IIMB Management Review (2016) 28, 126–135

available at www.sciencedirect.com

ScienceDirect

journal homepage: www.elsevier.com/locate/iimb

Studying borrower level risk characteristics of education loan in India

Arindam Bandyopadhyay *

National Institute of Bank Management (NIBM), Pune, India

Received 29 March 2014; revised 3 March 2015; accepted 16 June 2016; available online 1 July 2016

KEYWORDS
Credit risk;
Education loan;
Financial institution;
Risk management

Abstract This paper empirically investigates the granular level risk of education loan using a cross section of data from 5000 borrowers obtained from four major public sector banks in India. The findings suggest that education loan defaults are mainly influenced by security, borrower margin, and repayment periods. The presence of guarantor or co-borrower and collateral significantly reduce default loss rates. The socioeconomic characteristics of borrowers and their regional locations also act as important factors associated with education loan defaults. The results suggest that by segmenting borrowers by probability of default and loss given default in a multidimensional scale, banks can adopt better risk mitigation and pricing strategies to resolve borrower problems.

Introduction

Bank credit has an important role in promoting the education and development of skilled professionals required by an emerging market economy such as India. Banks in India provide education loans in the form of term loans. The loans are meant for funding diploma, graduation, and medical (MBBS) studies, post-graduation including technical (Engineering, Computer Science, etc.) and professional courses (Master of Business Administration or MBA, Hotel Management, Aviation, etc.) offered in India and abroad and recognised by the university/government/autonomous institutions such as the Indian Institutes of Technology (IITs) and the Indian Institutes of Management (IIMs). The traditional education loan scheme can provide financial assistance of Rs 1 million for pursuing higher education in India. Similarly, an applicant can get up to Rs 2 million for studies abroad. The interest rate is around 13%, depending upon the amount of the loan. A concession is provided to girl students and also to economically poor students through a central government subsidy scheme. There are no processing fees. Repayment commences one year after completion of the course or six months after securing a job, whichever is earlier. The maturity period of the loan for studies in India (up to Rs 1 million) and studies abroad (Rs 2 million) is 5–7 years. Generally, no security is required for loans up to Rs 400 000. But for loan amounts ranging from Rs 400 000 to Rs 750 000, banks may seek third party guarantee. For loans above Rs 750 000, tangible collateral security of suitable value, along with the assignment of the future income of the student for payment of instalments, is required. The loans for vocational courses are unsecured loans generally in the range of Rs 20 000 to Rs 150 000 for those pursuing courses that have a tenure ranging from 2 to 3 months to 3 years. The moratorium period ranges from six months to one year.

* Tel.: +91 8551032624; fax: +91 2026834478.
E-mail addresses: arindam@nibmindia.org; arindb@rediffmail.com.

http://dx.doi.org/10.1016/j.iimb.2016.06.001
0970-3896

The level of credit risk in education loans has been very high in India in recent years. The system level gross non-performing assets for education loans is around 6% in comparison to 2.6% for its overall retail portfolio during the year 2012–2013. As a result, there has been a sharp decline in the growth rate of education loans in scheduled commercial banks (SCBs) in recent years. Banks are also seeking credit guarantee protection from the government.

The major causes of default of loans are:

Idiosyncratic borrower specific problems: The borrower specific risk arises due to reasons such as repayment problem, collateral risk, academic failure, drop out, personal problems, financial problems (hence lesser capacity to pay), and so on. The rise in the number of institutions also increases employability risk due to an increase in the supply of students. Institutional characteristics or reputation and course category can also determine the employment opportunity.

Systematic (market specific/external) factors: Loan defaults could also occur when students do not get employment/placements due to external risk. Following a financial crisis or a recession, companies may hire fewer employees. Further, lack of quality education that can ensure jobs is also another reason for higher default.

Further, wrong selection of beneficiaries, ineffective follow up of advances, and failure of debt collection machinery in banks have also contributed to non-performing assets (NPAs) in education loans.

Under the Basel II internal rating based approach (IRB) for measuring regulatory credit risk capital, education loans fall into the category of "other retail exposures". The probability of default (PD) and loss given default (LGD) are the most important drivers of credit risk. Though there are numerous studies and publications on retail credit scoring, very few studies have been conducted in an emerging market economy like India to assess education loan default risk by applying econometric models. Boyes, Hoffman, and Low (1989) demonstrate how maximum likelihood estimates of default probabilities can be obtained using the data on borrower specific personal characteristics, economic variables and financial variables from credit card applicants. Altman, Avery, Eisenbeis, and Sinkey (1981) have reviewed the literature on application of statistical techniques in the development of retail score cards. Greene (1992) developed a statistical scoring model for discrete choice and explained its utility in predicting consumer loan default and loan approvals. Fritz, Luxenburger, and Miehe (2007) describe the retail score card development process through a linear combination of several input variables such as socio-demographic information, borrower financial information, account data, collateral characteristics, and external credit history data. Such scoring models are able to predict the future default and survival probability of a customer and thereby could assist in the lending process. Roszbach (2004) developed a bivariate Tobit model to predict future defaults and loan survival time for new retail applicants. Such models allow banks to better predict the risk of a customer and make more realistic evaluations of the returns.

