Journal of Student Financial Aid

Volume 32 | Issue 3 Article 3

12-1-2002

Student Loan Defaults and Enrollment Persistence

Michael Podgursky

Mark Ehlert

Ryan Monroe

Donald Watson

Student Loan Defaults and Enrollment Persistence

By Michael Podgursky, Mark Ehlert, Ryan Monroe, Donald Watson, and John Wittstruck

Michael Podgursky is Middlebush Professor of Economics and Department Chair, and Ryan Monroe and Donald Watson are Research Associates, for the Department of Economics, University of Missouri - Columbia. Mark Ehlert is Research Assistant Professor for the Department of Educational Leadership and Policy Analysis, University of Missouri - Columbia. John Wittstruck is Deputy Commissioner of Higher Education, Missouri Department of Higher Education.

We wish to thank the Missouri Department of Higher Education for assistance and financial support. An early draft of this paper was presented at the 17th Annual NASSGAP/ NCHELP Financial Aid Research Conference in St. Louis, MO on June 2000. The views expressed in this paper are those of the authors and do not necessarily represent the views of the Missouri Department of Higher Education.

This article provides an estimate of a model of student loan defaults using a rich panel data file. The file was constructed by merging administrative data on student loans, higher education enrollment and performance, and ACT test data for a large cohort of first time, full-time, degree-seeking students who entered Missouri two- and four-year public higher education institutions in the Fall 1992 semester. These loan recipients were tracked forward to December 1999 to determine which ones defaulted on their loans.

The authors identify a variety of individual characteristics associated with loan defaults; however, the variable with the largest effect on the default odds ratio is continuous enrollment. Within windows ranging from four to eight semesters, students who are continuously enrolled or who complete their program are far less likely to default than are students who drop out during the same period The authors also assess the predictive power of their statistical model "out of sample" on a subsample of student borrowers and illustrate the potential use of the model in targeting default prevention resources to students most at risk of default.

The combined effects of increasing college enrollments and increasing costs for college attendance are generating a rapidly growing demand for funds to finance higher education. While some students can finance their postsecondary education with grants, scholarships, or family wealth, many students can only attend college by borrowing through various student loan programs. The federal government, through its Federal Family Education Loan and Federal Direct Student Loan programs, is the largest provider of student loans for postsecondary education students. In fiscal year 1998, the U.S. General Accounting Office reported that the nearly 8.1 million students received federal student loans, and these students borrowed roughly $53 billion (U.S. General Accounting Office [GAO], 2001).

Not all college students who borrow through the federal loan programs repay their loans. A federal student loan is considered to be "in default" if a borrower's scheduled loan repayment is 270 days or more overdue and the borrower has not died, become permanently and totally disabled, or qualified for a loan deferment or forbearance. As the number and size of federal student loans has increased, the cost of loan defaults

has increased as well. In fiscal year 1998, for example, the cost to the federal government of defaulted loans was $2.1 billion (GAO, 1999). Because student loan defaults result in higher program costs to taxpayers, policymakers should be interested in identifying characteristics of borrowers that can predict loan defaults. Identification of these characteristics can inform the development of loan default intervention strategies.

Prior Research on Student Loan Defaults

Past studies of student loan defaults were based on two types of data. Stockham and Hesseldenz, (1979), Myers and Siera (1980), and Greene (1989), examined the characteristics of borrowers and defaulters using administrative data from a single institution, while Wilms, Moore, and Bolus (1987), and Knapp and Seaks (1992) considered administrative data from a set of selected institutions in a single state. A second group of studies used retrospective household survey data from the National Postsecondary Student Aid Study (NPSAS) (Dynarski, 1994; Flint, 1997; Volkwein et al., 1998). All of these studies used characteristics of borrowers and institutions to predict students who defaulted on their loans.

