Predicting quality failures in higher education

Alex Griffiths notes the limitations of a data-driven, risk-based approach for predicting failure

Since 2010 the UK has seen rapid growth in the number of new higher education providers. This growth, aided by reduced barriers to entry to the higher education sector, and concerns over the quality of the new provision it has brought, has been a key driver in successive UK governments pushing for the introduction of a data-driven, risk-based approach to regulating quality in higher education (BIS, 2011; Quality Assessment Review Steering Group, 2015). For the regulator, the Quality Assurance Agency for UK Higher Education (QAA), to prioritize its oversight activity based on freely available performance data has its attractions as high quality providers are allowed to prosper when freed from the burden of unnecessary regulation, and low quality provision is quickly targeted and addressed, and all of this is achieved at a reduced cost to the taxpayer.

A data-driven, risk-based approach, however, relies on one central assumption: that the available data is actually helpful in prioritizing the regulator's activity. Whether or not this is the case has been the focus of an ESRC-funded PhD at King's College London. Our analysis suggests that there is no way to reliably prioritize higher education providers for review despite the wealth of available performance data.

Research design

The research was premised on the fact that we had the outcome of all QAA reviews comparable to today's approach and access to vast amount of historic performance data. This allowed us to investigate whether those providers who were judged 'unsatisfactory' after a review could have been identified in advance using data available at the time. If so, then, in principle, a data-driven, risk-based approach to quality assurance could have been used effectively in the past and our research findings could help inform future risk-based approaches. If it proved impossible to identify high risk providers, even with the benefit of hindsight, our research would suggest that any risk-based approach is unlikely to succeed in the future.

We made use of modern machine-learning techniques to, in effect, try every possible weighted combination of indicators to separately develop the best predictive model for universities, further education colleges and 'alternative' providers. To be as comprehensive as possible we considered not just the indicators in their given form, but also how each provider's performance had changed over time and, where appropriate, standardized indicators by academic year to account for sector-wide shifts in performance over time.

Results

Across all the provider types very few indicators had a strong correlation with the outcome of QAA reviews. Those that did supported the prediction of a small number of 'satisfactory' providers but were of limited use for predicting 'unsatisfactory' providers.

For universities we had 1,700 indicators derived from a wealth of data sources including student surveys, the outcome of previous reviews, complaints raised with the QAA, and staffing, student, research, applications, finance, and overseas activity data.

Figure 1 shows the predicted probability of a university being found 'unsatisfactory' prior to the review, ordered from most to least likely, mimicking the order in which the QAA may be expected to prioritize each university, and the subsequent review finding.

Despite the abundance of data the best model was very poor at predicting the outcome of QAA reviews. Had the QAA carried out their reviews in order of the predicted probabilities, 174 out of the 184 reviews that took place would have been required to discover all 'unsatisfactory' provision and 92.5% of those universities reviewed would have been judged 'satisfactory'. Moreover, with the predicted likelihood of being judged 'unsatisfactory' differing little between universities natural variation in scores would play a large part in the perceived risk posed by each university. Finally, when applied to new data the model produces some questionable results.

The results were similar for further education colleges. The best model required nearly all providers to be prioritized before all of the 'unsatisfactory' provision judgement would have been discovered. However, when the model was tested on new reviews which have taken place since the analysis was conducted, the resulting predictions were worse than chance. The QAA would have be better off doing to the exact opposite of what the model suggested.

Alternative providers offered the greatest promise. There was a clear pattern for younger providers with no prior experience of regulatory reviews and limited funds were significantly more likely to be judged 'unsatisfactory' than more established alternative pro-