Flint (1997), Knapp and Seaks (1992), Volkwein and Szelest (1995), and Woo (2002) examined the role of individual student background and institution characteristics in predicting student loan default. Gross, Cekic, Hossler, and Hillman (2009) conducted a survey of studies of student loan default. They have summarised several multivariate empirical studies of student loan default conducted between 1991 and 2007 to investigate crucial factors of student loan default. These factors have been categorised under: a) students characteristics; b) institution category (type, area, educational outcomes etc.); c) level of student debt; and d) students' employment and income and total debt position. Lochner, Stinebrickner, and Suleymanoglu (2013) have used survey and administrative data from the Canadian Student Loan Programme and have considered demographic characteristics (age, gender, and aboriginal status), educational background, income and other financial resources in their study. They find that income level, access to savings and family support, educational attainment, and various demographic factors have influence on student loan repayment behaviour. The findings from these studies have guided us in our study in the framing of the hypotheses, the methodology, and the choice of variables.

In this paper, we have studied the micro level risk of education loan using a large set of historical customer level loan data sample from four major public sector banks in India. Our main objective is to study the performance of loans over time and identify key risk factors of such loans across various geographies and constitutions. Accordingly, we have collected data of performing as well as defaulted accounts to compare their behaviours and identify key risk factors. We have examined education loan default as explained by various characteristics associated with the loan (loan amount, interest rate and repayment period), and security positions (margin given, security, etc.). For this, we have employed multivariate statistical techniques to control multiple factors that contribute to default risk. We also check how various borrower characteristics (age, marital status, presence of guarantor/co-borrower, etc.), geographic locations of borrowers (rural, semi-urban, urban, and metro), course related factors (domestic vs. overseas education and placement record) and rating of the education institutes explain risk of default in student loans.

The rest of the paper is structured as follows. In the next section, we describe the empirical methodology that has been adopted in this paper. The third section lays out the data, construction of variables, and descriptive statistics. The fourth section presents the empirical results. This section discusses the key findings of our study and their relevance. The fifth section discusses the main conclusions of the paper.

Empirical methodology

In order to understand the borrower risk pattern, we have performed univariate tests such as mean comparison *t* test and rank sum test on some key target variables (default incidents, default losses, etc.) and across various sub-categories (e.g. rural vs. urban customers, domestic vs. overseas courses and secured vs. unsecured loans). We have also tested pairwise correlation between various test variables (e.g. correlation between loan amount, margin paid, co-borrower income, course type and placement records with incidents of default).

Multivariate analysis—logistic regression estimation for predicting default risk

We have used logistic regression technique to examine how various borrower specific as well as loan specific factors (qualitative as well quantitative) along with situational factors (location and local factors) influence the probability of a loan default.

Logistic regression is a useful tool for analysing data that include binary dependent variables like default and non-default status of an account (coding of NPA vs. not NPA into 1, 0 mode). Logistic regression is non-linear transformation of the linear regression model that determines the probability that the borrower will default because of other causative factors such as absence of security, margin, lack of placement, and so on. However, unlike ordinary least square (OLS), it does not require assumptions about normality. The dependent variable is log odds ratio Default (D) over Non-Default (ND) or logit. Note that *Prob (D) + Prob (ND) = 1*.

The probability of the outcome is measured by the odds of occurrence of an event. If the symbol "P" is the probability of a default event, then (1-P) is the probability of it not occurring.

$$\text{Odds of success} = P/1-P$$

Logit equation: $F(\beta' X_i)$

$$= P(Y_i = 1) = \frac{1}{1+e^{-(\beta_0+\beta_1 X_1+\beta_2 X_2+\ldots+\beta_k X_k+\varepsilon_i)}} \tag{1}$$

P_i is the probability of default of the ith borrower.

So, $E(Y_i|X_i) = P_i = \text{Prob}(Y_i = 1)$.

The joint effects of all explanatory variables put together on the odds are:

$$\text{Odds of default} = \frac{\text{P}}{1-\text{P}} = e^{\beta_0+\beta_1 X_1+\beta_2 X_2+\beta_3 X_3+\ldots+\beta_k X_k.}$$

Taking the logarithms of both sides

$$\ln\left[\frac{\text{Prob}(Default)}{\text{Prob}(Solvency)}\right] = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \ldots + \beta_k X_{k.} + \varepsilon \tag{2}$$

Through regression exercise, we try to find estimates for β_k parameters so that the logistic function best fits the data.

Logistic regression technique is appropriate for estimating the log of the odds of default as a linear function for loan attributes. The risk factors and logic for selection of variables are summarised in Table 1.

Logistic regression uses maximum-likelihood estimation to compute the coefficients for the logistic regression equation. The maximum likelihood (MLE) function that has been optimised with respect to the parameters is defined as:

$$L(\beta) = \ln[l(\beta)] = \sum_{i=1}^{n}\{y_i \ln[P(x_i)] + (1-y_i)\ln[1-P(x_i)]\} \tag{3}$$

The likelihood function is an equation for the joint probability of the observed events as a function of β.