In general, these studies found that student demographics, family background, ability to pay, type of institution attended, and whether or not students completed their postsecondary programs of study were significantly related to loan defaults. Of particular interest for this study was the consistent finding that program or degree completion tended to reduce significantly the probability of loan default (Dynarski, 1994; Volkwein et al., 1998). These studies included a self-reported dichotomous variable indicating program completion. However, neither Dynarski nor Volkwein et al. had data available that would allow examination of the dynamics of enrollment patterns and their effects on loan defaults. Our study contributes to this literature by examining the effect of enrollment dynamics on default risk using a new longitudinal data file on Missouri higher education students.

Three Sources of Data

The data for this study came from three sources. The first was the Enhanced Missouri Student Achievement Study (EMSAS) database. Since 1986, the Missouri Coordinating Board for Higher Education has maintained a longitudinal database of students enrolled in Missouri public higher education (McDaniel, 1988). EMSAS records were merged with student loan records provided to us by the Missouri Student Assistance Resource Services (MOSTARS) to create a longitudinal data set of borrowers in Missouri public higher education. Finally, the student

records were matched to ACT college admissions test scores for 1990-91.

Additionally, enrollment and loan information were compiled for all students enrolled in public two- and four-year institutions in the Fall 1992 semester. In that semester, there were 202,140 students enrolled in Missouri two- and four-year public institutions. Thirty-two percent of these students had student loans or would subsequently borrow. We examined student administrative records from Fall 1992 through December 1999 to determine enrollment patterns, highest degree earned, and loan defaults. To focus on students who were or should have been in repayment, students who were enrolled in Missouri public two- or four-year institutions in the Spring or Fall 1999 semesters were excluded from the study dataset. In the sample of borrowers, 12.7 percent had defaulted on at least one loan by December 1999.

The cross-tabulations in Table 1 show that the default rate was strongly related to the level of the highest degree completed. The default rate was highest for the less-than-two-year certificates (20 percent) and lowest for post-baccalaureate students (2.5 percent). The default rate for those who did not complete a degree (20 percent) was higher than for any group except the less than two-year graduates.

The dropout group is of particular concern because it was big. Forty-five percent of the Fall 1992 student borrowers did not earn a degree or certificate from a Missouri public institution by June 1999.[1] These dropouts accounted for 81 percent of all loan defaulters, and 73 percent of the dollar-weighted share of loans.

Loan Default Model and Estimates

To understand better the effects of enrollment persistence and dropping out on loan defaults, we focused on a more homogeneous cohort of first-time, full-time, degree-seeking students who entered a public two- or four-year institution in Missouri the Fall 1992 semester. Because students can receive student loans at any point in their educational career, it is useful to concentrate on loans from a specific period of time. Our student

[1] The share of non-graduates in the Missouri sample is much larger than that in the NPSAS household survey (Dynarski, 1994). Part of this may be due to the different time periods and sampling frames. However, part of the difference may be due to the 31 percent non-response rate in the student loan subsample of NPSAS. Dropouts may have been less likely to respond to a household survey. Non-response is not a problem with the Missouri administrative data.

TABLE 1
Default Rates and Dropouts:
Experience of the 1992-93 Missouri Public Two- and Four-Year Cohort

Degree Level	Number of Students Who Borrowed	Percentage of Students Who Borrowed	Average Amount Borrowed	Dollar-Weighted Percentage of Amount Borrowed	Default Rate	Percentage of Borrowers Who Defaulted	Dollar-Weighted Percentage of Defaulted Amount Borrowed
Certificate <2 years	514	0.9%	$6,137	0.6%	20.2%	1.5%	1.1%
Associate's degree	3,427	6.1%	$7,940	4.7%	10.4%	5.0%	4.8%
Bachelor's degree	19,699	34.9%	$11,237	38.5%	4.1%	11.3%	17.6%
Post-baccalaureate (includes post-baccalaureate certificates and masters, doctoral, and first professional degrees)	4,006	7.1%	$18,569	12.9%	2.5%	1.4%	3.9%
No degree	28,726	51.0%	$8,649	43.2%	20.2%	80.8%	72.7%
Total	56,372	100.0%	$10,192	100.0%	12.7%	100.0%	100.0%

loan data is limited to only schools for which the Missouri Coordinating Board for Higher Education is the guarantor. Prior to 1995, the loan data included all Missouri public two- and four-year institutions. In 1995, eight institutions moved to the Federal Direct Student Loan Program, causing data for these schools to be unavailable. As a result, we analyze only loans disbursed prior to 1995.[2]

