Grades across universities over time

GERAINT JOHNES[*]

Lancaster University

KWOK TONG SOO[†]

Lancaster University

June 2015

Abstract

This paper examines the determinants of degree outcomes in a sample of UK universities from 2004 to 2012. We use stochastic frontier methods to account for differences in efficiency across universities and over time. The quality of the student intake and the university's research performance are the main determinants of degree outcomes. There is no evidence of grade inflation at the boundary between upper and lower second class honours. Decomposing the determinants of degree outcomes, we find that good entry grades yield a higher return in traditional universities than in new universities. Although high quality universities award more good degrees, we find little evidence that universities of different quality differ in their propensity to inflate grades over time.

[*] Department of Economics, Lancaster University Management School, Lancaster LA1 4YX, United Kingdom. Tel: +44(0)1524 594215. Email: g.johnes@lancaster.ac.uk
[†] Department of Economics, Lancaster University Management School, Lancaster LA1 4YX, United Kingdom. Tel: +44(0)1524 594418. Email: k.soo@lancaster.ac.uk

1   INTRODUCTION

"Everyone at this graduation is first in his class. What a grand and glorious accomplishment! The diplomas are all here and since every one of you is first in his class, you can simply come up and get them in any old order you please".

Lao She, *Cat Country.*

In any sphere of activity where standards are set, assessors face the challenge of ensuring the constancy of the level of performance that is required to meet the given standard. In certain contexts this is a straightforward task, but in others, where objective measurements are difficult or when the context within which the evaluation is taking place is changing, assessment to a consistent standard is more difficult. In such circumstances, incentives may be present that serve to discourage assessors from being consistent, and the outcome will result from a tension between these incentives and forces working in the opposite direction; this is notably the case where success in meeting the standard might be used as a performance indicator to which reward is attached. In the field of monetary policy, for example, such conditions might lead to price inflation. In the field of education, which is the focus of the present paper, they might lead to grade inflation.

Grade inflation has been a perennial source of concern for policy makers. In the United Kingdom, the Higher Education Academy has recently launched its 'Transforming Assessment in Higher Education' pilot scheme (http://bit.ly/X06tLP) amidst substantial press coverage.[3] This project aims to refine the process of student assessment, thereby compensating for the fact, noted in several studies, that the ability of the current system of degree classification to distinguish between students' performance has declined (Universities UK, 2007, 2012; Higher Education Academy, 2012). This has led the GPA advisory group to call for a national review of degree classifications after a five year period of trialling alternative mechanisms (Grove, 2015). But equally in other countries, the tendency for grade

---

[3] The Sunday Times (22 September 2013) refers to the scheme as an attempt to introduce more refined grade point averages (GPAs) alongside the traditional British degree classification – though it also reports that grade inflation is a concern in countries that use GPAs. More recently, Universities Minister Jo Johnson has announced a Teaching Excellence Framework that will include 'incentives for the sector to tackle degree inflation and ensure that hard-won qualifications hold their value'. He also proposes to provide 'extra granularity through a grade point average' (http://bit.ly/1gaGPjt).

inflation to occur has been noted. In the US, Stuart Rojstaczer has written about this tendency in the Washington Post (28 January 2003), and has subsequently recorded the phenomenon on his website www.gradeinflation.com. He finds that, since the 1980s, average GPA scores across US colleges has risen from below 2.9 to 3.1. In related work, Rojstaczer and Healy (2010) detail how this increase in average scores varies across types of institution.

If the increase in grades serves to compress more and more students into a smaller number of grade classifications, then clearly the process of grade inflation entails a loss of information to users – which may include graduate schools and employers. This, in turn, can result in a deterioration in the quality of employer-employee matches, thereby harming economic efficiency. While grade inflation concerns a nominal variable, it can therefore have real costs.[4]

Yet it is far from clear that the increases in assessment scores that are observed in practice are indeed the result of grade inflation. A number of other factors may be at work, and it is desirable that we should, as far as possible, seek to separate these out. For example, technological advances and improvements in quality assurance procedures may have led to pedagogical improvements in universities that enable them to add more value to their students than was previously the case. Likewise, similar improvements in primary and secondary education may mean that universities can build on students that enter higher education with stronger foundations of knowledge and understanding. Moreover, changes in the efficiency with which students sort themselves across universities may further contribute to rising performance over time. Our goal in this paper is to shed as much light as possible on the factors that underpin the apparent improvements in students' performance over time, and hence to establish the extent to which this is attributable to grade inflation.

The remainder of the paper is structured as follows. In the next section, we present a brief review of the relevant literature. This is followed by a consideration of our methodological approach and the data sources used in the empirical analysis. We then report the results of this analysis, and conclude with a restatement of our findings and some suggestions for future research.

---

[4] Johnson (2003), Pressman (2007) and Babcock (2010) all suggest that a further real cost of grade inflation is that students tend to work less hard, and learn less, when good grades are relatively easy to come by.

2 LITERATURE REVIEW

Marks (2002) argues that norms are public goods. These norms include those that attach to the evaluation of work that students submit for assessment – they apply across whole cohorts of students and are thus non-rival and non-excludable. In common with other public goods, there are incentives for consumers to free ride. In the context of norms, people are generally happy for others to subscribe to standards, but if adherence is costly they are happy to seek ways of circumventing them themselves. Higher education institutions aspiring to be classed among the elite may wish to set high standards in order to enhance their prestige, but at the same time yield to the temptation to manipulate metrics by which they are judged.

McKenzie and Tullock (1981) provide an interesting rationale for this in their analysis of how tuition fees in the US rose during the rapid post-Sputnik expansion of higher education. Institutions that subsequently found difficulty in recruiting sufficient students were unable to reduce tuition fees since their costs had risen in line with their revenues; they had therefore to find a different way of pricing themselves into the market, and increasingly lenient grading practices may have been used to achieve this. Freeman (1999) has tested this idea by examining the relationship between grading standards and labour market prospects in different disciplines. His findings support the hypothesis that grading is more lenient in disciplines where the price of higher education needs to be relatively low in order to compensate for a weak labour market for graduates. Other studies that report that average grades have been rising at different rates in different subject areas include Sabot and Wakeman-Linn (1991), Anglin and Meng (2000), and Jewell et al. (2013). The first of these is particularly interesting since it provides evidence for a 'demand curve' where the strictness of grading practice is regarded as a price.

The tension between the public good nature of norms and the incentive to cheat on those norms has led several authors to view grade inflation in a game theoretic context. Such studies include those by Correa (2001) and Johnes (2004). The latter considers a simple model in which there are two students, each of whose utility is determined by their rank order in grading. So the ith student enjoys utility of $U_i = x_i - x_j, j \neq i$, where x_i denotes the ith student's grade. Suppose that x_i depends on some absolute measure of the value of i's work, x_{0i}, but that it is also influenced by pressure that the students bring to bear on the assessor, so

that $x_i = x_{0i} + \delta_i + \sigma\delta_j$, $j \neq i$, where $0 < \sigma < 1$. Here the δ terms are binary variables reflecting student-specific pressure, and students each choose the value of this variable to maximise their utility. It is readily seen that the Nash equilibrium is one in which each student sets this equal to one, thus creating pressure for grade inflation.

Chan et al. (2007) have developed the idea of grade inflation as a game further, by considering the schools to be the players. Schools signal the quality of their graduates to employers using the grades that they award. This signal is made up in part of a component that reflects students' abilities but in part also of a component that is due to schools wanting to help their own students in the face of competition from other schools. Employers know this, but they do not know how much weight is assigned to each component. It is therefore in schools' interest to inflate grades.

In the same spirit, work by Achen and Thurnau (2009) investigates pressures that university departments face vis-à-vis each other. Where modules form part of the compulsory core of a degree programme, they expect to find relatively low grades; where, on the other hand, the modules are optional, and thus need to compete with other modules for student registrations, they expect grading to be more lenient. Where assessment relies on subjective judgements, grading is expected to be more lenient, because assessors seek thereby to avoid the costs imposed by students who try to query low grades. Finally, departments that are struggling to recruit students are likely to offer more lenient grading. Some support for each of these hypotheses is found using data from the University of Michigan.

The last of Achen and Thurnau's hypotheses might, by extension, be viewed as implying that more prestigious universities are, given their strong position in the market, less likely than others to succumb to the temptations of grade inflation. Yet Popov and Bernhardt (2013) argue otherwise. They develop a model in which good universities respond to a rising demand for highly skilled workers by relaxing the conditions under which they award top grades; they do this because they are seeking to optimise employment outcomes for their graduates, and they can exploit the fact that employers are not able to distinguish between relatively strong and weak students within the top grade. In support of their theory, Popov and Bernhardt provide evidence on grade inflation over a forty year period in a range of US institutions.