The model likelihood ratio (LR) statistic is:

$$\text{LR}[\text{i}] = [-2\text{LL}(\text{of beginning model})] - [-2\text{LL}(\text{of ending model})]\} \tag{4}$$

The LR statistic is distributed chi-square with k degrees of freedom, where k is the number of independent variables. The pseudo R^2 or McFadden's R^2 measures the degree of fitness of the model.[1]

Multivariate analysis—Tobit regression estimation to predict LGD

We have used Tobit censored regression technique to examine the determinants of education loan LGD in a multivariate analysis framework. The dependent variable is LGD which is loss given default = 1-historical recovery rate (RR). The historical RR is the sum of the cash flow received from defaulted loans divided by total loan amount due at the time of default (EAD). The construction of the variable has been explained in Table 1. The hypotheses have been built based on the literature discussed in the previous section.

Since many loans have zero loss rates, we have censored the dependent variable LGD (at zero level). A Tobit regression performs maximum likelihood technique (MLE) and gives us unbiased estimates of coefficients as compared to OLS technique (Amemiya, 1981; Greene, 1997; Maddala, 1983).

In modelling the LGD rate of a defaulted loan, the Tobit censored regression model (censored from below) is a better suited model as it takes into account bounds on the dependent variable through truncation. It uses a latent variable y* to model boundary cases such that:

$$y_i^* = \beta.x_i + \varepsilon_i \tag{5}$$

$Y_i = y^*$ if $y^* > 0$ & $y_i = 0$ if $y^* \le 0$ where the error term $\varepsilon_i \sim IN(0, \sigma^2)$.

Thus, yi = max (0, y*) for left side censoring. Note that this method does not drop "0" observations because "0" LGD values are also important in the predictive equation.

The likelihood function for the Tobit model can be written as:

$$L(\beta) = \sum_{y_i>0}^{n}\frac{1}{\sigma}f\left(\frac{y_i-\beta x_i}{\sigma}\right)\sum_{y_i\le 0}^{n}\frac{1}{\sigma}F\left(-\frac{\beta x_i}{\sigma}\right) \tag{6}$$

When the likelihood function (mentioned above) has been maximised (MLE method) with respect to β and σ, we get the estimates of the regression parameters (Maddala, 1989; Greene, 1997, pp. 962–966; Wooldridge, 2002, pp. 517–520).

Data, variables and summary statistics

We created a data template to collect borrower level education loan data from 10 participating banks. We selected such

[1] Pseudo R^2 or McFadden's $R^2 = 1 - \frac{LL_M}{LL_0}$, where LL_M is maximum likelihood and LL_o is the initial likelihood ratio.

Table 1 Description of variables used in the study and hypotheses.

Variable	Full form and definition	Argumentation/Logic for selection of variable
DDEF	Default Dummy; DDEF = 1 if account is NPA and DDEF = 0 if it is in standard category	The higher the odd, i.e., Prob(DDEF = 1/DDEF = 0), the higher is the chance of default
LGD	$\text{Loss Given Default} = 1 - \frac{\sum \text{Recoveries}}{\text{EAD}}$ Recoveries = Total post-default principal and interest recovery; Exposure of default (EAD) = Principal outstanding + Interest due at default year	The higher the LGD, the higher is the severity of losses due to borrower default
LOANSIZE	Represents the loan size; it is measured either by taking natural log of loan outstanding or loan limit	Loan size has influence on the default risk
GENDERD	Dummy variable capturing gender information. GENDERD = 1 if the borrower is male & GENDERD = 0 if the borrower is female	This captures borrower character which may explain default behaviour
MARRITALD	Dummy capturing marital status. Dummy = 1, if borrower is married; = 0 if unmarried	Socioeconomic factors influence default risk; older students may have higher financial obligations
LNAGE	Natural log of age capturing the age of the borrower	Capturing borrower specific character that might have influence on default
Borrower margin (BMARGIN)	Captures borrower's own contribution in financing course expenses. It is estimated as follows: $\text{BMARGIN} = 1 - \frac{\text{Limit Sanctioned}}{\text{Course Expenses}}$	The higher the borrower margin, the lower is the probability of default
REPAYMNTH	Repayment period of the loan in number of months	Tenure of the loan explains the default risk of the loan
SECURITYD	Dummy variable capturing whether loan is secured (=1) or un-secured (=0)	Presence of collateral security has risk reduction effect
Presence of guarantee (GUARDUM)	Dummy = 1 if guarantee/co-borrower is attached to the loan, Dummy = 0 if there is no guarantee	Presence of guarantee reduces the default risk as well as loss given default
LNEAD	Log of EAD measures the size of outstanding at default	The greater the default outstanding amount, the higher is the severity of loss
BORRLOCD1_R, BORRLOCD2_SU, BORRLOCD3_U & BORRLOCD4_M	Four location dummies (1, 0 form) representing borrower location in rural, semi-urban, urban and metro areas. These dummies capture the local situational factors.	Local factors have influence on probability of default of a borrower
INDLOCD1_C, INDLOCD2_N, INDLOCD3_E, INDLOCD4_NE, INDLOCD5_W, INDLOCD6_S & INDLOCD7_UT	Seven region dummies (1,0 form) capturing borrower location in different parts of India: Central, North, East, North East, West, South and in Union Territories (UTs)	Regional location factors have influence on likelihood of default
COURSE_CATD1-COURSE_CATD10	Ten course type dummies (1, 0 form) to capture course differences.	Employability and job market success of the courses matter in predicting default of a borrower