The sample in Table 2 includes all Fall 1992 first-time, full-time, degree-seeking students who received a loan before 1995 and who were not enrolled in the Fall or Spring semesters.[3] Enrollment persistence was measured by creating a

[2] An analysis of students in schools that stayed in the Federal Family Education Loan Program shows that most students who borrow do so in their first two years at college. In addition, very few students default on a loan taken after their first two years without also defaulting on their initial loans.

[3] These 4,711 individuals are a random 90 percent subsample of the full set of records. Ten percent of the records were excluded for the out-of-sample prediction exercise described later in the article. Information used to establish first-time, full-time, and degree-seeking status is collected from students by each public higher education institution.

series of categorical variables that take the value of one if the student was continuously enrolled or completed a degree within the indicated time frame. For example, the six-semester continuous enrollment indicator takes the value of one if the student was enrolled continuously for *at least* six semesters or completed the degree within six semesters.

Our model is a more complicated specification than earlier studies that simply included a dummy variable indicating whether or not a student eventually graduated or completed a

TABLE 2
Mean Characteristics of 1992-93 First-Time, Full-Time, Degree-Seeking Freshmen at Missouri Two- and Four-Year Public Institutions, by Student Loan Defaulter and Non-Defaulter Status as of December 1999

	All	Defaulters	Non-Defaulters
Default rate			
(Dependent variable)	14.3%	100.0%	0.0%
Enrollment History			
Continuous enrollment or graduation—			
≥2 semesters	78.9%	81.3%	78.5%
≥4 semesters	49.6%	33.0%	52.4%
≥6 semesters	40.7%	17.8%	44.6%
≥8 semesters	29.0%	10.1%	32.1%
Institutional Characteristics			
Enrolled in—			
2-year, community college	17.7%	27.5%	16.1%
4-year, non-selective	33.8%	45.0%	31.9%
4-year, selective	48.5%	27.5%	52.0%
Student Characteristics			
Caucasian	83.9%	68.9%	86.4%
African American	10.6%	26.3%	8.0%
Other minority	5.5%	4.8%	5.6%
Male	44.4%	51.4%	43.3%
ACT - composite score	14.2	9.94	14.9
ACT - missing	33.5%	49.0%	30.9%
Mean age (in 1999)	26.8	28.4	26.6
Time since program separation—			
2 to 6 semesters	42.9%	14.7%	47.5%
7 to 9 semesters	24.6%	21.7%	25.1%
10 to 15 semesters	32.5%	63.6%	27.3%
Number of borrowers	4,711	673	4,038

program. In part this was due to data limitation in other studies: the NPSAS was retrospective, not longitudinal, thus it was not possible to reconstruct the actual enrollment dynamics leading to default. From a policymaker's point of view, panel data is potentially useful because enrollment patterns and degree completion *within specified time panels* are what decision-makers actually observe unfolding in real time. To estimate the effect of continuous enrollment we estimate a logit model of loan defaults. Our logit model is of the following form:

$$\Pr(D=1) = L(X\beta + T\gamma + C_t\delta) \quad L' < 0$$

where D is a vector of dummy variables where the i^{th} element, corresponding to the i^{th} student, takes the value one for those who default on their loans and L(.) denotes the logistic function.[4] X denotes a matrix of institutional and individual control variables. T measures the number of semesters since the students graduated from or were last enrolled in a Missouri public higher education institution. Students who have been in repayment longer will have had a longer time to default on their loans, thus we expect $\gamma > 0$. Our interest is primarily in C_t, a vector of indicator variables that take the value of one if the i^{th} student is continuously enrolled for t periods. Prior research leads us to expect that $\delta < 0$, that is, that students who drop out are at greater risk of default than those who do not. Our panel data allow us to observe enrollment patterns unfold over time.