Several developments in the technology of educational production may have served to improve performance. For example, Bracey (1994) suggests that the introduction of word-processing software allowed students to improve the presentation of their work. Innovations of this kind are sporadic and one-shot, however, and do not do much to account for a consistent trend over a prolonged period. Nevertheless, technology more generally might evolve gradually in such a way that it contributes to increasing achievement over time, and this raises the question of how best to capture such effects in empirical analyses.

Efforts to evaluate the extent of grade inflation in the US include contributions by Kolevzon (1981) and Compton and Metheny (2000). Both claim to find evidence that grade inflation is indeed present. In the UK there has been some evidence of grade inflation at secondary school level. Exercises that have presented cohorts of students, years apart, with identical problems suggest that while some of the apparent gain in performance is genuine, some has also been illusory (Tymms and Fitz-Gibbon, 2001).

Evidence of grade inflation at the level of higher education has been harder to come by, however. Johnes and McNabb (2002) and Johnes (2004) adopt an educational production function approach, using stochastic frontier analysis to model a measure of the university average of students' grades, using a panel of universities as the data set. The explanatory variables include average student scholastic achievement on entry to university and a variety of other indicators of student demographics. Changes in these variables are likely to account in part for changes, over time, in university average grades. Likewise, changes in the efficiency with which universities convert inputs to outputs, measured by the non-normal component of the residual in the stochastic frontier estimator, likely accounts for some of the change in grades over time. The model specification also includes a full set of time dummies, and the authors argue that changes in the coefficients on these dummies reflect grade inflation – since they cannot be due to changes in input quality or in efficiency. The results suggest that there was some grade inflation over a short period during the mid-1980s, but not in the 1990s.

The empirical work reported in the remainder of this paper updates and builds on the Johnes and McNabb approach. In particular, it allows us to investigate whether there has been any further grade inflation in the UK higher education system since the early part of the last

decade. In addition, we are able to perform a comparison between the conflicting findings of Achen and Thurnau (2009) and Popov and Bernhardt (2013) on the relationship between university quality and grade inflation.

3 DATA AND METHODS

The data used in the empirical analysis that follows come from the Sunday Times University Guide, published annually in September each year. For each higher education institution in the UK they provide information on a variety of variables. These variables include the student-staff ratio, library expenditure, expenditure on other facilities, and research performance (as measured by the Research Assessment Exercise). The data set also contains information about the average performance of students in each institution both on entry (measured by the mean tariff score achieved in the upper secondary school exit qualifications, typically A levels) and on exit (measured by the proportion of students graduating with first or upper second class honours degrees; henceforth, good degrees). All monetary values have been converted into real terms using the consumer price index with base year in 2005. Descriptive statistics for these variables are reported in Table 1, which shows, among other things, that there is substantial variation in the variables used in the analysis, both over time (comparing the 2004 to 2008 sample and the 2009 to 2012 sample), and across universities (comparing the pre-92 universities to the post-92 universities). Figure 1 shows a clear rising trend in the percentage of good degrees in our sample at all points of the distribution (with the possible exception of the upper tail).

[Table 1 here]
[Figure 1 here]

We shall be interested in establishing the relationship between university output – as measured by the proportion of students graduating with good degrees – and input. In order to ensure that the appropriate input variables are used, we study output in a given year as a function of, *inter alia*, the quality of students recruited three years earlier, reflecting the

duration of a typical bachelors programme in the UK.[5] Our data span the period from the 2003-04 academic year to the 2011-12 academic year.

We follow the lead of Johnes and McNabb (2002) and Johnes (2004) in modelling the production of good degrees in a stochastic frontier framework. This allows us to evaluate the effects of (i) changing inputs, including changes in the quality of the student intake over time, (ii) changing efficiency in the conversion of inputs to outputs, reflecting genuine changes in productivity in the higher education sector, and (iii) through the use of time dummies, grade inflation. By using time dummies to capture grade inflation, we are defining grade inflation to be any change in grades that is not explained by changes in the inputs or efficiency[6].

The inputs used in the production function are those discussed above: entry grades, student-staff ratio, library expenditure, expenditure on other facilities, and research performance. We believe these are important inputs in the production of university degrees; Smith and Naylor (2001) and Wossmann (2003) have used similar types of variables in analysing the determinants of undergraduate performance in the UK (Smith and Naylor) and international student performance in science and mathematics (Wossmann)[7]. Similarly, de Paola and Scoppa (2011) show for an Italian public university that larger class sizes have a negative effect on student performance. Our focus is therefore on university characteristics as inputs rather than student characteristics, although some recent evidence (Nicoletti and Rabe, 2013) has shown that family characteristics play an important role in school outcomes[8].

---

[5] There are exceptions. Medicine and architecture, for example, are longer cycle degrees. Some students also undertake industrial placements or spend time studying abroad, both of which have the effect of lengthening their period of study. But, reflecting the fact that there is already a considerable measure of subject specialisation at upper secondary school in England and Wales, three years study for the bachelors degree remains the norm.

[6] Given the level of aggregation in our data we are unable to identify the reason for grade inflation, whether it is more lenient marking, technological change that allows different forms of assessment that enable higher marks, or some other reason. Identifying the reason for grade inflation is clearly an important issue for future research.

[7] Data are also available on the outcomes from the National Student Survey (NSS), which is a measure of student satisfaction with degree provision. This survey started in 2005, so restricts our sample to the 2005-2012 period. Including the proportion of students expressing satisfaction on the overall measure of the NSS as an explanatory variable yields similar results to those reported. However, it may be questioned whether the NSS score is an input into the degree production process, or an output. We therefore do not include the NSS score in the reported results.

[8] The Higher Education Policy Institute (HEPI) has conducted a survey of student experiences in UK higher education beginning in 2006. The survey includes information on students' self-reported private study time, which may be used as an additional input in the production function. However, data availability is limited for the time period we consider.

Stochastic frontier models were introduced into the literature by Aigner et al. (1977), and have been widely used in the economics of education as a means of evaluating cost functions (see, for example, Izadi et al., 2002; Johnes and Johnes, 2009). The stochastic frontier approach is based on the idea that no university can outperform the production frontier; as a result, barring noise, all observations lie within the frontier. To implement this, the error term in the regression model is divided into two components: the classical error term, and a one-sided term which represents inefficiency. Therefore, the stochastic frontier approach is different to conventional regression methods, in which the best-fit line passes through the centre of the data, and does not allow universities to differ in terms of their efficiency[9]. With the benefit of a panel data set, a variety of models become available for analysis (Greene, 2005). The results reported below are obtained using a true random effects specification in which

$$y_{it} = \alpha_i + \beta' \mathbf{X}_{it} + \gamma_t + v_{it} + u_{it} \tag{1}$$

$$v_{it} \sim N[0, \sigma_v^2] \text{ and } u_{it} = |U_{it}| \text{ where } U_{it} \sim N[0, \sigma_u^2]. \tag{2}$$

The dependent variable, y_{it}, is the logged proportion of degrees awarded at first or upper second class honours level, and $\mathbf{X}_{it}$ denotes the vector of logged explanatory variables; in each case the i and t subscripts represent university and year. The γ_t are time dummies; an increase in the value of coefficients of these dummies as t rises would reflect a measure of grade inflation. This specification of the model allows efficiency to be measured by the u_{it} term, but allows institution-specific variation that is due to unobserved heterogeneity other than efficiency differentials to be captured by the α_i. This represents an advance over the previous literature, which is unable to separate efficiency differences from other forms of heterogeneity across institutions (Greene, 2005).

The α_i may be estimated with either fixed effects or random effects used to capture variation across universities. Although the true fixed effects model is desirable in that it allows for university-specific effects to differ across universities, we have made use of the true random effects model because the true fixed effects model suffers from an identification problem.

---

[9] The specifications used in our main table of results (Table 2) are re-run using conventional random effects methods to compare the results. These results are reported in the Appendix, and show that the two different estimation methods yield similar results.

That is, the signal-to-noise ratio ($\lambda = \sigma_u/\sigma_v$) in the true fixed effects model is very large, so that the log-likelihood becomes quite flat, leading to numerical maximisation problems (see Belotti et al 2013). As a result, the parameter estimates from the true fixed effects model are not very stable. In addition, Greene (2005) shows that the predicted efficiency scores from the true fixed effects and true random effects models are very similar to each other.

The half-normal specification of the efficiency term represents a choice on the part of the analyst; u_{it} must be non-normal in order to distinguish it from v_{it}, but a range of possibilities exists, including gamma, exponential and truncated normal. In our experience the choice makes little difference to results, and since half-normal is fairly standard in the literature we choose this option[10]. It is worth noting that the assumption that efficiency follows a distribution that is non-normal is just that – an assumption. In the present analysis, we would expect this assumption to be satisfied as a consequence of the fierce competition that exists between institutions of higher education in the UK; this competition should lead most institutions to be relatively efficient in converting their inputs into good degree results, with a small number of less efficient outliers.

