Using generalised boosting models to evaluate the UCAS tariff

Tim Gill Research Division

Introduction

The Universities and Colleges Admissions Service (UCAS) is a UK-based organisation providing the application process for almost all British universities. The UCAS tariff points system is used by universities to help select students for entry to their courses. Each grade in a qualification has a certain number of UCAS points allocated to it, which are then summed to provide an overall tariff points score for each student. The assumption made is that two students with the same UCAS tariff gained from different qualifications are of the same ability, or have the same potential to achieve at university.

This article uses a statistical technique known as generalised boosting models (GBMs) to evaluate the use of the UCAS tariff as a predictor of degree outcome. GBMs are able to analyse complicated interactions between large numbers of variables (such as UCAS tariff points from different qualifications) to produce more accurate predictions. By running GBMs on a set of data including degree class and UCAS tariff points in different qualifications, it is possible to make predictions about the degree class. If these predictions are no better (or only slightly better) than the predictions from using only the total UCAS tariff score then this would mean the UCAS tariff could not be improved by including the extra information – in other words, the equivalences between qualifications assumed by the UCAS tariff are reasonably accurate.

This investigation was also undertaken for different qualifications separately to see whether any effects found were different.

Data

The data for this research was provided by the Higher Education Statistics Agency (HESA)[1]. The data consisted of all full-time graduates who were 17–19 years old when they started a first degree (expected not to last more than three years) in the academic year 2010/11 in a UK HE institution, and completed it in the academic year 2012/13. Thus students entering for degrees lasting more than three years (e.g., in Medicine, Dentistry, Veterinary Science and in many language or Engineering courses) were excluded. Included in the database was information on the prior qualifications taken by students at Level 3, including type of qualifications, subjects, grades achieved and total UCAS tariff points. Where students re-sat an examination only, the highest grade was kept and only qualifications that were graded with at least a pass were included.

After some initial investigation of the data it was found that for some students the UCAS tariff included in the database did not match the tariff calculated (by the author) from their grades achieved in prior qualifications. This is likely to be because some students achieved grades in other, minor, qualifications that were not included in the prior qualifications database (e.g., Key Skills). These qualifications are likely to have, at most, only a small impact on degree performance, so it was decided to use the UCAS tariff calculated from the grades in the prior qualifications database as the basis for predicting degree performance. To prevent the distribution of tariffs having a very long tail, potentially distorting the analysis, students had their overall tariff capped at 700 (equivalent to 5 grade A*s at A level).

Students whose degree status was 'Classification not applicable' or 'Missing' were excluded from the data, as these students had dropped out or had not yet completed their degree. This meant that the total number of students included in the analysis was 83,468.

Method

The main aim of this research was to compare the accuracy of the predictions of final degree outcomes from students' UCAS tariff scores, with the predictions from a more complex model which takes into account which qualifications were taken and the combination of qualifications taken by students that contribute to the tariff scores. A GBM was used to generate the more complex predictions. Brief instructions for how to use a GBM in the current context now follow. For a more detailed explanation see Elith, Leathwick and Hastie (2008) and Ridgeway (2012):

1. Split the available data on all individuals into a training data set and a test data set. The training data set will be used to build the statistical model and estimate parameters. The test set will be used to evaluate the model and prevent over-fitting. That is, it is used to prevent the statistical model focussing on characteristics of the data that are unlikely to be repeated in future data sets.
2. Make an initial prediction of outcomes for all individuals in the dataset. In the context of this research this might be the overall probability of achieving a First-class honours degree (hereafter called a 'First') amongst all students.
3. Estimate some simple adjustments to the model to improve its predictive power[2]. For the model described here, this involves the use of regression trees. These work by searching for the partition of the data that leads to the greatest increase in predictive power. For example, the model might divide the data three ways; between students with 300 or more UCAS tariff points from A levels, those

1. Source: HESA Student Record 2010/11 and 2012/13. Copyright Higher Education Statistics Agency Limited 2013. HESA cannot accept responsibility for any inferences or conclusions derived from the data by third parties.