From a policymaker's point of view, panel data is potentially useful because enrollment patterns and degree completion within specified time panels are what decision-makers actually observe unfolding in real time.

The estimates from our default model are presented in Table 3. Rather than present the estimates of β from the logistic model, which are not readily interpretable, we report exp(β), which is the estimated effect on the odds-ratios for each covariate. Since the overall default rate in the sample is .14, this implies a default odds ratio Pr(D=1) / (1 – Pr(D=1)) of .14/.86 or roughly 1:6 default odds (i.e., one defaulter for every six non-defaulters). If the estimated exp(β) > 1, a unit change in the variable raises the odds of default; if estimated exp(β) < 1 the covariate lowers default odds. For example, if the estimated effect of a covariate is .25, then this variable is associated with a reduction in the default odds at the sample mean to .041 (.25 x .14/ .86) or roughly one default for every 23 non-defaults (1:23). The statistical significance of the underlying estimated β-coefficients is indicated in Table 3 by asterisks.[5]

[4] The logistic function is of the following form: $L(X) = e^X / (1+ e^X)$.

[5] Failure to reject the null hypothesis that the β-coefficient is zero implies an odds-ratio effect of one.

TABLE 3
Odds Ratio Estimates from Logit Regression Model of Student Loan Defaults: Two- and Four-Year College Students

	(1) No Covariates	(2) Full Model	(3) No Covariates	(4) Full Model	(5) No Covariates	(6) Full Model	(7) No Covariates	(8) Full Model
Enrollment History								
Continuous enrollment or graduation—								
≥2 semesters	1.192*	1.162	--	--	--	--	--	--
≥4 semesters	--	--	.447***	.686***	--	--	--	--
≥6 semesters	--	--	--	--	.270***	.476***	--	--
≥8 semesters	--	--	--	--	--	--	.237***	.302***
Institutional Characteristics								
2-year, community	--	--	--	--	--	--	--	--
4-year, non-selective	--	.990	--	.892	--	.853	--	.840
4-year, selective	--	.607***	--	.538***	--	.517***	--	.477***
Student Characteristics								
Caucasian	--	--	--	--	--	--	--	--
African American	--	3.448***	--	3.503***	--	3.451***	--	3.303***
Other minority	--	1.309	--	1.274	--	1.281	--	1.326
Female	--	--	--	--	--	--	--	--
Male	--	1.470***	--	1.442***	--	1.426***	--	1.368***
ACT - composite score	--	.964***	--	.964**	--	.966**	--	.976
ACT - missing	--	.689	--	.687	--	.706	--	.853
Age (in 1999)	--	1.027***	--	1.028***	--	1.031***	--	1.034***
Time since program separation—								
2 to 6 semesters	--	--	--	--	--	--	--	--
7 to 9 semesters	--	2.817***	--	2.765***	--	2.731***	--	3.308***
10 to 15 semesters	--	5.785***	--	4.939***	--	4.059***	--	4.531***
Number of borrowers in sample	4,711	4,711	4,711	4,711	4,711	4,711	4,711	4,711

Significant at .1(*), .05(**), and .01(***)

The variables in Table 3 are grouped by our persistence measure C_t. The first two columns report estimates for a model that includes a control for graduation or two or more semesters of continuous enrollment. Columns three and four report estimates for a model that includes four semesters or more of continuous enrollment or graduation, and so forth. Before examining the persistence coefficients, it is useful to examine the effects of the institutional and individual covariates. The estimates

in column (8) are similar to those in the other columns. The results show that students in non-selective four-year colleges are equally likely to default as community college students, whereas the default odds ratio for four-year students at selective institutions (even after controlling for individual ACT scores) is only 48 percent that of community college students. Thus, selectivity is one institutional characteristic that seems to matter in the sample. However, some researchers have found that after controlling for individual characteristics, institutional characteristics do not matter (e.g., Knapp and Seaks, 1992). A study of default risk on a subsample of NPSAS found no significant effect of school selectivity on default risk (Flint, 1997).