It is possible to restrict the one-sided error term to be constant across all time periods for each institution such that $u_{it} = u_i$. We choose not to impose this restriction, and so we allow flexibility in the model to accommodate changes in efficiency over time for each institution. The u_{it} are recovered using the method of Jondrow et al. (1982) and, given the log specification of the model, may be expressed as efficiency scores using the transformation $\exp(-u_{it})$.

Once the coefficients of the preferred equation are estimated, we conduct a three way Blinder-Oaxaca (1973) decomposition of the gap between groups in the proportion of good degrees awarded *at the frontier*, $y_1 - y_2 - (u_1 - u_2)$, where the subscripts are group indicators. Following Daymont and Andrisani (1984) and Jann (2008), the gap in outcomes between groups may be expressed as a three way decomposition

$$(X_1 - X_2)\beta_2 + (\beta_1 - \beta_2)X_2 + (\beta_1 - \beta_2)(X_1 - X_2) \tag{3}$$

---

[10] We have also performed the analysis assuming an exponential distribution for the efficiency term. The results are similar to those reported in Table 2 using the half-normal distribution, and are available from the authors upon request.

where the calculation is performed at mean values of the explanatory variables. The first term shows the endowment effect: the part of the gap that can be explained by differences in characteristics between groups. The second term shows the coefficients effect: the part of the gap that can be explained by differences in the returns to endowments between groups. The third term shows the interaction effect: the part of the gap that can be explained by observations in group 1 having a higher return on those characteristics for which they have higher means. Where an explanatory variable is a dummy, we normalise the effects relative to the grand mean to make the decomposition invariant to the choice of omitted category (see Yun 2005, Jann 2008). In the analysis section that follows, we provide three way decompositions in order to evaluate the source of the gap in the propensity to award good degrees, first between time periods, and secondly between different types of universities.

We make use of the three way decomposition instead of the more common two way decomposition because the two way decomposition is most often used in the discrimination literature, where the gap between groups is decomposed into an explained component and an unexplained component, the latter of which is usually attributed to discrimination. Whilst not uninformative, we believe that the three way decomposition we perform in this paper provides clearer insight into the determinants of differences between the different groups analysed.

4   RESULTS

4.1   Main results

Estimated coefficients for the model are reported in Table 2. The simplest specification includes as explanatory variables only the average entry grades of students (three years before their cohort graduates) and the time dummies. The coefficient on entry grades is positive but not statistically significant. However, the magnitude of the coefficient is intuitively plausible. An average intake which has three A levels each of grade A has a tariff score that (at midpoint) is 18% higher than an intake which has three A levels each of grade B, so the coefficient of 0.15 implies that a university with the former intake profile will

graduate around 1.6 percent more of its students with good degrees ($0.18 \times 0.15 \times 0.6097 = 0.01646$).

[Table 2 here]

Few of the time dummies are significant. From 2008 onwards, however, there is a rising trend in the coefficients, and these become significantly different from zero in 2011 and 2012, suggesting some grade inflation. Indeed, by 2012, the proportion of good degrees awarded has risen by over 3.7 percentage points[11], other things being equal. While we cannot infer causality, it is interesting to note that this has coincided with a period of intensifying competition in the higher education sector in England, with undergraduate tuition fees rising from around £1000 per year to £3000 from the 2006 entry cohort, and later up to £9000 from the 2012 entry cohort.

In the remaining columns of the table, we add further explanatory variables. In the second column, it is readily observed that research score affects the dependent variable significantly and in an intuitively plausible direction. However, the student-staff ratio, library spending and facilities spending do not have statistically significant effects on the percentage of good degrees. Entry grades are now significantly and positively related to the percentage of good degrees, with a slightly larger coefficient than in column 1. The time dummies are again for most part statistically insignificant. The coefficients are negative in the earlier part of the sample, becoming positive in 2011 and 2012 (albeit not significantly so), reflecting the pattern observed in column 1 of the table. What this suggests is that, because the other explanatory variables are themselves increasing over time, controlling for this increase eliminates the grade inflation result from column 1.

In column (3) of Table 2, we add a set of interaction terms between entry grades and the time dummies. Our intention here is to test the hypothesis that entry grades have become more important over time as a determinant of degree outcomes. This might be the case if sorting of students into degree programmes improves over time, strengthening peer effects and hence reinforcing the positive impact of higher education in institutions that recruit high quality intakes. The increased competitiveness of higher education in England, noted earlier, may

---

[11] This is equal to the coefficient for 2012 multiplied by the sample mean: $0.061 \times 60.97\% = 3.72\%$.

well have contributed to prospective students becoming more selective in their choices of where and what to study, in the sense of choosing courses and universities which are more appropriate to their abilities and interests.[12]

The coefficients on the interaction terms are mostly significantly positive, especially for the years between 2008 and 2011. The size of the coefficient increases over time; entry grades are becoming a stronger and more significant predictor of degree class. This indicates that there is evidence for improved sorting. At the same time the coefficients on the time dummies become significantly negative. This suggests that, after controlling for the improved sorting over time, there is no evidence of grade inflation, or indeed, that there is some evidence of grade *deflation*. Entry grades and research scores remain significantly and positively related to the percentage of good degrees.

Finally, the last column of Table 2 shows the results of the model with all explanatory variables but no time dummies. These results are shown since the Blinder-Oaxaca decomposition between time periods which we perform next does not allow for the inclusion of time dummies. The results excluding the time dummies are broadly comparable to those in column (2) which include them.

Figure 2 plots the distribution of efficiency scores over time, omitting the outliers to highlight the main features. Median efficiency is high in all years (approximately 0.93), and shows no obvious trend. There may be some evidence that the distribution became more compressed between 2008 and 2010; however this was reversed in 2011 and 2012.

[Figure 2 here]

---

[12] Future research may take note of a change in university admissions procedures in 2009 that provides a neat discontinuity and which may have affected the quality of matching between students and institutions. In that year, the Universities and Colleges Admissions Service, which coordinates applications to undergraduate study at British higher education institutions, introduced a system known as 'adjustment', which allows students to switch their choice of university *after* they get their A level results; this option is likely to be pursued by students who perform better than expected, and may lead to an improvement in sorting that was not possible in earlier years. Adjustment may have influenced the results observed in the present paper – the interaction term becomes significant only in the years after its introduction – but it has taken several years for this option to become widely appreciated and exercised. As recently as 2011, only a few hundred students exercised the option, but the numbers have risen rapidly since.

4.2 Blinder-Oaxaca decomposition

In Table 3, we report the results of a variety of three way decompositions. In the first column, we focus on the gap between the degree results variable in the years from 2004 to 2008 and in the years from 2009 to 2012[13]. The gap amounts to 3 percent in favour of the later period. Differences in endowments (especially research and entry grades) more than explain the total gap, while the later period in fact has worse returns to endowments than the earlier period, although this is due to a much lower constant term in the later period. The positive interaction between endowments and coefficients indicates that where the later period has a higher return, it is in variables in which the later period is better-endowed, and is significant for entry grades, library spend, and research score.

[Table 3 here]

In the remaining columns of Table 3 we provide decompositions for various specifications of the model, focusing this time on the gap between pre-1992 universities and other institutions.[14] Column (2) reports results without time dummies, column (3) adds the time dummies, and column (4) includes both time dummies and time-entry grade interactions. In 1992 a large number of higher education institutions were given university status for the first time. These were mainly polytechnics whose mission was focused on teaching, and they contrast sharply with institutions that had university status before 1992 in that the latter have missions that emphasise both teaching and research. While the distinction between pre-1992 and post-1992 institutions has blurred at the edges in the decades since this change, the two mission groups serving research intensive institutions (the Russell Group and the 1994 Group) both comprise only pre-1992 universities; the distinction between pre- and post-1992 institutions therefore remains relevant.

Before moving on to the discussion of the decomposition of the difference in outcomes between pre-92 and post-92 universities, we briefly refer back to Table 1. This reports descriptive statistics of the two time periods under consideration (2004 to 2008, and 2009 to

---

[13] Recall that this is measured on the frontier, so that we are studying a decomposition of $y_1 - y_2 - (u_1 - u_2)$.

[14] Note that the analysis of the gap between time periods in column 1 of Table 3 is, by necessity, conducted using a specification of the frontier model without time dummies. For the analysis of the gap between different types of university, the time dummies can be included; hence results from three different specifications of the model are reported in the remaining columns.

2012), and for the two groups of universities (pre-92 and post-92 universities). It can be seen that all values are higher for the later period than for the earlier period apart from the student-staff ratio which was unchanged. Similarly, pre-92 universities outperform post-92 universities in all the measures, although the differences are larger in some variables than in others. Figure 3 shows the percentage of good degrees over time by pre-92 and post-92 universities, showing that in both cases there is a rising percentage of good degrees over time. Figure 4 shows the efficiency scores of the two groups of universities; pre-92 universities are on average slightly more efficient than post-92 universities, and exhibit smaller variance both across institutions and over time.