2. Predictive power refers to how accurately the model predicts the final degree outcome for each student.

with less than 300 UCAS points from A levels and those not taking any A levels. Then, within each subgroup, the model looks for further partitions of the data that improve the predictive power. The number of partitions allowed within the tree is pre-determined. The prediction within each subgroup then becomes the average outcome within the subgroup. For example, the overall probability of a First might be 0.20; after the first partition the probabilities of a First might be 0.30 for those with 300 or more A level UCAS points, 0.18 for those with fewer than 300 A level UCAS points and 0.15 for those not taking A levels.

4. Partially accept these adjustments to the predictions and update the model. Instead of accepting the adjustments from Step 2 completely, the model will only be adjusted by a fraction of the amount suggested. This fraction is known as the *learning* rate, and its value can be between 0 and 1, but is usually set to 0.01 or below. For example, if the suggested adjustment for those achieving 300 or more A level UCAS points is 0.1 (that is, an increase in probability of a First of 0.1), then if the learning rate is set to 0.01 the model would adjust the prediction by 0.001 (i.e., increase the probability to 0.201). The point of setting the learning rate to be so low is that, even if we are using many predictors, and even if we are considering the differences between small subgroups, it ensures that no individual adjustment results in the overall model predictions matching too closely to currently available data.
5. Return to Step 2 and repeat, using the adjusted predictions, for a specified number of iterations. Thus the model will again search for the best partitions. The number of iterations is pre-determined and is usually in the thousands.
6. Evaluate the number of iterations at which the model had the greatest predictive accuracy. Then apply the adjustments up to and including this iteration to any new data set to make predictions.

GBMs have been shown to improve the accuracy of predictions, compared with other predictive methods such as linear regression or neural networks (see Ridgeway, 2013). There are also specific reasons why the method may be particularly appropriate in the context of this research. Firstly, the models automatically handle missing data, which is useful in this situation with students taking different combinations of qualifications. Secondly, they automatically find the most important interactions between variables, rather than having to run many complex regression models in order to try and determine which interactions are important. Finally, they also have built-in mechanisms to avoid over-fitting of data, which is important when analysing complex data.

The variables included in the GBMs were the overall UCAS tariff points, the total tariff points achieved in each qualification and the mean tariff points achieved in each (relevant) qualification. The qualifications included were A level, Advanced Subsidiary (AS) level, A level (double), AS level (double), A level (9 unit award), Extended Project, International Baccalaureate (IB), BTEC Diploma, Certificate and Award, Oxford, Cambridge and RSA (OCR) National Extended Diploma, Diploma and Certificate, Cambridge Pre-U Certificate, Cambridge Pre-U Global Perspectives and Research (GPR) and Cambridge Pre-U Short Course. For some of the qualifications the mean tariff points score was not included, because all students taking the qualification took the same number of subjects (for example, Extended Project, Cambridge Pre-U GPR) and so it did not add any further information than the total tariff score.

The analysis compared the predictions from the GBM with the predictions from using the UCAS tariff only. Significant improvements in the predictive power from using the GBM would suggest that using the UCAS tariff to predict degree outcomes is not the ideal model. Two different predictions were made using each method: the probability of a student achieving a First; and the probability of achieving at least an Upper Second-class honours degree (hereafter called an 'Upper Second'). The UCAS tariff predictions were generated by a logistic regression with a smoothing spline. This allowed the relationship between predictor and outcome to vary from the standard log function for a logistic regression.

As well as an analysis of all students together, the predictions from the different models for those taking particular qualifications were compared. This was done to give an indication of how well aligned the tariff points are for different qualifications. For a particular qualification, if the predictions from the GBM are much better than those generated by using only the total UCAS tariff, this would suggest that the tariff points are not well aligned because knowledge of the qualification improved the predictive power of the model. This analysis was limited to three qualifications with large numbers of candidates; A levels, BTECs and the IB. Students were classified as follows: those taking only A levels and AS levels were categorised as 'A levels only'; those taking BTEC qualifications only were categorised as 'BTECs only'; those taking the IB were categorised as 'IB only'; all other students were categorised as 'Mixed'.