Men are significantly more likely to default than women, and each year of age raises the default odds ratio.

Further, Table 2 also shows that men are significantly more likely to default than women, and each year of age raises the default odds ratio. The table also shows that the default rate for African Americans is significantly higher than for Caucasians and for other racial/ethnic minorities. Other studies have not included measures of student aptitude or ability, an omission noted by Flint (1997). Higher ACT scores are significantly associated with lower defaults in columns (2), (4), and (6), but not (8).

The dummy variables for semesters since program separation show that the window used to assess default matters. The U.S. Department of Education reported cohort default rates based on defaults two years after separation from an institution (GAO, 1999). The coefficients in Table 3 suggest that for our Missouri student cohort, such an approach significantly underestimates actual default rates. The probability of default rises sharply after 2-6 semesters.[6]

The primary focus of this research, however, is on the persistence variables. Continuous enrollment appears to have a very strong association with default rates, larger than any other variable in our study. Column (1) of Table 3 reveals that two semesters of continuous enrollment explain very little about default risk. Beyond two semesters, however, enrollment persistence becomes a very powerful predictor of defaults. At four semesters, the default odds ratio of continuously enrolled students falls to just 45 percent of the rate for those students who have dropped out or otherwise interrupted their education. By six semesters the default odds ratio of the continuously-enrolled student falls to 27 percent that of a non-continuous peer, and by eight semesters it falls further still, to just 24 percent.

[6] We find that the default rate tends to level off after 10 semesters, which is why we used this cutoff for our step function.

When the covariates are added to the model, the basic pattern remains. At four semesters, the odds ratio for continuously enrolled students is 69 percent that of dropouts. This ratio falls to just 30 percent when we compare eight-semester continuously enrolled students to their non-continuous peers.

It is possible that the powerful effect of continuous enrollment may simply be a statistical artifact arising from pooling two- and four-year students in the same sample. Thus, the model is re-estimated on a sample of students enrolled at four-year institutions only. Table 4 shows these results. The odds

TABLE 4
Odds Ratio Estimates from Logit Regression Model of Student Loan Defaults: Four-Year College Students Only

	(1) No Covariates	(2) Full Model	(3) No Covariates	(4) Full Model	(5) No Covariates	(6) Full Model	(7) No Covariates	(8) Full Model
Enrollment History								
Continuous enrollment or graduation—								
≥2 semesters	1.232*	1.206	--	--	--	--	--	--
≥4 semesters	--	--	.391***	.736***	--	--	--	--
≥6 semesters	--	--	--	--	.236***	.534***	--	--
≥8 semesters	--	--	--	--	--	--	.169***	.266***
Institutional Characteristics								
4-year, selective	--	.629***	--	.611***	--	.615***	--	.570***
4-year, non-selective	--	--	--	--	--	--	--	--
Student Characteristics								
Caucasian	--	--	--	--	--	--	--	--
African American	--	3.589***	--	3.654***	--	3.645***	--	3.537***
Other minority	--	1.671**	--	1.607**	--	1.607**	--	1.746
Female	--	--	--	--	--	--	--	--
Male	--	1.524***	--	1.508***	--	1.498***	--	1.430***
ACT - composite score	--	.969*	--	.970	--	.970	--	.982
ACT - missing	--	.628	--	.630	--	.640	--	.816
Age (in 1999)	--	1.072***	--	1.075***	--	1.076***	--	1.075***
Time since program separation—								
2 to 6 semesters	--	--	--	--	--	--	--	--
7 to 9 semesters	--	2.706***	--	2.668***	--	2.649***	--	3.352***
10 to 15 semesters	--	6.277***	--	5.423***	--	4.552***	--	4.769***
Number of borrowers in sample	3,874	3,874	3,874	3,874	3,874	3,874	3,874	3,874

Significant at .1(*), .05(**), and .01(***).

