Using Predictive Modelling to Identify Students at Risk of Poor University Outcomes

Tim Maloney, Economics Department, AUT
Using Integrated Data for Social Sector Policy & Practice
22 February 2016, Wellington

Presentation Overview

- Predictive Risk Modelling (PRM) is used to identify students at risk of course non-completion in the first year and non-retention in the second year at university.

- Two purposes of this study:
    1. To explore the practicality of using PRM to intervene on vulnerable students when they first arrive at university.
    2. To estimate the specific factors that are particularly predictive of these poor early outcomes.

- This analysis utilises conventional administrative data routinely collected as part of the admissions process.

Literature Review

- Extensive empirical literature on the potential determinants of poor university outcomes. See our paper for more details.

- We contribute to this literature in a number of ways:
  1. We have access to a wider range of individual, course and programme characteristics.
  2. We apply formal Predictive Risk Analysis to both course non-completions and student non-retention.

- Maximum likelihood probit analysis is used to estimate the probability of course non-completion in the first year and student non-retention in the second year. We use out-of-sample analysis to validate the effectiveness of this risk tool.

The Data

- Administrative data were taken from 2009 through 2012 on all students enrolled in a bachelors degree programme for the first time.

- Full sample has 18,638 individual and 101,948 course-specific observations.

- The sample is randomly split between equal-sized 'estimation' and 'validation' samples.

- Table 1 provides descriptive statistics on the range of data available for this analysis.

Table 1
Descriptive Statistics and Variable Definitions
Full Sample

Variable	Definition	Mean (std. deviation)
Dependent Variables		
Non-Completion	1 if first-year course is completed; zero otherwise	0.154 (0.361)
Non-Retention	1 if student returns to university in the second year; zero otherwise	0.226 (0.418)
Year of Cohort		
Year 2009	1 if student first enrols in the year 2009; zero otherwise	0.225 (0.478)
Year 2010	1 if student first enrols in the year 2010; zero otherwise	0.253 (0.435)
Year 2011	1 if student first enrols in the year 2011; zero otherwise	0.240 (0.427)
Year 2012	1 if student first enrols in the year 2012; zero otherwise	0.281 (0.450)
Ethnicity		
Asian	1 if student reports ethnicity as Asian; zero otherwise	0.242 (0.428)
European	1 if student reports ethnicity as European; zero otherwise	0.392 (0.488)
Māori	1 if student reports ethnicity as Māori; zero otherwise	0.098 (0.297)
Pacifica	1 if student reports ethnicity as Pacifica; zero otherwise	0.112 (0.316)
Others	1 if student reports other ethnicity; zero otherwise	0.080 (0.272)
Unknown	1 if students reports no ethnicity; zero otherwise	0.076 (0.265)

Country of Origin

New Zealand	1 if student reports New Zealand as country of origin; zero otherwise	0.695 (0.460)
China	1 if student reports China as country of origin; zero otherwise	0.086 (0.280)
India	1 if student reports India as country of origin; zero otherwise	0.016 (0.127)
Korea	1 if student reports Korea as country of origin; zero otherwise	0.022 (0.147)
Vietnam	1 if student reports Vietnam as country of origin; zero otherwise	0.013 (0.113)
Others	1 if student reports other country of origin; zero otherwise	0.155 (0.362)
Unknown	1 if students reports no country of origin; zero otherwise	0.013 (0.111)

Personal Characteristics

Female	1 if student is female; zero if male	0.601 (0.490)
Part-Time	1 if student is enrolled part-time; zero full-time	0.293 (0.455)
Language	1 if student reports a first language; zero otherwise	0.578 (0.494)
English	1 if student reports English as first language; zero otherwise (conditional on reporting a first language)	0.707 (0.455)
Domestic	1 if student receives domestic funding; zero otherwise	0.879 (0.326)
Age	Mean age	22.075 (6.322)

High School Information

Known NCEA Score	1 if NCEA score is available from last year of school; zero otherwise	0.444 (0.497)
Actual NCEA Score	Actual NCEA score (conditional on availability of score)	155.107 (62.860)
Literacy/Numeracy	1 if student took literacy and numeracy test in school; zero otherwise	0.238 (0.426)
School Decile	Mean school decile (conditional on availability of school decile)	6.846 (2.812)