[Figure 3 here]
[Figure 4 here]

In column (2) of Table 3 the gap in the proportion of good degrees awarded between the two groups of institutions is pronounced, at nearly 24 percent in favour of pre-92 universities. This is roughly equally divided between the endowments, coefficients, and interaction effects. Among the university characteristics, it is differences in entry grades and research quality which have the largest and most significant effects on the difference in good degrees. However, the really big difference between pre-92 and post-92 universities is in the return to university characteristics. There is a much higher return to entry scores in pre-92 universities, and a much smaller constant term. This is also reflected in the interaction terms: pre-92 universities have higher returns to characteristics in which they are well-endowed (such as entry grades). This suggests that students with good A level scores should attend pre-1992 universities because their talent is rewarded more highly there. This observation is clearly related to the sorting mechanism alluded to earlier.

In columns (3) and (4) of Table 3, the inclusion of the time dummies and interaction terms results in the interaction between endowments and coefficients playing the predominant role in explaining the gap between pre- and post-92 universities. Differences in endowments play almost no role in these specifications. We once again see a large effect of differences in the coefficient on entry scores, as well as on research scores, with pre-92 universities benefitting from higher payoff to both these characteristics than post-92 universities. This comes through in the interaction terms as well: for the characteristics that pre-92 universities excel in (entry grades and research scores), the return to the pre-92 universities is higher than to the post-92

universities. The conclusion is once again that good students do better in pre-92 universities than in post-92 universities.

4.3   Additional results

We next extend the basic model to analyse the conflicting findings of Achen and Thurnau (2009) and Popov and Bernhardt (2013). To recap, Achen and Thurnau find that a stronger market position would lead to less pressure for grade inflation, whereas Popov and Bernhardt argue that it is the better universities that would lower standards to give their students more favourable job market outcomes. We make use of two alternative measures of university quality: an indicator for whether the university is a pre-92 or a post-92 university, and an indicator for whether the university is a member of the Russell Group[15] of leading universities. These indicators are interacted with the year dummies, and the coefficients of the interaction terms will shed light on whether it is higher-quality institutions that suffer more grade inflation over time, or vice versa.

[Table 4 here]

Table 4 reports these results; column (1) for the post-92 indicator, and column (2) for the Russell Group indicator; including both indicators in the same regression yields similar results. Controlling for all the other determinants of degree outcomes, post-92 universities award fewer good degrees than pre-92 universities. The magnitude of this coefficient is large; a post-92 university would typically award 15.4 percent fewer good degrees than a pre-92 university with identical characteristics. This is an important indicator that post-92 universities award fewer good degrees than pre-92 universities beyond the fact that they have fewer inputs into the production of good degrees (see Table 1). There is some suggestion that post-92 universities are more susceptible to grade inflation than pre-92 universities since the coefficients on the interaction terms are increasing over time; however, they are statistically significant only in the earlier part of the sample period, where the coefficients are negative. This suggests that, controlling for university characteristics, post-92 universities had a

---

[15] "The Russell Group represents 24 leading UK universities which are committed to maintaining the very best research, an outstanding teaching and learning experience and unrivalled links with business and the public sector" (The Russell Group, 2013).

tendency to award fewer good degrees than pre-92 universities between 2004 and 2008; however, this tendency has to some extent been reversed in more recent years.

Column (2) of Table 4 reports the same exercise for Russell Group universities. The Russell Group indicator is positive and significant; Russell Group universities award more good degrees than other universities, even after controlling for other inputs into the production of good degrees. A member of the Russell Group would typically award 12.2 percent more good degrees than a non-member with the same characteristics. However, the interactions between the Russell Group indicator and the year dummies are not statistically significant; neither do they show any clear trend.

Care must be taken in interpreting the significant coefficients on the post-92 and Russell Group indicators in Table 4. It may indeed reflect differences in the strictness of grading, but it may also reflect compositional differences in the subjects offered by different types of universities; it may simply be that Russell Group universities have more students in subjects which award a larger proportion of good degrees[16]. Our dataset does not allow us to investigate which of these possible explanations is the more valid, but this is an important issue for future research.

All the foregoing analysis has made use of the percentage of good degrees (first class and upper second class) as our dependent variable. However, as discussed in the Introduction, as the percentage of good degrees has increased, it may cease to be a discriminating measure of good performance. Hence we replicate the specifications in the first three columns of Table 2 using the (natural log of) percentage of first class degrees as the dependent variable. Data for this extension were obtained from the HESA (Higher Education Statistics Agency) Students in Higher Education Institutions database. On average, in our sample, 13.5 percent of degrees awarded are first class degrees, and the correlation between the percentage of first class degrees and the percentage of good degrees is 0.76. Figure 5 shows that the percentage of first class degrees has also been rising over time.

[Figure 5 here]
[Table 5 here]

---

[16] This is a more general point; compositional differences in terms of subject mix occur across all universities, not just between different groups of universities.

The results are reported in Table 5. Column (1) shows that, when we control only for entry grades, there is strong evidence of grade inflation; the coefficients on the year dummies are increasing over time, and highly significant. When we control for other determinants of degree outcomes in column (2), evidence of grade inflation still exists, but is somewhat weaker; the coefficients on the year dummies are smaller than in column (1), albeit still highly significant. In column (3), when we include interactions between entry grades and the year dummies to allow for greater sorting of students by ability, there is little evidence of grade inflation, or of increased sorting. Of the other explanatory variables, entry grades, library spend and research scores are all positively and significantly related to the percentage of first class degrees awarded. Overall, our results using data on the determinants of first class degrees are similar to those using data on good degrees.

5   CONCLUSIONS

Grade inflation grabs headlines. It provides empirical illustration of Goodhart's law – where a measure is used as a performance indicator, it becomes subject to manipulation (Goodhart, 1981). It is also a highly emotive subject – educated people are loath to see their qualifications devalued by subsequent relaxation of standards. However, ascertaining the extent to which grade inflation occurs is far from straightforward.

In this paper we have investigated the issue of whether grade inflation has occurred in higher education in the UK in recent years, using a stochastic frontier approach to take into account different levels of efficiency across universities. The evidence to support the existence of grade inflation is, at best, patchy. The quality of the student intake to universities has typically been rising over this period, and there have been changes in other factors that might reasonably be supposed to affect degree performance too. Moreover, the efficiency with which higher education institutions convert their inputs into outputs has remained fairly constant over time. The intensification of competition in the higher education sector appears to have resulted in an increase in the efficiency of the matching process between students and universities, and there is evidence that this too may have led to an improvement in

outcomes[17]. However, whilst better universities award more good degrees, we find little evidence that different groups of universities exhibit different degrees of grade inflation over time. To the extent that there has been grade inflation over the period covered by our study, this seems to be evident mainly at the boundary between upper second class and first class degrees.

All our results control for entry grades achieved by secondary school leavers – whether those are the result of A-levels or other qualifications. This has an important implication. What we show is that universities do not contribute to grade inflation *over and above* any grade inflation that might affect these entry grades. Clearly, investigating the presence of grade inflation in secondary schools is an important issue for future research.

The emotive nature of the content suggests that ours is unlikely to be the last word on the subject. This is good, not least because changes in the way university admissions are organised mean that there will in future be the opportunity for more rigorous consideration of some of the issues. In particular, as noted earlier, the introduction of adjustment offers the possibility of a discontinuity that could, with a few more years of data, be instructive.

---

[17] However, Dillon and Smith (2015) find little evidence of gains resulting from re-sorting students amongst institutions in the United States.

REFERENCES

Achen, A. and Thurnau, A. F. (2009). 'What are grades made of?' *Journal of Economic Perspectives*, Vol. 23, No. 3, pp. 77-92.

Aigner, D.J., Knox Lovell, C.A. and Schmidt, P. (1977). 'Formulation and estimation of stochastic frontier production function models', *Journal of Econometrics*, Vol. 6, No. 1, pp. 21-37.

Anglin, P.M. and Meng, R. (2000). 'Evidence on grades and grade inflation at Ontario's universities', *Canadian Public Policy*, Vol. 26, No. 3, pp. 361-368.

Babcock, P. (2010). 'Real costs of nominal grade inflation? New evidence from student course evaluations', *Economic Inquiry*, Vol. 48, No. 4, pp. 983-996.

Belotti, F., Daidone, S., Bardi, G. and Atella, V. (2013). 'Stochastic frontier analysis using Stata', *The Stata Journal*, Vol. 13, No. 4, pp. 719-758.

Blinder, A.S. (1973). 'Wage discrimination: Reduced form and structural estimates', *Journal of Human Resources*, Vol. 8, No. 4, pp. 436-455.

Bracey, G.W. (1994). 'Grade inflation', *Phi Delta Kappan*, Vol. 76, No. 4, pp. 328-331.

Chan, W., Hao, L. and Suen, W. (2007). 'A signalling theory of grade inflation', *International Economic Review*, Vol. 48, No. 3, pp. 1065-1090.