Several studies found that students who graduate or complete a program have lower default rates than those who do not.

ratio coefficients for continuous enrollment are close to those in Table 3. In column (8), for example, the default odds ratio for a four-year student enrolled for eight continuous semesters is just 27 percent that of a similar student who has dropped out.

It may be the case that our coefficients on continuous enrollment or graduation in Tables 3 and 4 simply reflect the effect of graduation. Several studies found that students who graduate or complete a program have lower default rates than those who do not (Flint, 1997; Knapp and Seaks, 1992; Dynarski, 1994). To examine this hypothesis, the basic model is re-estimated distinguishing continuous enrollment and graduation. This new model includes three mutually exclusive dummy variables: continuously enrolled non-graduate, continuously enrolled until graduation, and not continuously enrolled until graduation. For example, if the continuous enrollment dummy refers to four semesters, the set of graduation dummy variables indicate graduation by the fourth semester or earlier.

The odds-ratio effects are shown in Table 5. As in other studies, graduation has a statistically significant and sizeable effect. Students who graduate within eight semesters of

TABLE 5
Odds Ratio Estimates from Logit Regression Model of Student Loan Defaults: Distinguishing Continuous Enrollment and Graduation

	Pooled 2- and 4-Year Colleges			4-Year Colleges Only
	(1)	(2)	(3)	(4)
Continuously enrolled but not graduated				
≥4 semesters	.718***	--	--	--
≥6 semesters	--	.568***	--	--
≥8 semesters	--	--	.669*	.717
Graduated but not continuously enrolled	N/A	N/A	.089***	.026***
Continuously enrolled until graduation	.434***	.318***	.243***	.075***
Number of borrowers in sample	4,711	4,711	4,711	3,874

Note: In addition to the covariates shown, the models included the individual and institutional characteristics reported in Table 4.
Significant at .1(*), .05(**), and .01(***).

enrollment have a default odds ratio much lower than those who do not. However, the continuous enrollment dummy remains statistically significant. The default risk of students who are continuously enrolled for eight semesters, but who have not graduated, is only 67 percent of the risk for non-graduate dropouts. The last row of the table indicates the combined effect of continuous enrollment and graduation. The last column of Table 5 reports estimates for the four-year college sample only.

The estimates in Tables 3 and 5 show a very strong relationship between default risk and continuous enrollment and graduation. However, this does not necessarily mean that continuous enrollment or graduation *cause* lower default rates. The statistical relationship between default rates and continuous enrollment may simply reflect sorting on unobserved variables. For example, it may be that certain personal characteristics, such as effort or responsibility, are associated with a low propensity to default and with enrollment persistence. Then, over time, non-random attrition will produce a sample of continuously enrolled students with above-average effort and responsibility and thus a below-average propensity to default. To the extent that the measured individual covariates fail to control for these personal attributes, the observed relationship simply reflects a sorting of students.[7]

The estimates show a very strong relationship between default risk and continuous enrollment and graduation.

By its very nature, it is difficult to gauge the bias caused by unmeasured variables. However, the effect the measured covariates have on the persistence estimates can be assessed by comparing companion columns in Tables 3 and 4. Controlling for individual covariates reduces the effect of continuous enrollment by roughly one-quarter to one-third. In the full two- and four-year sample in Table 3, at eight semesters the default odds ratio effect is 24 percent without and 30 percent with the covariates. In other words, controlling for a set of six individual characteristics, most of which are significantly related to default risk, only lowers the estimated persistence odds ratio by six percentage points. This suggests that sorting is playing only a modest role in generating these results. It is possible, of course, that there are unmeasured correlates of default risk that would dramatically reduce the effect of persistent

[7] The sorting hypothesis assumes that these unmeasured attributes must not be causally related to the duration of enrollment. However, if college enrollment fosters effort or responsibility, then the causal interpretation still holds. This is an example of the more general statistical problem of heterogeneity and state dependence. See Heckman and Singer (1984).

enrollment. However, even if these unmeasured correlates accounted for one-half of the estimated reduction in the odds ratio, this would still imply that eight semesters of continuous enrollment reduced default risk by 35 percent.

