.7%	34	11.6%
		$1,001 to 3,000	1,086	21.0%	191	17.6%
		$3,001 to 4,000	500	9.7%	86	17.2%
		$4,001 to 6,000	551	10.6%	52	9.4%
		$6,001 to 9,000	705	13.6%	72	10.2%
		$9,001 or more	836	16.1%	40	4.8%
	Dependency Status					
		Dependent	1,567	30.3%	184	11.7%
		Independent	3,610	69.7%	385	10.7%
Received Exit Counseling						
	No		1,883	36.4%	350	18.6%
	Yes		3,294	63.6%	219	6.6%
Loan-related Variables						
	Number of Loans					
		1	827	16.0%	167	20.2%
		2 to 4	1,524	29.4%	243	15.9%
		5 to 6	710	13.7%	54	7.6%
		7 to 9	883	17.1%	53	6.0%
		10 or more	1,233	23.8%	52	4.2%
	Has Consolidation Loan					
		No	4,387	84.7%	501	11.4%
		Yes	790	15.3%	68	8.6%
	Total TG Loan Amount					
		$1 to 3,000	886	17.1%	194	21.9%
		$3,001 to 6,000	672	13.0%	106	15.8%
		$6,001 to 9,000	578	11.2%	96	16.6%
		$9,001 to 12,000	411	7.9%	44	10.7%
		$12,001 to 16,000	500	9.7%	40	8.0%
		$16,001 to 20,000	523	10.1%	25	4.8%
		$20,001 to 30,000	875	16.9%	35	4.0%
		$30,001 or more	732	14.1%	29	4.0%
	Highest Grade At Which Borrower Received A Loan					
		Freshman	1,541	29.8%	355	23.0%
		Sophomore	577	11.1%	67	11.6%
		Junior	425	8.2%	51	12.0%
		Senior	2,088	40.3%	83	4.0%
		Graduate	546	10.5%	13	2.4%

Appendix B

Standard Errors and Confidence Intervals

Variable Group	Variable	Reference Group	Coefficient	Standard Error	95% Confidence Interval of Change in Probability
	Intercept		-1.902	0.255	
College Success Variables					
Graduation Indicator	Graduated	Did not graduate	-0.708	0.210	-8% to -3%
Grade Point Average	0-1.00	2.01-2.50	0.466	0.158	2% to 11%
	1.01-2.00	2.01-2.50	0.178	0.139	-1% to 6%
	2.51-3.00	2.01-2.50	-0.531	0.173	-7% to -2%
	3.01-4.00	2.01-2.50	-1.627	0.298	-11% to -8%
College Preparedness					
High School Class Rank Percentile	0% - 25%	51% - 65%	0.242	0.183	-1% to 8%
	26% - 50%	51% - 65%	-0.031	0.173	-4% to 4%
	66% - 80%	51% - 65%	-0.070	0.193	-4% to 4%
	81% - 100%	51% - 65%	-0.590	0.245	-8% to -1%
	Missing	51% - 65%	0.103	0.175	-2% to 6%
Demographics					
Age of Borrower At Time of Entering Repayment (in years)	17-20	21-22	-0.066	0.143	-3% to 3%
	23-26	21-22	-0.063	0.157	-4% to 3%
	27-30	21-22	-0.206	0.218	-6% to 3%
	31-34	21-22	-0.253	0.302	-7% to 4%
	35 or older	21-22	0.352	0.230	-1% to 12%
Gender of Borrower	Female	Male	-0.438	0.102	-6% to -2%
Ethnicity of Borrower	African-American	Caucasian and Asian/ Pacific Islander	0.918	0.194	7% to 22%
	Hispanic		0.469	0.149	2% to 11%
Financial Aid Variables					
Expected Family Contribution (EFC)	Zero	$1 - 5,000	0.268	0.112	1% to 7%
	$5,001 and higher	$1 - 5,000	-0.684	0.215	-8% to -3%
	Missing	$1 - 5,000	-0.181	0.154	-5% to 1%
Dependency Status	Dependent	Independent and Missing	-0.199	0.123	-4% to 0%
Exit Counseling	Did not receive	Did receive	0.613	0.102	5% to 12%
Loan-Related Variables					
Highest Grade At Which Borrower Received a Loan	Sophomore	Freshmen	-0.515	0.165	-7% to -2%
	Junior	Freshmen	-0.337	0.192	-6% to 0%
	Senior	Freshmen	-0.744	0.195	-8% to -4%
	Graduate	Freshmen	-0.802	0.344	-10% to -1%

Sample Size: 5,177 Defaulters: 569 (11.0 percent)

-2 log likelihood: Intercept and covariates: 2,891

Chi-Square: 694.33 with 27 degrees of freedom (Pr > ChiSq = <.0001)

C Statistic: 81.6 percent

Baseline probability of default (intercept only): 13.0 percent

* Statistically significant at the 0.05 level

** Statistically significant at the 0.01 level

*** Statistically significant at the 0.001 level

Note: The value “African-American” includes 342 borrowers whose ethnicity was reported as “Black,” 18 borrowers of ethnicity “Native Indian/Alaskan” and 15 borrowers of ethnicity “Other.”

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.

Steiner, M. and Teszler, N (2005). Multivariate Analysis of Student Loan Defaulters at Texas A&M University. Austin, TX. January.

TG and the Council for the Management of Educational Finance (2003). *A Clear and Present Danger to Institutional and Student Success: A training model for embedding student loan default aversion within strategic enrollment management*. Austin, TX.

U.S. Department of Education (2001). *The Cohort Default Rate Guide*. Washington D.C. http://www.ifap.ed.gov/drmaterials/finalcdrg.html#mulifiles .

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.