Compton, D.M. and Metheny, B. (2000). 'An assessment of grade inflation in higher education', *Perceptual and Motor Skills*, Vol. 90, No. 2, pp. 527-536.

Correa, H. (2001). 'A game theoretic analysis of faculty competition and academic standards', *Higher Education Policy*, Vol. 14, No. 2, pp. 175-182.

Daymont, T.N. and Andrisani, P.J. (1984). 'Job preferences, college major and the gender gap in earnings', *Journal of Human Resources*, Vol. 19, No. 3, pp. 408-428.

De Paola, M. and Scoppa, V. (2011), 'The effects of class size on the achievement of college students', *The Manchester School*, Vol. 79, No. 6, pp. 1061-1079.

Dillon, E.W. and Smith, J.A. (2015). 'The consequences of academic match between students and colleges', IZA Discussion Paper No. 9080.

Freeman, D.G. (1999). 'Grade divergence as a market outcome', *Journal of Economic Education*, Vol. 30, No. 4, pp. 344-351.

Goodhart, C.A.E. (1981). 'Problems of monetary management: the UK experience', in A.S. Courakis (ed.) *Inflation, depression and economic policy in the West*, Washington: Rowman and Littlefield, pp. 111-146.

Greene, W. (2005). 'Reconsidering heterogeneity in panel data estimators of the stochastic frontier model', *Journal of Econometrics*, Vol. 126, No. 2, pp. 269-303.

Grove, J. (2015). 'Rising interest shown in grade point average degree classification trial', *Times Higher Education*, 28 May.

Higher Education Academy (2012). *A marked improvement: transforming assessment in higher education*, York: Higher Education Academy.

Izadi, H., Johnes, G., Oskrochi, R. and Crouchley, R. (2002). 'Stochastic frontier estimation of a CES cost function: the case of higher education in Britain', *Economics of Education Review*, Vol. 21, No. 1, pp. 63-71,

Jann, B. (2008). 'The Blinder-Oaxaca decomposition for linear regression models', *The Stata Journal*, Vol. 8, No. 4, pp. 453-479.

Jewell, R.T., McPherson, M.A. and Tieslau, M.A. (2013). 'Whose fault is it? Assigning blame for grade inflation in higher education', *Applied Economics*, Vol. 45, No. 9, pp. 1185-1200.

Johnes, G. (2004). 'Standards and grade inflation', in G. Johnes and J. Johnes (eds) *International Handbook on the Economics of Education*, Cheltenham: Edward Elgar, pp. 462-483.

Johnes, G. and Johnes, J. (2009). 'Higher education institutions' costs and efficiency: taking the decomposition a further step', *Economics of Education Review*, Vol. 28, No. 1, pp. 107-113.

Johnes, G. and McNabb, R. (2002). 'Academic standards in UK universities: more for less or less for more?' Paper presented at the Royal Economic Society Conference, 27 March.

Johnson, V.E. (2003). *Grade inflation: a crisis in college education*, New York, Springer-Verlag.

Jondrow, J., Knox Lovell, C.A., Materou, I.S. and Schmidt, P. (1982). 'On the estimation of technical efficiency in the stochastic frontier production function model', *Journal of Econometrics*, Vol. 19, No. 2-3, pp. 233-238.

Kolevzon, M.S. (1981). 'Grade inflation in higher education: a comparative study', *Research in Higher Education*, Vol. 15, No. 3, pp. 195-212.

Marks, D. (2002). 'Academic standards as public goods and varieties of free-rider behaviour', *Education Economics*, Vol. 10, No. 2, pp. 145-163.

McKenzie, R.B. and Tullock, G. (1981). *The new world of economics*, Homewood: Irwin.

Nicoletti, C. and Rabe, B. (2013). 'Inequality in pupils' test scores: How much do family, sibling type and neighbourhood matter?' *Economica*, Vol. 80, No. 318, pp. 197-218.

Oaxaca, R. (1973). 'Male-female wage differentials in urban labor markets', *International Economic Review*, Vol. 14, No. 3, pp. 693-709.

Popov, S.V. and Bernhardt, D. (2013). 'University competition, grading standards and grade inflation', *Economic Inquiry*, Vol. 51, No. 3, pp. 1764-1778.

Pressman, S. (2007). 'The economics of grade inflation', *Challenge*, Vol. 50, No. 5, pp. 93-102.

Rojstaczer, S. and Healy, C. (2010). 'Grading in American colleges and universities', *Teachers College Record*, ID15928.

Sabot, R. and Wakeman-Linn, J. (1991). 'Grade inflation and course choice', *Journal of Economic Perspectives*, Vol. 5, No. 1, pp. 159-170.

Smith, J. and Naylor, R. (2001). 'Determinants of degree performance in UK universities: a statistical analysis of the 1993 student cohort', *Oxford Bulletin of Economics and Statistics*, Vol. 63, No. 1, pp. 29-60.

The Russell Group (2013), 'The Russell Group' [online]. Available at http://www.russellgroup.ac.uk/ [Accessed 9 January 2014].

Tymms, P. and Fitz-Gibbon, C. (2001). 'Standards, achievement and educational performance: a cause for celebration?' in R. Phillips and J. Furlong (eds.) *Education, reform and the state: 25 years of politics, policy and practice*, London: Routledge Falmer, pp. 156-173.

Universities UK (2007). *Beyond the honours degree classification: Burgess Group final report*. London: Universities UK.

Universities UK (2012). *Bringing it all together: Introducing the HEAR*, London: Universities UK.

Wossmann, L. (2003). 'Schooling resources, educational institutions and student performance: the international evidence', *Oxford Bulletin of Economics and Statistics*, Vol. 65, No. 2, pp. 117-170.

Yun, M.-S. (2005). 'A simple solution to the identification problem in detailed wage decompositions', *Economic Inquiry*, Vol. 43, No. 4, pp. 766-772.

TABLE 1
DESCRIPTIVE STATISTICS.

Panel A: Full Sample (N = 968)

Variable	Mean	Std. Dev.		
% good degrees	60.97	10.854		
Entry grades	292.3	76.78		
Student-staff ratio	17.50	3.50		
Library spend	747.7	316.1		
Facilities spend	263.9	114.5		
Research score	48.54	22.12		

Panel B: Two time periods (2004-2008), and (2009-2012)

	2004-2008 (N = 518)		2009-2012 (N = 450)	
Variable	Mean	Std. Dev.	Mean	Std. Dev.
% good degrees	59.41	10.73	62.76	10.73
Entry grades	278.8	72.43	307.87	78.76
Student-staff ratio	17.57	3.68	17.41	3.280
Library spend	659.6	272.1	849.1	332.6
Facilities spend	226.3	94.03	307.1	120.5
Research score	41.13	26.66	57.08	9.941

Panel C: Pre-92 and Post-92 universities

	Pre-92 universities (N = 509)		Post-92 universities (N = 459)	
Variable	Mean	Std. Dev.	Mean	Std. Dev.
% good degrees	68.08	9.045	53.08	6.282
Entry grades	346.16	63.72	232.6	33.89
Student-staff ratio	15.45	2.930	19.76	2.554
Library spend	862.5	359.2	620.3	192.1
Facilities spend	294.8	108.6	229.6	111.0
Research score	64.11	11.34	31.29	17.96

Notes: Entry grades are based on the UCAS points system in which an English pre-university qualification with three or four subjects is awarded 120 points for an A, 100 points for a B, 80 points for a C, 60 points for a D, and 40 points for an E. Library and facilities spend have been converted into real terms using the Consumer Price Index with base year in 2005. The research score is obtained from the Research Assessment Exercise and expressed as a percentage.

TABLE 2
RESULTS OF THE TRUE RANDOM EFFECTS MODEL.