How Well Can We Predict Defaults?

The research results have identified a number of statistically significant correlates of defaults, several of which appear to have quite large effects on the default odds ratios. However, predicting which individuals actually default is difficult. To explore default prediction issues further, the model was used to predict defaults "out of sample." The estimation sample used for Table 3 was a random 90 percent subsample of the full sample. We set aside the remaining 10 percent of observations to test the predictive validity of the model.

Assessing the "goodness of fit" of a logit or probit model, particularly when the sample is unbalanced (i.e., has a share of zeros or ones that is close to one) is not a straightforward matter (Windmeijer, 1995; Greene, 2000). To predict the occurrence of a default with our estimated logit model, it is necessary to choose a cutoff value (P^*) for the fitted probability above which we predict a student default. For example, a naïve model that predicts no one defaults ($P^* = 1$) would correctly classify 87.2 percent of the observations in our prediction subsample. In fact, if we vary the cutoff score P^* so as to maximize the percentage of observations classified correctly, we can do no better than the 87.2 percent correct classification score of the naïve model (see Table 6A).

But the naïve model or the best fit calibration is unlikely to be attractive to policy makers because the cost of false negatives (i.e., predicting an individual will not default when, in fact, he or she does) is quite high relative to the cost of most pre-default interventions (e.g., letters, phone calls, seminars). In fact, it is likely that policy makers would prefer a cutoff set sufficiently low so as to identify many, if not all, true defaulters. Table 6B reports the results of using a P^* cutoff that correctly predicts 90 percent of the defaulters in our estimation sample. Unfortunately, the reduction in false negatives comes at the cost of a higher rate of false positives: 76 percent of borrowers classified as defaulters are not. At first glance, this suggests that our model has little practical value for policy makers. However, given that resources spent on default prevention are costly, it might still be efficient to use the information provided by the statistical model to target resources to those most likely to default, even if the model generates a high rate of false positives.

TABLE 6A
Out-of Sample Default Predictions with a Naïve Model (P*=1, Zero Predicted Defaults)

	True		Total
Predicted	Default (+)	No Default (-)	
Default (+)	0	0	0
No Default (-)	67	456	523
Total	67	456	523

TABLE 6B
Out-of Sample Default Predictions with 90 Percent of True Defaults Predicted in Estimation Sample (P*=.0818)

	True		Total
Predicted	Default (+)	No Default (-)	
Default (+)	63	204	267
No Default (-)	4	252	256
Total	67	456	523

Consider a hypothetical example. Suppose that a particular loan intervention costs $150 and lowers default risk by 10 percent. Further assume that the average default is $8,000. If this intervention is applied to the entire prediction sample, here are the costs and benefits:

Cost: 523 x $150 = $78,450

Benefit: 67 x .1 x $8,000 = $53,600

Benefit/Cost = .683

Alternatively, when the econometric model for eight semesters of continuous enrollment is used to target resources to the most likely defaulters (i.e., those with a $P^* > .0818$, where P^* is chosen in the estimation sample to correctly classify 90 percent of defaulters) it would have these cost/benefit results:

Cost: 267 x $150 = $40,050

Benefit: 63 x .1 x $8,000 = $50,400

Benefit/Cost = 1.258

In other words, using the model to eliminate the predicted negatives (most of which are true negatives), lowers costs by considerably more than it lowers benefits, thus yielding a favorable benefit-cost ratio.

Conclusion

This article presented estimates of a model of student loan defaults based on a rich panel dataset constructed by merging administrative data on student loans, higher education enrollment and performance, and ACT college admissions test scores for a large cohort of first-time, full-time, degree-seeking students. These students entered Missouri two- and four-year public higher education institutions in the Fall 1992 semester and received student loans to help finance their postsecondary education expenses. The students were tracked forward to December 1999, and those who defaulted on student loans were identified.