The reason for using the GBM method was to find out if predictions could be improved by including extra information, such as the different qualifications taken, the grades achieved and the combinations of qualifications. Whilst it would not be plausible to use such complex models in reality, it would be possible to change the tariff equivalencies for different qualifications. An analysis of the accuracy of tariff equivalencies for a number of qualifications is undertaken in a separate report (Gill, 2015).

Results

The first stage in using GBMs is to determine the best model. This involves changing a number of different factors that affect how the model runs; specifically the number of trees, the shrinkage factor and the interaction depth. Essentially, these determine how far the model searches in order to find the best outcome. By increasing the number of trees or the interaction depth the model will either investigate more trees, or more branches within each tree. Reducing the shrinkage factor means that a smaller proportion of the adjustments from each iteration will be applied before the next iteration, so the model updates at a slower rate. Changing these factors may improve the model outcomes, but beyond a certain point the improvements are too small to be worthwhile. A number of different models were run to determine at what level to set these factors to produce a good model within a reasonable time. This led to a selection of a model with 3,000 trees, a shrinkage factor of 0.01 and an interaction depth of 3 for both of the different predictions.

There are a lot of different variables feeding into the GBMs, so it is of interest to look at which of the variables had the most influence on the prediction. Table 1 presents the top 5 variables in order of relative influence, for the probability of a First, whilst Table 2 does the same for the probability of achieving at least an Upper Second:

Table 1: Relative influence of variables in GBM (predicting probability of achieving a First-class degree)

Variable	*Relative influence (%)*
A level mean	65.2
A level total	9.6
AS level mean	8.5
IB total	8.0
AS level total	3.1

Table 2: Relative influence of variables in GBM (predicting probability of achieving at least an Upper Second-class degree [2:1])

Variable	*Relative influence (%)*
A level mean	60.3
A level total	18.2
IB total	8.1
AS level mean	4.3
BTEC Diploma	2.5

Thus, according to the GBM, by far the most important variable in terms of predicting degree outcomes was the A level mean tariff points. This suggests that the current UCAS system, where achievement is based on total UCAS tariff points, could be improved by using a mean points score instead (at least in terms of A levels). The current system apparently over-values the performance of students who perform less well in a larger number of A levels, compared with students doing better in fewer A levels. However, it may be that admissions tutors are aware of this and therefore take account of it when making offers.

This effect is also illustrated in Figures 1 and 2, which show the relationship between the likelihood of achieving a First or at least an Upper Second (as measured by the log of the odds according to the GBM) and the value of the A level mean and A level total variables. The figures demonstrate that there is a fairly good linear relationship between the A level mean variable and the likelihood. However, for the A level total points variable, beyond a certain value the likelihood does not increase as the total increases (and actually falls in Figure 1).

To see whether the GBMs improved the prediction accuracy we compared the prediction of degree performance to the actual outcome, for the model using the UCAS tariff only and for the model using the