banks that had significant exposure to education loan schemes so that data obtained from them were robust and representative of the entire population of such schemes. Amongst these 10 banks that we approached, 6 banks provided us with detailed borrower level data. Finally, we selected four banks from which to draw our sample data as the information given by them was as per our requirements. Amongst them, three are large banks and one is a mid-sized bank. The total business of these banks covers almost 50% of the entire education loan advances in India. These four banks provided us with detailed account level data with the borrower code and characteristics of each applicant (age, gender, education, etc.), the date on which the loan was sanctioned, the loan limit, size of the disbursement, interest rate, tenure, margin given, course information (domestic or overseas), course expenses, placement record (not in all cases), institution rating (record not available in many cases), information about guarantor or co-borrower, the status of each loan (good or bad), post-defaulted recoveries of defaulted loans, and so on.

The sample data of 5000 borrowers are randomly drawn from the loan information data given by these four banks. All these banks have education loan schemes for both higher education and vocational/skill development education and training courses. These banks have significant exposure to education loan schemes and hence the data obtained from them are robust and representative of the entire population of such schemes in India.

Table 1 gives a brief description of the variables used in our micro risk analysis along with an argumentation for the choice of each variable. Table 2 provides a brief description of the sample data used in the study. It gives the profile of the education loan borrowers and their distribution in India. Table 2 shows that 12.17% of the borrowers are default borrowers in the sample data. The average size of the loan limit is Rs 340 000 (approximately) and outstanding is Rs 265 000 with a median size of Rs 186 000.

The median age of the borrowers is 24 years with a standard deviation of 12 years. Location wise, 42.99% of the sample borrowers live in urban India, 22.16% in metros, 18.62% in rural areas and the remaining 16.23% in semi-urban areas. As far as regional distribution of borrowers is concerned, 57.605% of education loan borrowers are concentrated in the southern part of India (in the states of Tamil Nadu, Andhra Pradesh, Kerala, and Karnataka) followed by 16.25% in the north, 11.53% in the west and 10% in the east. We have also seen that the default contributions are higher from urban and rural areas. Our sample data are quite representative of the actual aggregate population distribution of education loans in India.

Furthermore, we have classified the education courses into 10 different categories (e.g., 1: Teaching; 2: Engineering & B tech; 3: Medical,..., and 10: Nursing courses as in Table 3). Most of the education loans are for Engineering (BE and B Tech) courses (51.6%), followed by Master of Business Administration (MBA) and Management courses (14%), other technical and professional courses (8.03%), University post-graduate courses (7.67%) and Teaching courses (5.70%). We have further examined whether course category explains the loan default risk. Table 3 gives additional statistics about course-wise default rates (default proportions). The default rates are highest in Teaching and Training courses (course category 1:

Table 2 Overall sample descriptive statistics.

Variable	N	Mean	Median	Standard deviation	Minimum	Maximum
Loan limit (Rs Lac)	4998	0.3399	0.240	0.3786	0.008	2.30
Loan outstanding (Rs Lac)	4957	0.2654	0.186	0.2947	0.0018	2.702
LNEAD	470	−0.3332	−0.3377	0.991	−4.573	2.5455
DDEF	5000	0.1217	0.00	0.327	0	1
GENDERD (=1, Male; =0, Female)	4918	0.662	1	0.473	0	1
MARRITALD (=1, Married; =0, Other)	4584	0.1555	0	0.3624	0	1
AGE	4875	28	24	12	17	64
LNAGE	4875	3.26	3.18	3.27	2.83	4.16
Repayment months (REPAYMNTH)	4961	107	111	26	12	250
BMARGIN	683	0.2190	0.1483	0.8407	0	0.80
SECURITYD (=1, Secured; =0, not)	4946	0.115	0	0.32	0	1
GUARDUM (=1, with guarantee; 0, not)	2619	0.2146	0	0.41	0	1
BORRLOCD1_R	4941	0.1862	0	0.39	0	1
BORRLOCD2_SU	4941	0.1623	0	0.37	0	1
BORRLOCD3_U	4941	0.4299	0	0.43	0	1
BORRLOCD4_M	4941	0.2216	0	0.49	0	1
INDLOCD1_C	5000	0.027	0	0.163	0	1
INDLOCD2_N	5000	0.1625	0	0.369	0	1
INDLOCD3_E	5000	0.10	0	0.30	0	1
INDLOCD4_NE	5000	0.0168	0	0.1285	0	1
INDLOCD5_W	5000	0.1153	0	0.3195	0	1
INDLOCD6_S	5000	0.57605	1	0.4942	0	1
INDLOCD7_UT	5000	0.0018	0	0.0423	0	1

Units in Rs million, others in numbers.