The model identified several individual characteristics associated with loan defaults; these characteristics were consistent with previously published studies conducted with different samples of borrowers and defaulters. In our sample, men were more likely to default than women, African Americans were more likely to default than Caucasians, and borrowers who attended selective four-year universities were less likely to default than borrowers who attended other types of public two- and four-year colleges, even after controlling for individual ACT scores.

Continuous enrollment or program completion appeared to be the variable with the largest effect on the default odds ratio. Students who were continuously enrolled or who completed their program were far less likely to default than were students who dropped out. Moreover, this result was not driven solely by program completion. Students who did not graduate but were continuously enrolled had a substantially lower probability of defaulting when compared to similar non-graduates with interrupted enrollment spells.

These results are potentially useful for policy makers because enrollment and program completion are variables that are readily observed. Student aid administrators are not omniscient; they cannot know whether a student will eventually graduate. However, they do observe whether a student is enrolled or has completed a program within a specified time interval. This research also highlights a potentially useful role for state higher education agencies. Student aid administrators are typically only aware of whether a student is enrolled at their own institution. However, there is a good deal of mobility of students between institutions, particularly between two- and four-year institutions. Thus, it would be useful for state education agencies to provide information to aid administrators as to the

enrollment or non-enrollment status students who have left their institutions.

The strong association between enrollment status and default risk may reflect sorting (i.e., default-prone students are also more likely to drop out) but may also be causal. The variable with the largest effect on the default odds ratio was continuous enrollment or program completion. Continuous enrollment may change a student's behavior concerning loan repayment. To the extent that the results reflect the latter, they highlight a benefit of institutional programs that encourage student retention. The strength of the association between continuous enrollment and its robustness to the inclusion of student-level covariates that are associated with loan default suggests that at least some, and perhaps most, of the relationship is causal. However, this is clearly a matter for future research.

References

Dynarski, M. (1994). Who defaults on student loans? Findings from the National Postsecondary Student Aid Study. *Economics of Education Review, 13*(1), 55-68.

Flint, T. (1997). Predicting student loan defaults. *Journal of Higher Education, 68*(3), 322-54.

Greene, L. (1989). An economic analysis of student loan default. *Educational Evaluation and Policy Analysis, 11*(1), 61-68.

Greene, W. (2000). *Econometric Analysis.* (4th ed.). New York: Prentice Hall.

Heckman, J. & Singer, B. (1984). Econometric duration analysis. *Journal of Econometrics, 24*, 63-132.

Knapp, L., & Seaks, T. (1992). An analysis of the probability of default on federally guaranteed student loans. *Review of Economics and Statistics, 74*(3), 404-11.

McDaniel, C. (1988). *The Missouri student achievement study: A report of the results of the first two years.* Missouri Coordinating Board for Higher Education. Jefferson City, MO.

Myers, G., & Siera, S. (1980). Development and validation of discriminant analysis models for student loan defaulters. *Journal of Student Financial Aid, 10*(1), 9-17.

Stockham, D., & Hesseldenz, J. (1979). Predicting national direct student loan defaults: Role of personality data. *Research in Higher Education, 10*, 195-205.

U.S General Accounting Office, (2001). *Major management challenges and program risks.* Washington, D.C.: U.S. Government Printing Office (GAO-01-245).

Volkwein, J., Szelest, B., Cabrera, A., & Napierski-Prancl, M. (1998). Factors associated with student loan default among different racial and ethnic groups. *Journal of Higher Education, 69*(2), 206-37.

Wilms, W., Moore, R., & Bolus, R. (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.

Windmeijer, F. (1995). Goodness of fit measures in binary choice models. *Econometric Reviews, 14*, 101-116.

Zimmerman, D. & Miles, B. (1994). Substituting direct government lending for guaranteed student loans: How budget rules distorted economic decision-making. *National Tax Journal, 47*(4), 773-87.