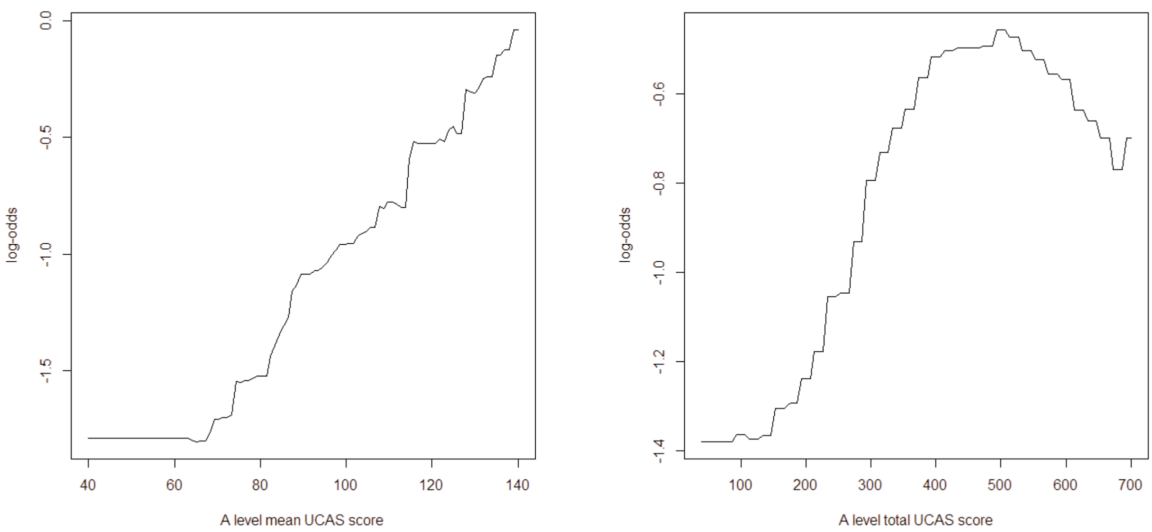

Figure 1: Log odds of achieving a First-class degree, for given values of A level/A level mean

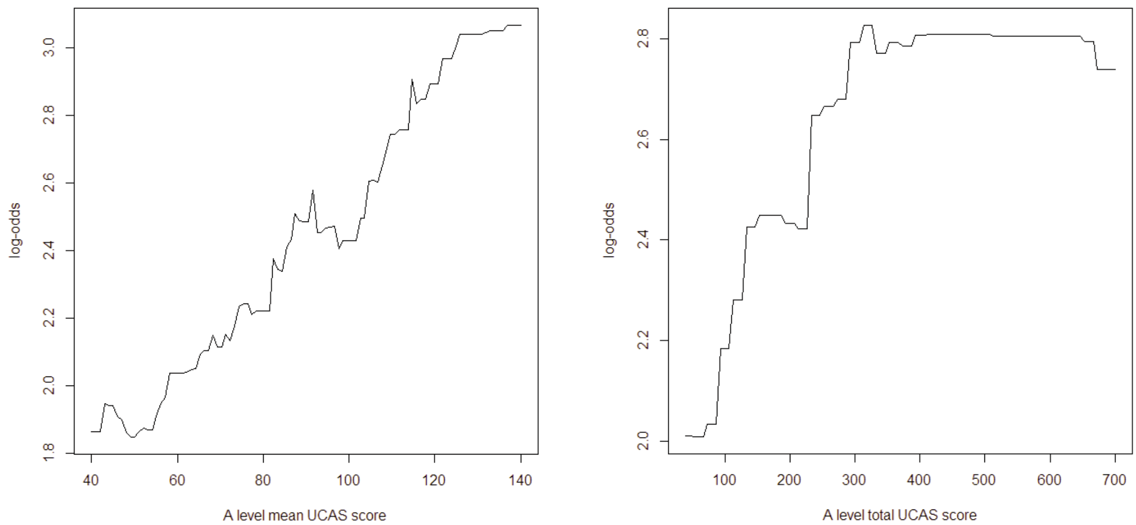

Figure 2: Log odds of achieving at least an Upper Second-class degree, for given values of A level/A level mean

GBM. We used two measures to evaluate how well the models predicted outcomes overall; the correct classification rate and the proportion of deviance explained.

The correct classification rate was calculated as the percentage of candidates where the model prediction of whether they would achieve, for example, a First (that is, whether their predicted probability was above 0.5) matched whether they actually achieved this. This measure is easy to understand but has some weaknesses. For example, this is a binary measure so it doesn't take account of whether a student was very close to being correctly classified (e.g., achieves a First, probability of a First of 0.49) or not (e.g., achieves a First, probability of a First of 0.10).

An alternative measure is the deviance. This is based on the likelihood of students achieving their actual outcome, given the model (it is in fact minus two times the log of this value). So if the model predicts a probability of achieving a First of 0.25 (i.e., unlikely to get a First) for a particular student and they do not achieve this, their likelihood will be 0.75 and their deviance will be -2*log(0.75)=0.575. However, if that student did achieve a First, the likelihood will be 0.25 and their deviance will be -2*log(0.25)=2.77. Therefore, the lower the level of deviance, the better the model is at predicting the outcome. The overall deviance was calculated by summing the deviance across all students. One advantage of using this measure is that different models can be compared, with a lower value indicating a better model fit. The final measure used here to compare different models was the percentage improvement in deviance of each model compared to the 'null' model, which just assigns the overall probability of achieving a First to all students (i.e., a model which is a very poor predictor of outcomes).