Table 3 Course-wise default proportion.

Variable: DDEF (Dummy for defaulted borrowers)	N	DDEF = 1 [Dummy for defaulted borrowers] (%)
		Mean
Teaching and Training:COURSE_CATD1	226	49.11
Engineering and B Tech:COURSE_CATD2	2044	7.974
Medical:COURSE_CATD3	164	7.927
Graduation:COURSE_CATD4	189	12.17
Management:COURSE_CATD5	552	11.23
Post-Graduate in Univ:COURSE_CATD6	304	18.42
Other Diploma:COURSE_CATD7	67	25.37
Other and Vocational:COURSE_CATD8	318	11.01
Law and CA:COURSE_CATD9	19	0.00
Nursing:COURSE_CATD10	79	17.72
Total	3962	12.47

Table 4 Comparison between solvent and defaulted education loan borrowers.

Variable	Solvent group		Defaulted group		T test for difference
	Mean	N	Mean	N	
Loan limit in Rs Lac (LOANSIZE)	0.366	4389	0.152	608	0.214*** (13.30)
GENDERD (=1, Male; =0, Female)	0.6685	4330	0.6156	588	0.053*** (2.55)
MARRITALD (=1, Married; =0, Other)	0.1284	4051	0.3621	533	−23.37*** (−14.30)
Borrower age (LNAGE)	3.243	4297	3.435	577	−0.192*** (−13.72)
Repayment months (REPAYMNTH)	108.44	4353	100.07	608	8.36*** (7.52)
Borrower margin (BMARGIN)	0.1901	639	0.1152	41	0.075*** (2.46)
SECURITYD (=1, Secured; =0, not)	0.1173	4339	0.099	607	0.0185* (1.335)

Note: This table shows the outcome of two-sample un-paired t tests with equal variances
*Significant at 5–10%.
** Significant at 5% or better.
***Significant at 1% or better.
Units in Rs million, others in numbers.

49.11%). This number is statistically significant when compared to other courses. This has been confirmed by an unpaired univariate t-test (mean difference 0.389 with t = 17.85). The other high default rates are observed in Diploma courses (course category 7, default rates: 25.37%) and in University post-graduate courses (course category 6, default rates: 18.42%). Defaults are relatively lower in Engineering (course category 2: 7.97%) and Medical courses (course category 3: 7.93%). There are no defaults in Law and professional Chartered Accountancy courses (category 9). This probably captures the employability and job market success of the courses in explaining education loan defaults.

Table 4 compares the profile of defaulted and non-defaulted (solvent) borrowers in our sample data of education loans. There is a statistically significant difference between the two classes of customers. Solvent groups are distinctly safer than their defaulted counterparts. One can notice that average loan size is significantly higher in the solvent group than their defaulted counterparts. The proportion of female (61.56%) and married (36.21%) borrowers is relatively higher in the defaulted group than the solvent group of borrowers. On the other hand, the average borrower margin is significantly higher in case of the solvent group of borrowers (19.01%) in comparison to the defaulted ones (11.52%).

Table 5 describes the interest rates on various sizes of education loans and borrower margin positions. It is evident that the bigger the size of the loan limit, the higher is the borrower margin attached to the loan to reduce the credit risk. The smaller sized loans (less than Rs 400 000 category) are charged with relatively higher interest rate as most of them are unsecured where borrower margins are small. These conclusions are based on their mean and median values.

Table 5 Relationship between loan size, borrower margin and rate of interest.

Loan limit	Margin				Interest rate			
Size bins	N	Mean (%)	Median (%)	SD	N	Mean (%)	Median (%)	SD (%)
<400 thousand	452	14.19	5.9	0.179	3720	14.43	13.75	0.397
400–750 thousand	158	38.06	21.52	1.706	723	13.01	13.5	0.0169
750 k–1 million	23	29.96	27.88	0.181	161	13.14	13.3	0.0176
≥1 million	50	36.86	33.58	0.197	365	13.07	13.3	0.018

Note: The sample size difference occurs depending upon the availability of the observations; k = thousand.

Table 6 Loss experiences of defaulted education loan.

Variable	N	Mean	Median	SD	Min.	Max.
LGD Loss given default	470	0.62696	0.7677	0.3734	0	1
EAD Defaulted loan exposure	470	0.11301	0.07134	0.12758	0.00103	1.275

Units in Rs million, others in numbers.

Table 7 Loss given default (LGD) difference—due to the presence of security and guarantee.