Dep var: ln(% good degree)	(1)	(2)	(3)	(4)
Ln(Entry grades)	0.148	0.186	0.108	0.228
	(0.121)	(0.027)***	(0.051)**	(0.063)***
Ln(Student-staff ratio)		-0.041	-0.056	-0.043
		(0.099)	(0.031)*	(0.029)
Ln(library spend)		0.038	0.038	0.029
		(0.051)	(0.017)**	(0.016)*
Ln(facilities spend)		0.002	0.001	0.007
		(0.015)	(0.012)	(0.014)
Ln(research score)		0.023	0.050	0.034
		(0.009)***	(0.022)**	(0.008)***
2005	-0.003	-0.003	-0.240	
	(0.006)	(0.007)	(0.132)*	
2006	-0.012	-0.014	-0.270	
	(0.007)*	(0.008)*	(0.184)	
2007	-0.013	-0.027	-0.414	
	(0.012)	(0.017)	(0.215)*	
2008	-0.010	-0.026	-0.458	
	(0.015)	(0.016)*	(0.223)**	
2009	0.009	-0.022	-0.881	
	(0.016)	(0.019)	(0.338)***	
2010	0.019	-0.016	-0.913	
	(0.018)	(0.021)	(0.380)**	
2011	0.036	0.004	-0.795	
	(0.018)**	(0.019)	(0.355)**	
2012	0.061	0.027	-0.472	
	(0.018)***	(0.021)	(0.359)	
Entry grades*2005			0.043	
			(0.023)*	
Entry grades*2006			0.046	
			(0.032)	
Entry grades*2007			0.070	
			(0.038)*	
Entry grades*2008			0.078	
			(0.039)**	
Entry grades*2009			0.150	
			(0.058)***	
Entry grades*2010			0.156	
			(0.065)**	
Entry grades*2011			0.139	
			(0.061)**	
Entry grades*2012			0.087	
			(0.062)	
N	968	968	968	968
Universities	116	116	116	116
σ_u	0.0880	0.0847	0.0841	0.0767
	(0.0079)***	(0.0078)***	(0.0066)***	(0.0092)***
σ_v	0.0188	0.0226	0.0238	0.0349
	(0.0085)**	(0.0063)***	(0.0043)***	(0.0060)***
λ	4.676	3.754	3.529	2.194
	(0.0151)***	(0.0132)***	(0.0097)***	(0.0143)***

Notes: * significant at the 10% level; ** significant at the 5% level; *** significant at the 1% level. Estimation method is Greene's (2005) True Random Effects model with heteroskedastic-robust standard errors in parentheses. $\lambda = \sigma_u/\sigma_v$.

TABLE 3

THREE WAY BLINDER-OAXACA DECOMPOSITION OF THE RESULTS.

		(1) Group 1 = after 2008	(2) Group 1 = Pre-92	(3) Group 1 = Pre-92	(4) Group 1 = Pre-92
overall	Group 1	4.165	4.254	4.252	4.270
		(0.014)***	(0.009)***	(0.008)***	(0.009)***
	Group 2	4.133	4.016	4.059	4.047
		(0.012)***	(0.024)***	(0.018)***	(0.025)***
	difference	0.032	0.238	0.192	0.223
		(0.018)*	(0.025)***	(0.019)***	(0.027)***
	endowments	0.085	0.082	-0.022	-0.006
		(0.010)***	(0.020)***	(0.031)	(0.041)
	coefficients	-0.159	0.058	-0.010	-0.017
		(0.031)***	(0.034)*	(0.028)	(0.042)
	interaction	0.106	0.099	0.224	0.246
		(0.026)***	(0.031)***	(0.036)***	(0.055)***
endowments	Ln(Entry grades)	0.016	0.035	0.006	0.012
		(0.005)***	(0.017)**	(0.013)	(0.016)
	Ln(Student-staff ratio)	0.000	-0.000	0.006	0.005
		(0.001)	(0.012)	(0.009)	(0.010)
	Ln(library spend)	0.005	0.009	0.005	0.005
		(0.005)	(0.007)	(0.008)	(0.008)
	Ln(facilities spend)	-0.004	-0.000	-0.004	-0.002
		(0.004)	(0.005)	(0.005)	(0.005)
	Ln(research score)	0.068	0.038	-0.035	-0.026
		(0.009)***	(0.008)***	(0.024)	(0.029)
	time			-0.000	0.0234
				(0.004)	(0.0298)
	interact			-0.0236	
				(0.0321)	
coefficients	Ln(Entry grades)	0.891	0.887	1.501	1.532
		(0.358)**	(0.369)**	(0.330)***	(0.358)***
	Ln(Student-staff ratio)	0.196	-0.085	-0.090	-0.054
		(0.123)	(0.182)	(0.151)	(0.146)
	Ln(library spend)	0.435	-0.115	0.024	0.027
		(0.205)**	(0.173)	(0.199)	(0.190)
	Ln(facilities spend)	0.126	0.151	0.194	0.154
		(0.107)	(0.125)	(0.128)	(0.129)
	Ln(research score)	0.376	0.090	0.344	0.434
		(0.156)**	(0.058)	(0.098)***	(0.154)***
	Constant	-2.183	-0.871	-1.980	-2.105
		(0.322)***	(0.414)**	(0.425)***	(0.484)***
	time			-0.003	0.0214
				(0.003)	(0.0289)
	interact				-0.0271
					(0.0301)

TABLE 3 (CONTINUED)
THREE WAY BLINDER-OAXACA DECOMPOSITION OF THE RESULTS.

		(1) Group 1 = after 2008	(2) Group 1 = Pre-92	(3) Group 1 = Pre-92	(4) Group 1 = Pre-92
interaction	Ln(Entry grades)	0.016	0.064	0.108	0.110
		(0.007)**	(0.027)**	(0.024)***	(0.026)***
	Ln(Student-staff ratio)	-0.000	0.007	0.008	0.005
		(0.001)	(0.016)	(0.013)	(0.013)
	Ln(library spend)	0.018	-0.005	0.001	0.001
		(0.008)**	(0.008)	(0.010)	(0.009)
	Ln(facilities spend)	0.008	0.008	0.011	0.008
		(0.006)	(0.007)	(0.007)	(0.007)
	Ln(research score)	0.065	0.025	0.096	0.121
		(0.027)**	(0.016)	(0.028)***	(0.043)***
	time			0.001	-0.0217
				(0.004)	(0.0301)
	interact				0.0225
					(0.0298)
N		968	968	968	968
N_1		450	509	509	509
N_2		518	459	459	459

Notes: * significant at the 10% level; ** significant at the 5% level; *** significant at the 1% level. Heteroskedastic-robust standard errors in parentheses. Each column of the table divides the sample into two groups. In column (1), Group 1 is from 2009 to 2012 and Group 2 is from 2005 to 2008. In columns (2) to (4), Group 1 is for pre-92 universities while Group 2 is for post-92 universities. The first panel shows the overall difference in the outcome variable (log percentage good degree) and decomposes it into endowment, coefficient and interaction effects. The rest of the table shows the detailed decomposition by variable. Time and interact are the subsumed effects of all time dummies and the interactions between time dummies and entry grades, where the time dummies and interaction terms are normalised relative to the grand mean (Jann 2008).

TABLE 4
INTERACTING YEAR WITH MEASURES OF UNIVERSITY QUALITY

Dep var: ln(% good degree)	(1)		(2)	
	Coefficient	S.E.	Coefficient	S.E.
Ln(Entry grades)	0.108	(0.037)***	0.156	(0.030)***
Ln(Student-staff ratio)	-0.044	(0.024)*	-0.028	(0.027)
Ln(library spend)	0.028	(0.015)*	0.025	(0.016)
Ln(facilities spend)	-0.003	(0.012)	0.005	(0.011)
Ln(research score)	-0.006	(0.022)	0.021	(0.008)***
2005	0.008	(0.006)	-0.006	(0.007)
2006	-0.000	(0.006)	-0.017	(0.009)*
2007	0.000	(0.009)	-0.025	(0.011)**
2008	0.003	(0.009)	-0.021	(0.011)*
2009	0.011	(0.010)	-0.015	(0.012)
2010	0.019	(0.011)*	-0.007	(0.013)
2011	0.032	(0.012)***	0.010	(0.013)
2012	0.048	(0.012)***	0.035	(0.013)***
Post-92 university	-0.154	(0.039)***		
Post-92 * 2005	-0.025	(0.012)**		
Post-92 * 2006	-0.031	(0.015)**		
Post-92 * 2007	-0.041	(0.017)**		
Post-92 * 2008	-0.037	(0.019)**		
Post-92 * 2009	-0.005	(0.036)		
Post-92 * 2010	-0.001	(0.038)		
Post-92 * 2011	0.009	(0.039)		
Post-92 * 2012	0.029	(0.041)		
Russell Group			0.122	(0.016)***
Russell Group * 2005			0.009	(0.009)
Russell Group * 2006			0.013	(0.012)
Russell Group * 2007			0.018	(0.021)
Russell Group * 2008			0.005	(0.015)
Russell Group * 2009			0.008	(0.015)
Russell Group * 2010			0.003	(0.015)
Russell Group * 2011			0.011	(0.017)
Russell Group * 2012			0.001	(0.018)
N	968		968	
Universities	116		116	
σ_u	0.0847	(0.0072)***	0.0858	(0.0072)***
σ_v	0.0191	(0.0055)***	0.0208	(0.0055)***
λ	4.437	(0.0118)***	4.126	(0.0116)***

Notes: * significant at the 10% level; ** significant at the 5% level; *** significant at the 1% level. Estimation method is Greene's (2005) True Random Effects model with heteroskedastic-robust standard errors in parentheses. $\lambda = \sigma_u/\sigma_v$.