Table 3 presents, for the probability of achieving a First-class degree, correct classification rates and proportion of deviance for all students together and then for those taking only the listed qualification(s). It should be noted that using the UCAS tariff prediction, none of the students had a prediction of more than 0.5. Thus the correct classification rate was just the percentage of students who did not get a First (83.68% overall). This was also the case for the GBM prediction for students taking BTECs only or IB only. Thus, this measure tells us very little for these subgroups of students.

Table 3: Comparison of prediction accuracy of UCAS only and GBMs (probability of achieving a First-class degree)

Qualification	*Students*	*Correct classification*		*Proportion of deviance explained*	
		UCAS prediction	*GBM prediction*	*UCAS prediction*	*GBM prediction*
A level only	71,270	83.72	83.78	0.0428	0.0550
BTEC only	3,190	91.19	91.19	0.0713	0.0951
IB only	1,930	78.76	78.76	0.0505	0.0711
Mixed	7,060	81.17	81.49	0.0649	0.0968
All	**83,450**	**83.68**	**83.75**	**0.0458**	**0.0604**

For the analysis of all students together, the improvement in the deviance measure from the model using UCAS tariffs rather than the null model was 0.0458. This was slightly less than the improvement when using GBM (0.0604). Similar differences were found for students taking the separate qualifications, although the difference was greater for students taking a mix of qualifications and for BTEC only students.

The results for the probability of achieving at least an Upper Second are presented in Table 4. Note that neither method predicted any IB only students to get lower than an Upper Second, so the correct classification rate is just the percentage achieving at least an Upper Second (82.38%).

The correct classification rate using the UCAS tariff only was 75.71%, improving to 76.49% using GBM. There was a very small improvement in the correct classification rate for A level students and none at all for IB students. However, for BTEC students the correct classification rate was substantially higher using the GBM prediction (58.94%) than using the UCAS tariff only (52.95%). Using the GBM improves the proportion of deviance explained measure from 0.0789 to 0.1043 overall. For BTEC only students there was a large improvement in this measure, from 0.0731 to 0.1805. There was also a large improvement in this measure for students taking 'Mixed' qualifications, from 0.1132 to 0.1739.

Table 4: Comparison of prediction accuracy of UCAS only and GBMs (probability of achieving at least an Upper Second-class degree)

Qualification	*Students*	*Correct classification*		*Proportion of deviance explained*	
		UCAS prediction	*GBM prediction*	*UCAS prediction*	*GBM prediction*
A level only	71,270	76.43	76.82	0.0759	0.0919
BTEC only	3,190	52.95	58.94	0.0731	0.1805
IB only	1,930	82.38	82.38	0.0819	0.1236
Mixed	7,060	76.89	79.40	0.1132	0.1739
All	**83,450**	**75.71**	**76.49**	**0.0789**	**0.1043**

Conclusion

The research presented in this article has shown evidence that using a GBM to predict degree performance based on attainment in Level 3 qualifications produces more accurate results than using a model based on the overall UCAS tariff only. This is likely to be because the GBM is able to cope better with the complexity of the data, such as the different qualifications and combinations of qualifications taken by students that contribute to the tariff score. It is difficult to assess the size of the improvement in the prediction accuracy because the measure used (proportion of deviance explained) is not easy to interpret. However, it is possible to use this measure to make comparisons between different qualifications in terms of the levels of improvement in prediction accuracy.

Thus, the GBM produced larger improvements in predictive accuracy for students taking BTECs only and for students taking a mix of qualifications, than for students taking A levels or IB. One possible reason for this could be because the current UCAS tariff equivalencies for these qualifications are not well aligned with A level tariffs, and therefore knowledge of the qualifications (and of the combinations of qualifications) taken by students improved the predictions. An assessment of the equivalencies of the UCAS tariff for different qualifications is undertaken in a separate report (Gill, 2015).

It is interesting that the models indicated that the most influential measure in terms of predicting future performance was the A level mean, rather than the A level total score. This is likely to be because of an attenuation effect at the top of the tariff range, where getting higher tariff scores by taking more qualifications is not indicative of higher ability levels (as demonstrated by Figures 1 and 2). For instance, students achieving 5 A* grades at A level (700 UCAS points) are probably not much more able than those achieving 4 A* grades (560 points).