Variable name	Mean	Mean	
	Secured loans (SECURITYD = 1)	Unsecured loans (SECURITYD = 0)	t-Test for difference
Loss given default (LGD)	0.2795	0.6665	−0.387*** (−7.16)
	Presence of Guarantee (DGUAR = 1)	*Absence of Guarantee (DGUAR = 0)*	
Loss given default (LGD)	0.527	0.6463	−0.119*** (−2.63)
No. of Accounts (N)	48	422	

Note: Two-sample t test with equal variances.
** Significant at 5–10%.*
*** Significant at 5% or better.*
****Significant at 1% or better.*

We also obtained detailed loss statistics for student loans (Table 6). The average size of the defaulted loan exposure is Rs 113 000 with a standard deviation of Rs 128 000. The average LGD rate of education loans is 62.70% (in terms of Mean). Loss given default rate estimates the magnitude of likely loss on the defaulted exposure. There is a significant statistical difference in the loss rate between secured and unsecured loans. The mean difference t-test results are reported in Table 7. This suggests that the presence of collateral significantly reduces the losses due to default in education loan. The presence of guarantee also reduces the loss rate due to default (see Table 7 results). Thus, banks can ask for security and guarantee cover as a safety measure to reduce their default losses. We also found that LGD varies across collateral categories depending upon their liquidity and marketability. The average LGD is 25.17% for liquid collaterals (e.g. Life Insurance Corporation (LIC) policies, Kisan Vikas Patra (KVP), deposit, bond, etc.) and 20.42% for property. However, the LGD rate is high if collateral is plant and machinery (LGD = 62.82%) and personal guarantee (LGD = 78.39%).

Empirical results

The primary objective of our multivariate analysis is to demonstrate the importance of borrower specific characteristics as well as institutional and regional factors in deterring the credit risk of education loans in India.

Risk factors in education loan default: multivariate logit regression results

First, we tested the key factors driving default probability in education loan. We ran several sets of logistic regres-

Table 8 Factors determining education loan default: logistic regression results dependent variable: DDEF is a binary dummy variable whose value = 1 if a loan is NPA and value = 0 if loan is standard one.

Independent variables	Regression set 1 coefficients	Regression set 2 coefficients
Loan limit (LOANSIZE)		−0.225** (−2.15)
Borrower age (LNAGE)	1.699*** (12.91)	5.747*** (5.47)
Customer's gender (GENDERD)	−0.3481*** (−3.44)	–
Repayment months (REPAYMNTH)	−0.0203*** (−9.93)	–
Presence of security (SECURITYD)	−0.3930** (−2.28)	–
Borrower margin (BMARGIN)	–	−2.507** (−2.18)
Rural borrower (BORRLOCD1_R)	1.1988*** (7.16)	–
Sub-urban borrower (BORRLOCD2_SU)	−0.3887* (−1.88)	–
Urban borrower (BORRLOCD3_U)	0.5917*** (3.72)	–
Metro borrower (BORRLOCD4_M)@	Dropped	–
Central India (INDLOCD1_C)	−1.896*** (−3.17)	–
North India (INDLOCD2_N)	−1.167*** (−6.22)	–
East India (INDLOCD3_E)	−0.8118* (−4.08)	–
North East India (INDLOCD4_NE)	−2.66*** (−2.63)	–
West India (INDLOCD5_W)	−0.9259*** (−4.46)	–
South India (INDLOCD6_S)$	Dropped	0.871** (2.50)
Union Territory (INDLOCD7-UT)	1.1415 (1.50)	–
Intercept	−5.478*** (−11.18)	−20.097*** (−5.95)
No. of observations (N)	4778	663
LR Chi^2 (d.f.)	543.12 (13)****	51.63 (4)***
Pseudo R^2#	0.155	0.174

Note: z-statistics are reported in the parentheses.
@ and $: These dummies have been dropped due to collinearity.
#Pseudo-R^2 statistic is the McFadden's-R^2.
*Significant at 5–10%.
**Significant at 5% or better.
***Significant at 1% or better.

sions and compared the results with respect to different subsamples provided by the banks. In logistic regression, we compared the borrowers in the standard category (87.83% of the sample) with their defaulted counterparts (12.17% of the total sample). A logit model has the flexibility to incorporate both the qualitative as well as the quantitative factors and is more efficient than the linear regression probability model. The logit regressions predict the probability of an education loan default based on a set of borrower specific factors, and situational factors (location and regional factors) obtained from the account level credit history of 5000 borrowers.

Two sets of regression results are reported in Table 8 which gives evidence of major default risk factors in education loan. The coefficients and t-statistics along with their significance level are presented as evidence. Our multivariate logistic regression results reported in Table 8 provide evidence of key factors affecting education loan default. We predict the odd ratio Prob (Default/Solvency) on various qualitative (gender, borrower location, security, etc.) and quantitative factors (loan size, age, repayment period, margin, etc.).

We find that male borrowers are relatively safer than their female counterparts. The higher the age of the borrower, the higher is the chance of default. In a separate regression, we found that married borrowers are riskier than unmarried ones. The likelihood of default is lower if the loan is secured and the borrower's own contribution (or borrower margin, denoted by BMARGIN) for the course is higher. The longer the repayment period, the lower is the chance of default. These results can be seen in column 2 and column 3 of Table 8 (see the coefficients of SECURITYD and REPAYMNTH).