Entrance Type

NCEA Admission	1 if student entered through NCEA level 3; zero otherwise	0.363 (0.481)
Special Admission	1 if student entered through Special Admission category; zero otherwise	0.130 (0.336)
Internal	1 if student entered through pre-degree at this University; zero otherwise	0.089 (0.285)
External	1 if student entered through study at another university; zero otherwise	0.150 (0.358)
Cambridge/IB	1 if student entered through Cambridge or International Bachelaureate; zero otherwise	0.014 (0.117)
Others	1 if student entered through some other category; zero otherwise	0.253 (0.435)

Course Information

Study Hours	Recommended hours of class and preparation time over the semester	180.539 (62.964)
Known Contact	1 if contact hours for the paper are reported	0.844 (0.362)
Contact Hours	Contact hours for the paper (conditional on reporting contact hours)	75.980 (32.234)
Class Size	Average class size in the course	38.279 (28.932)
Course Size	Total number of students enrolled in the course	562.194 (535.464)
Internet Content	1 if course is supported with internet content; zero otherwise	0.588 (0.492)
Level 4	1 if course is at level 4 (pre-degree); zero otherwise	0.005 (0.067)
Level 5	1 if course is at level 5 (first year); zero otherwise	0.836 (0.371)
Level 6	1 if course is at level 6 (second year): zero otherwise	0.156 (0.363)
Level 7	1 if course is at level 7 (third year); zero otherwise	0.004 (0.059)

Individual Academic Information

Number of Courses	Number of courses taken by the student	5.243 (2.301)
High Level	Proportion of level 6 or 7 courses taken by the student	0.138 (0.202)
Double Degree	1 if student is enrolled in a double-degree; zero otherwise	0.008 (0.088)
One Campus	1 if student is taking all courses on a single campus; zero otherwise	0.913 (0.283)

First-Year Programmes of Entry

BA	1 if student enrolled in Bachelor of Arts; zero otherwise	0.078 (0.268)
BBus	1 if student enrolled in Bachelor of Business; zero otherwise	0.282 (0.450)
BCIS	1 if student enrolled in Bachelor of Computer Information Science; zero otherwise	0.049 (0.216)
BCS	1 if student enrolled in Bachelor of Communication Studies; zero otherwise	0.068 (0.250)
Bde	1 if student enrolled in Bachelor of Design	0.074 (0.262)
BEdu	1 if student enrolled in Bachelor of Education	0.040 (0.197)
BEngT	1 if student enrolled in Bachelor of Engineering Technology; zero otherwise	0.029 (0.168)
BHS	1 if student enrolled in Bachelor of Health Science; zero otherwise	0.195 (0.396)
BIHM	1 if student enrolled in Bachelor of International Hospitality Management; zero otherwise	0.043 (0.204)
BMS	1 if student enrolled in Bachelor of Mathematical Science; zero otherwise	0.006 (0.079)
BSR	1 if student enrolled in Bachelor of Sports and Recreation; zero otherwise	0.060 (0.237)
Others	1 if student enrolled in another smaller programme; zero otherwise	0.075 (0.291)

Table 2
Estimated Results from Maximum Likelihood Probit Analysis on
Course Non-Completion and Student Non-Retention
Estimation Subsample

Variable	Course Non-Completion in First Year			Student Non-Retention in Second Year		
	Coefficient	Std. Error	dy/dx	Coefficient	Std. Error	dy/dx
Constant	-0.6693***	0.1148	-	-0.1374	0.2156	-
Year of Cohort						
Year 2009	-0.1069***	0.0217	-2.20%	-0.0134	0.0447	-0.36%
Year 2010	-0.0752***	0.0198	-1.54%	0.1030**	0.0424	2.74%
Year 2011	-0.1725***	0.0202	-3.54%	0.1112***	0.0428	2.96%
Ethnicity						
Asian	-0.1267***	0.0442	-2.60%	-0.1079	0.0901	-2.87%
European	-0.1830***	0.0483	-3.76%	0.0164	0.0984	0.44%
Māori	0.1666***	0.0515	3.42%	0.2200**	0.1064	5.85%
Pacifica	0.3247***	0.0501	6.67%	0.0967	0.1040	2.58%
Others	-0.0417	0.0511	-0.86%	-0.0525	0.1048	-1.40%