This suggests that the current tariff measure, based on total points score could be improved by taking account of this in some way.

Finally, it is worth considering to what extent admissions tutors (particularly those with many years' experience) are aware of some of these issues and account for them when making offers to students. They may, for instance, take some account of the number of qualifications contributing to a student's UCAS tariff score, or they may value points scores gained from some qualifications more than scores gained from other qualifications. This should go some way to making up for any lack of equivalence between UCAS tariff scores for different qualifications.

References

Elith, J., Leathwick, J.R. & Hastie, T. (2008) A working guide to boosted regression trees, *Journal of Animal Ecology*, *77*(4), 802–813.

Gill, T. (2015). *Assessing the equivalencies of the UCAS tariff for different qualifications*. *Cambridge Assessment internal report*. Cambridge: Cambridge Assessment.

Ridgeway, G. (2012) *Generalized Boosted Models: A guide to the gbm package*. Retrieved from: http://gradientboostedmodels.googlecode.com/git/gbm/inst/doc/gbm.pdf

Post-16 Mathematics qualifications: Differences between GCE A level, International A level, Cambridge Pre-U and Scottish examination questions

Ellie Darlington Research Division

Introduction

This article describes the application of a taxonomy in order to compare and contrast the mathematical skills required to answer examination questions from four different post-16 Mathematics qualifications taken by students both in the UK and overseas: A levels and Advanced Subsidiary (AS) levels, International A and AS levels, Cambridge Pre-U, and Scottish Highers and Advanced Highers. Though the precise content and structure of the different qualifications differ slightly, they are all qualifications which should provide students with a sound basis for university study in Mathematics. All UK universities accept these qualifications as prerequisites for their Mathematics courses. It is therefore of interest to establish whether the questions asked in the assessments of these qualifications require the same kinds of mathematical skills. If there are notable differences among the qualifications, this could suggest that there might be corresponding differences in how well prepared students are for studying Mathematics at university.

In recent years the number of UK schools offering alternative qualifications to General Certificate of Education (GCE) A level has increased. This perhaps may be attributable to head teachers' diminishing confidence in the A level system, with 67 per cent of those surveyed by the Office of Qualifications and Examination Regulation (Ofqual) in 2014 reporting that constant changes to the A level system were of concern. Furthermore, 12 per cent of head teachers surveyed said that they thought that international qualifications such as the International Baccalaureate (IB) and the Cambridge Pre-U were more challenging than A levels. A levels have been criticised for being "oblique at measuring academic ability" (de Waal & Cowen, 2007, p.8), with mathematicians in Higher Education (HE) claiming that it is easy for A level Mathematics students to "'learn the exam' rather than the subject" (Higton et al., 2012, p.58).

Furthermore, concerns are regularly voiced by educational researchers and university admissions and teaching staff regarding the preparedness of new undergraduate mathematicians. For example:

- a restructure of the modular system in A level Mathematics in 2006 resulted in complaints that there was diminishing content (Bassett, Cawston, Thraves, & Truss, 2009; Porkess, 2003, 2006) and that the newer examinations were easier (Qualifications and Curriculum Authority, 2007);
- the modular system of examinations has been criticised for failing to test students' synoptic understanding of Mathematics (Hodgson & Spours, 2004; Quinney, 2008; Wilde, Wrighton, Hayward, Johnson, & Skerrett, 2006);
- some have commented that the A level does not prepare students well for undergraduate Mathematics (Smith, 2004);
- the Engineering and Physical Sciences Research Council (EPSRC) has claimed that "mathematical A-levels are not as rigorous as they used to be." (EPSRC, 2004, p.17);
- the value of the top grade has been questioned, as some stakeholders have claimed that it can be "...achieved through high levels of accuracy rather than extended mathematical reasoning." (Smith, Mitchell, & Grant, 2012, p. 30); and
- claims have been made that standards are falling in the A level, that higher grades are becoming easier to obtain (Coe, 2011; Lawson, 1997).