The regression coefficients measure the influence of these variables on default risk. Estimated values of the coefficients reported in the second and third columns of Table 8 can be used to describe the probability of a borrower defaulting for a unit change in these parameters or a 1% change in the independent variables. Accordingly, using equation 5 we can estimate exp(β) which measures the effect of the independent variable on the "odds ratio". We find that the borrowers with security are 1.5 times more likely to remain solvent than those without security. Similarly, male borrowers are 1.42 times safer than their female counterparts. Similarly, a 10% increase in borrower margin (BMARGIN) decreases the odds of default by: $[\exp(-2.507)-1] \times 10 = (0.082-1) \times 10 = 9.18\%$. This measures the economic significance of variable BMARGIN reported in column 3 of Table 8.

Four locational (rural, urban, etc.) and seven regional (north, east, south, etc.) dummy variables allow the logit model to estimate the default impact among different locations. We have found that rural and urban borrowers are significantly riskier than the metro and semi-urban borrowers. This captures the local situational factors on risk of default. We also tested the effect of regional dummies on likelihood of default. It emerged that borrowers from south India are the riskiest. The borrowers from north, central and west region of India are significantly safer in comparison to borrowers located in the southern part of India.

We also tried to test whether merit and placement records have statistically significant risk reduction effect by introducing intercept dummies as additional variables in our regressions. Though the coefficient signs are negative, they are not statistically significant. In a separate regression (results not reported here), we have found that course category explains defaults. Default chances are higher in B.Ed. and other teacher training courses, whereas it is significantly lower in Engineering, B Tech. and Medical courses. We also found that previous education level explains likelihood of default (i.e., the higher the education level, the lower is the chance of default).

Determinants of LGD–Tobit regression results

We also assessed the factors contributing to education loan default loss rate, which is another important determinant of credit risk. Loss given default predicts losses as a proportion of the outstanding loan, in the event of a borrower going into default. It is an estimate of an amount of money, expressed as a percentage of the exposure amount that cannot be recovered by bank if a borrower defaults.

Loss given default is a key input in the measurement of expected and unexpected credit losses and hence of credit risk regulatory as well as economic capital.

We have run Tobit censored multivariate regression on LGD which investigates key determinants of credit loss (see equation 6). Our Tobit regression results reported in Table 9 reveal some interesting facts. The coefficients reported in the second column of Table 9 capture the level by which dependent variable LGD will be affected by one unit change in independent variable.

Table 9 Tobit regression explaining factors determining loss given default (LGD) of education loan (dependent variable: LGD)

Independent variables	Coefficients
Size of loan outstanding at default (LNEAD)	0.05238**
	(2.57)
Presence of guarantee (GUARD)	−0.1712***
	(−3.26)
Presence of collateral security (SECURITYD)	−0.4708***
	(−6.84)
Repayment months (REPAYMNTH)	−0.004***
	(−4.46)
Intercept	1.066***
	(11.97)
No. of observations (N)	468
LR Chi^2 fitness (d.f)	81.66 (4)***
Pseudo R^2#	0.119

Note: The dependent variable is left censored to zero (91 left-censored obs. at LGD ≤0).
#Pseudo-R^2 statistic is the McFadden's-R^2.
* Significant at 5–10%.
**Significant at 5% or better.
***Significant at 1% or better.

We find that smaller loans have better recovery and hence lower LGD percentage. The presence of guarantor/co-borrower and collateral security increases the chances of loan recovery and hence reduces LGD rates. The coefficients of these two variables in LGD regression are statistically significant. The longer the repayment period, the lower is the loss rate (significant negative effect on LGD).

We also checked whether the study course duration (in number of years) had any impact on LGD. We obtained statistical negative significance of this independent variable on LGD. This means that longer duration courses have lower LGD (coefficient = −0.1136 with t = −5.95).

Concluding discussion

Using a sample data of 5000 randomly selected borrowers from four representative public sector banks in India, we studied the micro level risk of education loan and identified key risk factors of such loans across various geographies and constitutions. Our study reveals some interesting facts about education loan defaults in India. We find that male borrowers are 1.42 times safer than their female counterparts. The higher the age of the borrower, the higher is the chance of default. Married borrowers are riskier than unmarried ones. The likelihood of default is lower if the loan is secured and the borrower's own contribution (or borrower margin) is higher. Borrowers with security are 1.5 times more likely to remain solvent than those without security. As age of the student increases, so does the likelihood of default (due to increasing financial commitments). The likelihood of default is lower if the loan is secured and the borrower's own contribution (or borrower margin) for financing the course is higher. The longer the repayment period, the lower is the chance of default. These results have important implications on loan appraisal and credit monitoring by the banks.