Country of Origin						
New Zealand	-0.4828***	0.0630	-9.91%	-0.7431***	0.1258	-19.78%
China	-0.4488***	0.0714	-9.21%	-0.8090***	0.1411	-21.53%
India	-0.4670***	0.0865	-9.59%	-0.7482***	0.1777	-19.91%
Korea	-0.3356***	0.0809	-6.89%	-0.4329***	0.1616	-11.52%
Vietnam	-0.6662***	0.1081	-13.68%	-1.5075***	0.2574	-40.12%
Others	-0.4833***	0.0650	-9.92%	-0.8868***	0.1299	-23.60%
Personal Characteristics						
Female	-0.1328***	0.0165	-2.73%	-0.0583*	0.0343	-1.55%
Part-Time	0.7546***	0.0179	15.49%	0.5935***	0.0466	15.80%
Language	0.0302	0.0250	0.62%	0.0304	0.0523	0.81%
English	0.0057	0.0269	0.12%	-0.0144	0.0568	-0.38%
Domestic	0.1852***	0.0434	3.80%	0.2495***	0.0880	6.64%
Under Age 18	0.1244	0.1098	2.55%	-0.2035	0.2180	-5.42%
Age 18	0.1854***	0.0645	3.81%	-0.1700	0.1243	-4.53%
Age 19	0.2233***	0.0646	4.58%	-0.0735	0.1245	-1.96%
Age 20	0.3473***	0.0649	7.13%	-0.0037	0.1250	-0.10%
Age 21	0.3418***	0.0658	7.02%	0.0469	0.1272	1.25%
Age 22	0.2411***	0.0679	4.95%	-0.0147	0.1323	-0.39%
Age 23	0.2555***	0.0695	5.25%	-0.082	0.1353	-2.18%
Age 24	0.1512**	0.0730	3.10%	-0.0653	0.1421	-1.74%
Age 25	0.1427*	0.0758	2.93%	0.0822	0.1439	2.19%
Ages 26 to 30	0.0764	0.0665	1.57%	0.0286	0.1251	0.76%
Ages 31 to 35	0.0264	0.0730	0.54%	-0.1324	0.1364	-3.52%
Ages 36 to 45	0.0838	0.0720	1.72%	-0.1190	0.1365	-3.17%

High School Information						
Known NCEA Score	0.3703***	0.0344	7.60%	0.1991***	0.0741	5.30%
Actual NCEA Score	-0.0047***	0.0002	-0.10%	-0.0024***	0.0005	-0.06%
Literacy/Numeracy	0.0819***	0.0259	1.68%	0.1404***	0.0473	3.74%
School Decile 1	0.4710***	0.0423	9.67%	0.084	0.0938	2.23%
School Decile 2	0.2370***	0.0419	4.87%	-0.1128	0.0899	-3.00%
School Decile 3	0.1823***	0.0374	3.74%	0.0225	0.0804	0.60%
School Decile 4	0.1674***	0.0326	3.44%	-0.1706**	0.0696	-4.54%
School Decile 5	0.0865**	0.0391	1.78%	-0.0602	0.0812	-1.60%
School Decile 6	0.0740**	0.0374	1.52%	-0.1361*	0.0778	-3.62%
School Decile 7	0.0886***	0.0340	1.82%	-0.1345*	0.0718	-3.58%
School Decile 8	0.1264***	0.0349	2.60%	0.0019	0.0724	0.05%
School Decile 9	0.1415***	0.0323	2.91%	-0.063	0.0664	-1.68%
School Decile 10	0.1607***	0.0289	3.30%	-0.1293**	0.0601	-3.44%
Entrance Type						
NCEA Admission	0.1752***	0.0363	3.60%	-0.0128	0.0768	-0.34%
Special Admission	0.0752***	0.0270	1.54%	-0.0827	0.0559	-2.20%
Internal	-0.0312	0.0310	-0.64%	-0.2004***	0.0653	-5.33%
External	-0.1542***	0.0274	-3.17%	-0.1752***	0.0556	-4.66%
Cambridge/IB	-0.3138***	0.0703	-6.44%	-0.3563**	0.1517	-9.48%