TABLE 5

PERCENTAGE OF FIRST CLASS DEGREES AS DEPENDENT VARIABLE

Dep var: ln(% first class degree)	(1)	(2)	(3)
Ln(Entry grades)	0.362	0.314	0.264
	(0.112)***	(0.099)***	(0.125)**
Ln(Student-staff ratio)		-0.054	-0.047
		(0.076)	(0.074)
Ln(library spend)		0.085	0.079
		(0.038)**	(0.040)**
Ln(facilities spend)		-0.001	-0.005
		(0.028)	(0.027)
Ln(research score)		0.087	0.116
		(0.023)***	(0.042)***
2005	0.020	0.020	-0.130
	(0.014)	(0.014)	(0.352)
2006	0.057	0.051	0.096
	(0.018)***	(0.019)***	(0.414)
2007	0.065	0.044	0.153
	(0.020)***	(0.022)**	(0.455)
2008	0.118	0.101	0.433
	(0.023)***	(0.026)***	(0.448)
2009	0.143	0.070	-1.086
	(0.027)***	(0.028)**	(0.653)*
2010	0.180	0.103	-0.615
	(0.027)***	(0.028)***	(0.670)
2011	0.258	0.185	-0.285
	(0.027)***	(0.028)***	(0.657)
2012	0.389	0.309	0.714
	(0.028)***	(0.032)***	(0.664)
Entry grades*2005			0.027
			(0.062)
Entry grades*2006			-0.008
			(0.073)
Entry grades*2007			-0.018
			(0.081)
Entry grades*2008			-0.057
			(0.078)
Entry grades*2009			0.201
			(0.113)*
Entry grades*2010			0.124
			(0.115)
Entry grades*2011			0.081
			(0.113)
Entry grades*2012			-0.072
			(0.114)
N	968	968	968
Universities	116	116	116
σ_u	0.1989	0.1880	0.1917
	(0.0342)***	(0.0293)***	(0.0277)***
σ_v	0.0649	0.0696	0.0639
	(0.0202)***	(0.0143)***	(0.0146)***
λ	3.066	2.7003	3.001
	(0.0530)***	(0.0424)***	(0.0410)***

Notes: * significant at the 10% level; ** significant at the 5% level; *** significant at the 1% level. Estimation method is Greene's (2005) True Random Effects model with heteroskedastic-robust standard errors in parentheses. $\lambda = \sigma_u/\sigma_v$.

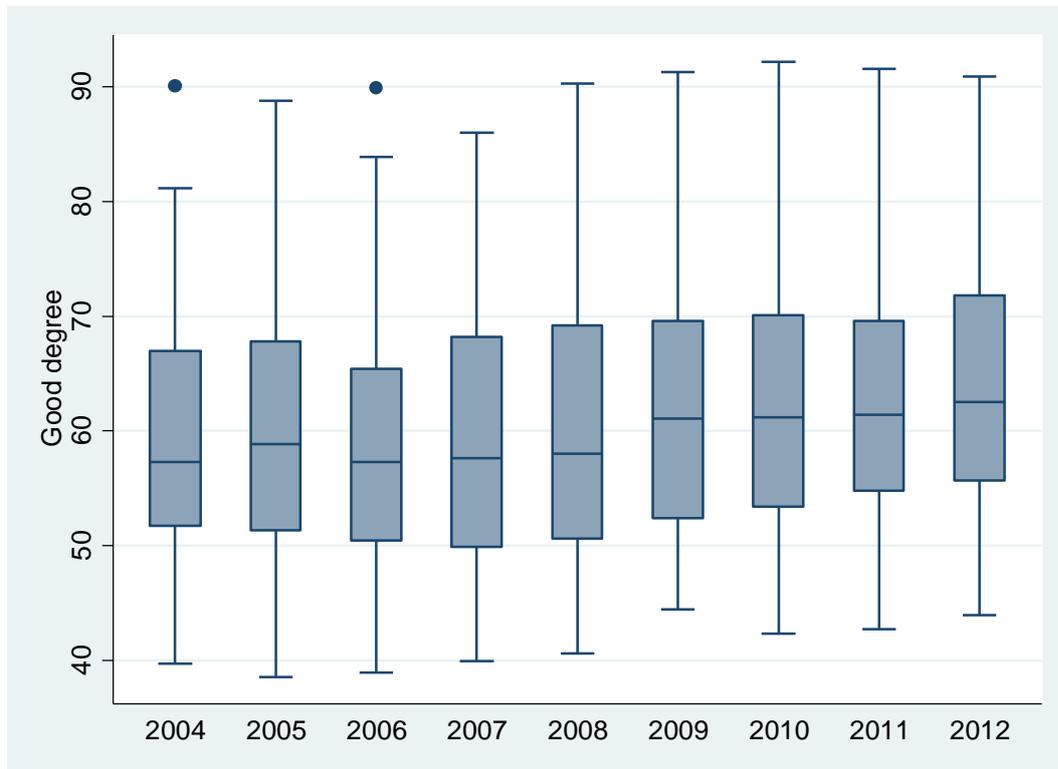FIGURE 1. Box plot of percentage of good degrees (first class and upper second class honours) over time.

FIGURE 2. Box plot of efficiency scores over time (using the model of column 3 of Table 2).
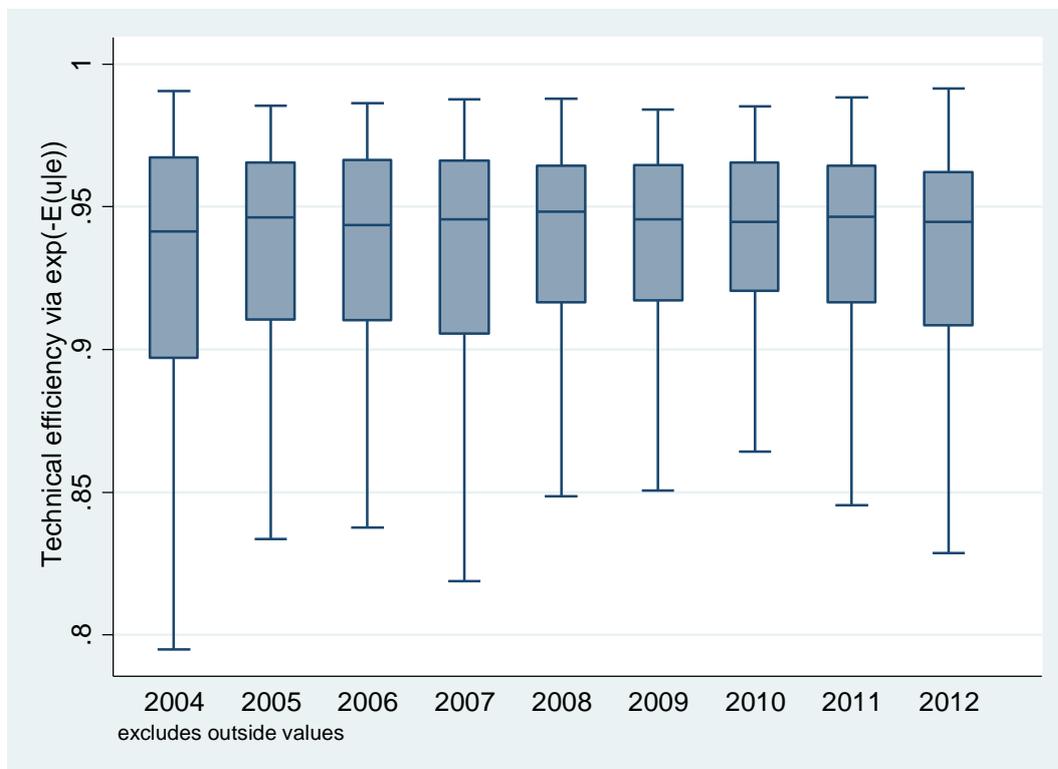
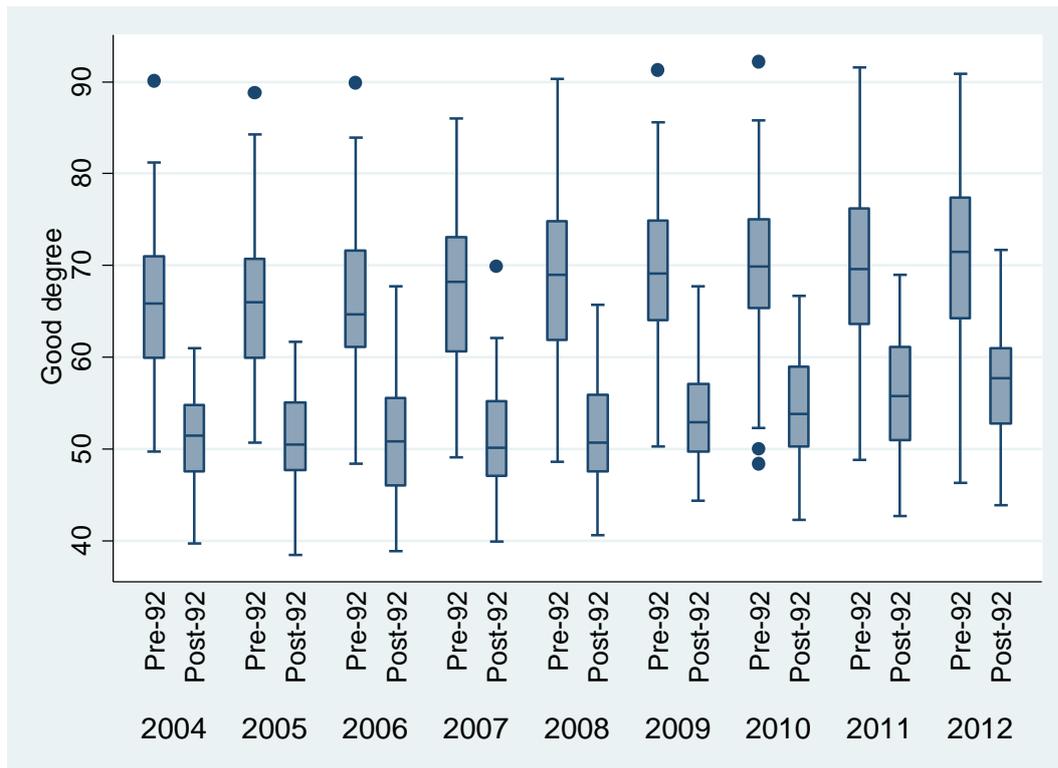FIGURE 3. Percentage of good degrees by pre-92 and post-92 universities.