Merit and placement records do not have statistically significant risk reduction effect. This may be due to the few data records available in banks. Previous education level explains likelihood of default (i.e., the higher the education level, the lower is the chance of default). Course category types explain borrower likelihood of defaults. Default chances are higher in B.Ed. and other teacher training courses whereas default chances are significantly lower in Engineering, B.Tech and Medical courses. Rural and urban borrowers are significantly riskier than the metro and semi-urban borrowers. This captures the local situational factors on risk of default. Default rates are significantly higher in the southern part of India and lower in northern India.

We have also assessed the factors contributing to education loan loss given default, which is another important determinant of credit risk. We found that smaller defaulted outstanding loans have better recovery and hence lower loss (LGD) percentage. Thus, the bigger the size of exposure at default, the higher is the severity of loss rate due to default. The presence of guarantor/co-borrower and collateral security increases the chances of loan recovery and hence reduces LGD rates. The longer the repayment period, the lower is the loss rate (significant negative effect on LGD). Longer duration courses have significantly lower LGD.

Our study thus suggests that strengthening credit risk assessment techniques, borrower risk assessment through credit rating, portfolio monitoring, due diligence in lending and institute performance measures can reduce credit risk in education loans. Merit, employability of course, and reputation of institutions should matter in loan appraisal to reduce the default risk. Creating awareness among the borrowers/co-borrowers for repayment of the dues as scheduled and building a repayment culture among the students is also part of the social responsibility for banks. Regular tracking of the student and follow-up may also reduce the risk of default. Employers should be sensitised regarding payment of equated monthly instalments (EMI) of education loan of their employees.

Moreover, by segmenting borrowers by probability of default and loss given default in a multidimensional scale, banks can adopt better loss mitigation and pricing strategy to resolve borrower problems. Borrowers with high probability of default and high loss severity can be segmented from lower credit risk borrowers.

Though the smaller loans are mostly unsecured, for bigger amounts, banks may ask for securities (in the form of fixed deposits (FD), LIC policies and property) and co-applicant as a guarantor to reduce the risk. We recommend that banks use yearly cohort default rates measures (e.g. transition matrix or NPA movements) to track the rating slippages to estimate the portfolio credit risk. This is to be done across regions, course types, institution-wise, and so on, to better understand and monitor portfolio risk. A portfolio approach may enable a bank to better monitor the risky customers and would allow for targeted collection efforts to resolve the default. Banks may prioritise the collection process for high risk accounts earlier in the delinquency cycle. Else, they may opt for credit guarantee protection from the government/private agency.

Acknowledgements

The author would like to thank and acknowledge V S Kaveri, M Manickraj, R Dasgupta, and Elizabeth James for their helpful comments and suggestions. The comments and suggestions of the two anonymous reviewers were extremely helpful. Remaining errors are the author's own.

References

Altman, E., Avery, R., Eisenbeis, R., & Sinkey, J. (1981). Applications of discriminant analysis in business, banking and finance. Greenwich, CT: JAI Press.

Amemiya, T. (1981). Qualitative response models: A survey. *Journal of Economic Literature*, *19*, 1483–1536.

Boyes, W. J., Hoffman, D. L., & Low, S. A. (1989). An econometric analysis of the bank credit scoring problem. *Journal of Econometrics*, *40*, 3–14.

Flint, T. A. (1997). Predicting student loan defaults. *Journal of Higher Education*, *68*(3), 322–354.

Fritz, S., Luxenburger, M., & Miehe, T. (2007). Implementation of an IRB compliant rating system. In M. Ong (Ed.), The Basel II handbook: a guide for financial practitioners (pp. 85–135). London: RISK Books.

Greene, W. H. (1992). A statistical model for credit scoring, working paper, Stern School of Business, New York University.

Greene, W. H. (1997). Econometric analysis (3rd ed.). Prentice Hall.

Gross, J. P. K., Cekic, O., Hossler, D., & Hillman, N. (2009). What matters in student loan default: A review of the research literature. *Journal of Student Financial Aid*, *39*(1), 19–29.

Knapp, L. G., & Seaks, T. G. (1992). An analysis of the probability of loan default on federally guaranteed student loans. *The Review of Economics and Statistics*, *74*(3), 404–411.

Lochner, L., Stinebrickner, T. R., & Suleymanoglu, U. (2013). The Importance of Financial Resources for Student Loan Repayment, NBER working paper no. w19716.

Maddala, G. S. (1983). Limited dependent and qualitative variables in econometrics. Cambridge: Cambridge University Press.

Maddala, G. S. (1989). Introduction to econometrics. Wiley.

Roszbach, K. (2004). Bank lending policy, credit scoring and the survival of loans. *The Review of Economics and Statistics*, *86*(4), 946–958. MIT Press.

Volkwein, J. F., & Szelest, B. P. (1995). Individual and campus characteristics associated with student loan default. *Research in Higher Education*, *36*(1), 41–72.

Woo, J. (2002). Factors affecting the probability of default: Student loans in California. *Journal of Student Financial Aid*, *32*(2), 5–23.

Wooldridge, J. M. (2002). *Econometric analysis of cross section and panel data*. MIT Press.