Course Information

Study Hours	0.0030	0.0265	0.06%	-	-	-
Known Contact	-0.0702**	0.0346	-1.44%	-	-	-
Contact Hours	-0.0101	0.0480	-0.21%	-	-	-
Class Size/10	0.0071**	0.0028	0.10%	-	-	-
Course Size/10	-0.0015***	0.0003	-0.03%	-	-	-
Internet Content	0.0187	0.0187	0.38%	-	-	-
Level 4	-0.2765***	0.1049	-5.68%	-	-	-
Level 6	-0.0218	0.0228	-0.45%	-	-	-
Level 7	-0.1097	0.1164	-2.25%	-	-	-

Individual Academic Information

Number of Courses	-	-	-	-0.0010	0.0119	-0.03%
6+ Courses	-	-	-	-0.7818***	0.0963	-20.81%
Double Degree	-0.1524	0.0936	-3.13%	-0.4927**	0.2423	-13.11%
One Campus	-0.0075	0.0247	-0.15%	0.0446	0.0559	1.19%

BA	0.0837***	0.0316	1.72%	0.3280***	0.0748	8.73%
BBus	-0.1791***	0.0469	-3.68%	-0.1242**	0.0719	-3.30%
BCIS	0.0167	0.0372	0.34%	-0.0737	0.0868	-1.96%
BCS	-0.2971***	0.0410	-6.10%	-0.1786**	0.0887	-4.75%
BDe	-0.3306***	0.0397	-6.79%	-0.1221	0.0818	-3.25%
BEdu	-0.5306***	0.0456	-10.89%	-0.2231**	0.0948	-5.94%
BEngT	0.0896*	0.0462	1.84%	-0.1237	0.1064	-3.29%
BHS	-0.3598***	0.0328	-7.39%	-0.033	0.0642	-0.88%
BIHM	-0.2196***	0.0407	-4.51%	-0.0913	0.0930	-2.43%
BMS	0.1758**	0.0727	3.61%	0.1497	0.1930	3.98%
BSR	-0.1145***	0.0366	-2.35%	0.1480*	0.0795	3.94%

Pseudo R^2	0.1339	0.1063
Log-Likelihood	-18,985.6	-4,417.61
Area Under the ROC Curve	0.7553	0.7125
n	50,932	9,301

Table 3
Percentage of Outcomes Correctly Identified
Validation Subsample

	Course Non-Completion in First Year	Student Non-Retention in Second Year
Top 1 Decile (top 10%)	29.25%	23.33%
Top 2 Deciles (top 20%)	47.57%	40.91%
Area Under the ROC Curve	0.7553	0.7125
n	50,932	9,301

- Show that PRM is more predictive of course non-completions than the current tool based on a first-year student survey. Our PRM was 46.5% and 53.4% more 'target effective' for the top decile and top two deciles, respectively.

Overall Conclusions

- Experimented with feasibility of using a Predictive Risk Model to identify students at risk of both course non-completions in the first year and non-retention in the second year at university.

- Found it to be both target and cost effective.

- More work can be done in this area. This includes improving access to data on a number of fronts (e.g., personal and family backgrounds and more refined NCEA results).

- The implementation of this predictive risk tool is the next step. Need to work through the various administrative and ethical issues in this area. Evaluation of any resulting interventions in another step in this project.

Extension #1

- Use improved NCEA data, and restrict sample to those with NCEA Level results. Current 'Rank Score' is positive and statistically significant.

- However, based on an arbitrary index (top 80 credits: 4=excellence, 3=merit, 1=achieved). *If* the goal is to use this NCEA information to predict first-year paper completion outcomes, this Rank Score should be computed differently. Regression analysis supports excellence and merit credits having the same value (double that of achieved credits). Furthermore, all NCEA credits are predictive. Show that a revised Rank Score would greatly improve its predictive power.

- Need to redo this analysis with letter grades from these papers. *If* the goal is to predict first-year paper grades, how would this change the ideal index?

Extension #2

- Emphasis on understanding the ethnic differences in paper completion rates and grade outcomes. Why do Maori and Pasifika students have worse outcomes than other ethnic groups? We'll use modern econometric and decomposition techniques to estimate the extent to which these ethnic gaps can be explained by school deciles, NCEA results, entrance types, gender and age at enrolment, and enrolment characteristics (e.g., part-time vs. full-time study, and choice of degree programmes).

- Test whether the importance of these factors vary by ethnic group. Extensions would look at specific exam results (e.g., the importance of literacy and numeracy results), and the possibility of merging this with the IDI where we could get better measures of school and household background factors.