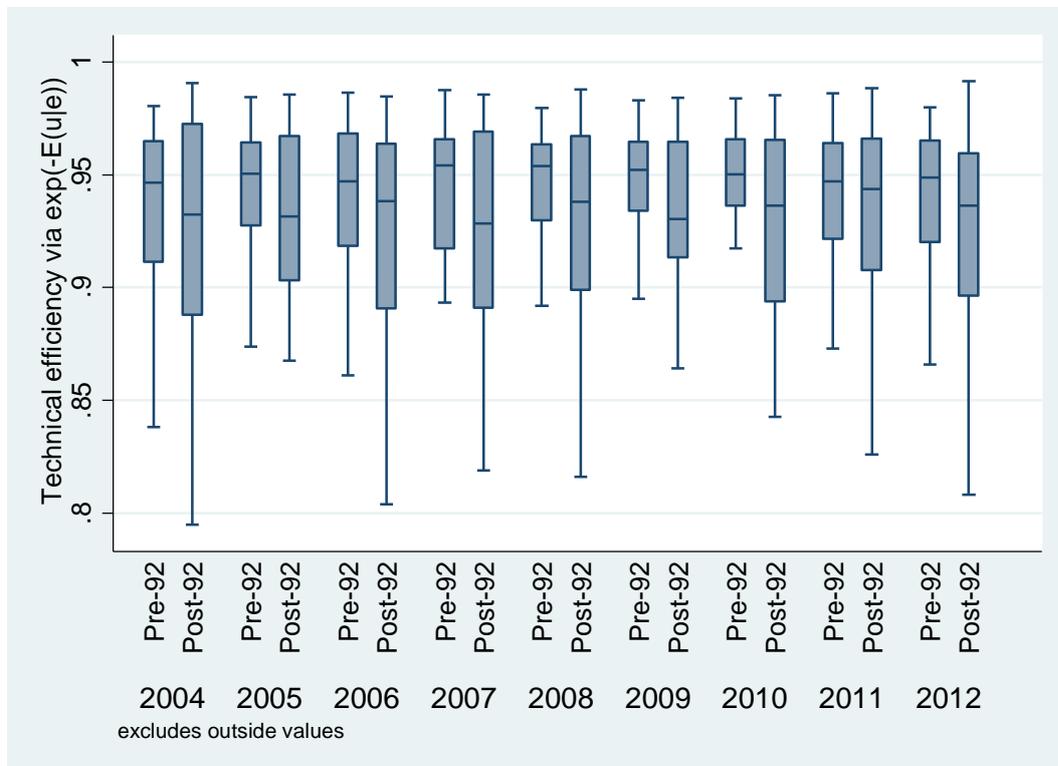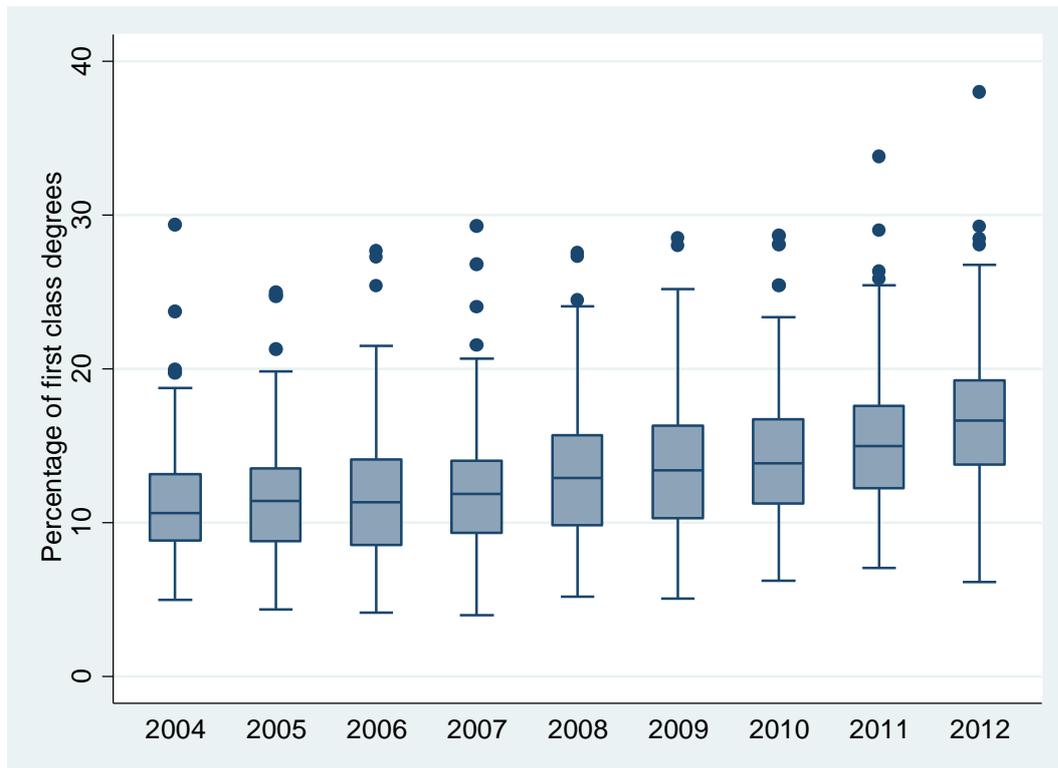FIGURE 4. Efficiency scores by pre-92 and post-92 universities.

FIGURE 5. Box plot of percentage of first class degrees over time.

APPENDIX
RANDOM EFFECTS RESULTS FOR THE SAME SPECIFICATIONS AS TABLE 2.

Dep var: ln(% good degree)	(1)	(2)	(3)	(4)
Ln(Entry grades)	0.404	0.316	0.217	0.309
	(0.043)***	(0.042)***	(0.069)***	(0.042)***
Ln(Student-staff ratio)		-0.073	-0.077	-0.065
		(0.032)**	(0.033)**	(0.029)**
Ln(library spend)		0.050	0.042	0.026
		(0.016)***	(0.018)**	(0.014)*
Ln(facilities spend)		-0.001	-0.005	-0.002
		(0.012)	(0.012)	(0.012)
Ln(research score)		0.029	0.061	0.035
		(0.009)***	(0.019)***	(0.007)***
2005	0.008	0.009	-0.102	
	(0.006)	(0.006)	(0.166)	
2006	-0.001	-0.004	-0.187	
	(0.009)	(0.010)	(0.258)	
2007	-0.021	-0.031	-0.350	
	(0.011)**	(0.011)***	(0.269)	
2008	-0.022	-0.029	-0.317	
	(0.011)**	(0.011)**	(0.268)	
2009	-0.003	-0.031	-0.895	
	(0.012)	(0.013)**	(0.357)**	
2010	-0.001	-0.028	-1.012	
	(0.013)	(0.013)**	(0.379)***	
2011	0.016	-0.008	-0.948	
	(0.012)	(0.013)	(0.362)***	
2012	0.033	0.009	-0.643	
	(0.013)***	(0.013)	(0.370)*	
Entry grades*2005			0.020	
			(0.029)	
Entry grades*2006			0.033	
			(0.045)	
Entry grades*2007			0.058	
			(0.047)	
Entry grades*2008			0.053	
			(0.047)	
Entry grades*2009			0.151	
			(0.062)**	
Entry grades*2010			0.172	
			(0.066)***	
Entry grades*2011			0.165	
			(0.063)***	
Entry grades*2012			0.114	
			(0.065)*	
N	968	968	968	968
Universities	116	116	116	116

Notes: * significant at the 10% level; ** significant at the 5% level; *** significant at the 1% level. Estimation method is random effects with heteroskedastic-robust standard errors in parentheses.