Rules versus Discretion in Committee Decision Making: An Application to the 2001 RAE for UK Economics Departments

Cherif Guermat, Richard D.F. Harris and Nigar Hashimzade

School of Business and Economics
University of Exeter

February 2004

Abstract:
The question of rules versus discretion has generated a great deal of debate in many areas of the social sciences. Recently, much of the discussion among academics and stakeholders about the assessment of research in UK higher education institutions has focused on the means that should be used to determine research quality. We present a model of committee decision-making when both rules and discretion are available. Some of the predictions of the model are tested empirically using the UK RAE 2001 results.

JEL Classifications: C25, D71, D78, D81.

Keywords: Rules; Discretion; Research Assessment Exercise; Probit.

Address for correspondence: Cherif Guermat, School of Business and Economics University of Exeter, Exeter EX4 4PU, UK. Tel: +44 (0) 1392 264434, Fax: +44 (0) 1392 263242. Email: C.Guermat@exeter.ac.uk.
We would like to thank David de Meza, Valentina Corradi, Dieter Balkenborg and Todd Kaplan for helpful comments.

1 Introduction.

The balance between rules and discretion has generated a great deal of debate among economists. Most of this debate has been in monetary policy making (Blackburn and Christensen, 1989; Fischer, 1990; and Clarida et al., 1999). In their seminal paper Kydland and Prescott (1977) demonstrated that policy rules are superior to discretion in monetary policy. Taylor (1993, p.197) stresses that there is a substantial consensus in macroeconomics that "policy rules have major advantages over discretion in improving economic performance."

In a different setting, discretion finds favour among management scientists. Many hold the view that allowing firms flexibility enhances performance (Kelman, 1961; Beyer and Trice, 1978; Marcus, 1988; Strebel, 1987), while rule-based culture stifles innovation within an organisation (Eisenhardt, 1989).

A third area of research in which discretion has been debated is the design of law and regulation. While strict laws and regulations ensure the protection of rights, fairness, and equal treatment, and offer greater levels of predictability, they also create rigidities and the erosion of trust, learning, and cooperation (Sitkin and Bies, 1994). It is found that more flexible approaches to regulation have many advantages (Bourgeois and Brodwin, 1984; Marcus, 1988). This appears to be confirmed by empirical studies, where more flexible regulation is found to lead to increased productivity (Majumdar, 1997). However, there is some empirical evidence that suggests that too much law is as problematic as too much discretion, and that an appropriate balance between legal rules and discretion is needed (Majumdar and Marcus, 2001).

The focus of this paper is on a different situation where a balance between rules and discretion is sought. We posit a decision maker (committee) that observes two noisy signals – a rule signal that is publicly observed and a discretion signal known only to the decision maker. The decision maker may be sensitive to political and social pressure and tries to find a balance between rules and discretion that minimises a specific loss function. Typically, committees always try to find a balance between rules and discretion within an environment of constraints and pressure. Political and social pressure aside, combining rules and discretion is more optimal than relying on a single measure. The reason is that both rule and discretion signals are noisy and, hence, there is a chance that the information contained in one signal is not available in the other. In the grading of research quality, rules offer greater transparency and equity, but rules may not account for certain characteristics. For example, suppose some sort of journal ranking is adopted and research papers are judged solely on the basis of where they were published. However, one can find top quality pieces of research in lower ranking journals. Clearly, these would not be picked up by the above rule, but can be easily accounted for via a discretionary measure.

We use the data from the 2001 UK Research Assessment Exercise to empirically examine some of the predictions of our model. We find evidence that both rules and discretion were used in the 2001 RAE for UK economics departments.

The rest of the paper is organized as follows. The next section presents a simple model of committee decision-making. Section 3 provides a brief review of journal quality ranking in the recent literature. Section 4 discusses the main features of the UK RAE 2001. Section 5 discusses the main data and modelling considerations. Sections 6 and 7 discuss the results of the estimated models and the last section concludes.

2 The Model

A committee has to give a verdict on a characteristic (quality) of a subject (department). The quality Q is not observed directly. Instead, noisy signals

$$R = Q + e \quad (1)$$
$$D = Q + p\omega + v \quad (2)$$

where

$$\begin{pmatrix} e \\ v \end{pmatrix} \sim (\mathbf{0},\sigma^2\mathbf{I}) \quad (3)$$

Signal R is a "rule" measure, such as publicly available quality indices, and is observed publicly. Signal D is a "discretionary" measure, observed only by the committee. The quantity p is some characteristic of a department, other than quality, that can influence the committee's discretion (for example, political pressure), and $\omega \in [0,1]$ is the degree to which the committee is submissive or vulnerable to this influence. We assume that p, ω, and σ^2 are known to the committee.

As true quality of the department is unobservable, the committee estimates it given available information. Conditional on p and ω, the optimal (unbiased minimum variance) estimator of Q is the ordinary least squares (OLS) estimator,

$$\hat{Q} = \frac{1}{2}(R + D - p\omega). \tag{4}$$

Higher values of σ^2 imply higher noise of observed quality measures and can thus be thought of as a measure of accuracy of the estimator. Clearly, an estimator of quality based solely on one of the two observed measures would be sub-optimal. Based on the estimated quality and other information the committee offers to the department the grade π^o based on a combination of rule and discretion:

$$\pi^o = \lambda R + (1 - \lambda)D. \tag{5}$$

The department anticipates to receive the grade π^a comprising publicly available quality measure R and full gain from its pressure on the committee:

$$\pi^a = R + p. \tag{6}$$

We posit the following loss function for the committee:

$$\mathcal{L} = \omega(\pi^a - \pi^o)^2 + (1 - \omega)(\pi^o - \hat{Q})^2. \tag{7}$$

The term $(\pi^a - \pi^o)$ in the first part of the loss function is the grade gap reflecting the difference between the grade the department anticipates to obtain and the grade offered by the committee. This can be thought of as a social constraint, because the committee wishes to, say, minimise the possible protest generated by a large grade gap.[1] The term $(\pi^o - \hat{Q})$ in the second part of the loss function represents the gap between the optimal grade and the grade offered by the committee. A committee that is completely insensitive to political pressure would set the grade offered equal to the optimal grade $\hat{Q}$. Higher values of ω correspond to higher sensitivity of the committee to pressure. Note that we use sensitivity ω in both (7) and (2), assuming that the committee has the same sensitivity to department and social pressure.

The committee chooses $\lambda \in [0,1]$, the optimal combination of rule and discretion, to minimize the expected value of the loss function $\mathcal{L}$. When the constraint $0 \leq \lambda \leq 1$ is not binding, the solution is interior and satisfies the first order condition:

$$\frac{d}{d\lambda} E\mathcal{L} = 0$$

or, using (7),

$$\omega \frac{d}{d\lambda} E\left[(\pi^a - \pi^o)^2\right] + (1 - \omega)\frac{d}{d\lambda} E\left[(\pi^o - \hat{Q})^2\right] = 0. \tag{8}$$

Using (1)-(6) we obtain:

$$\frac{d}{d\lambda} E\mathcal{L} = 2\lambda\left[(p\omega)^2 + 2\sigma^2\right] - 2\sigma^2(1 + \omega) \tag{9}$$

Hence, for optimal λ we have

$$\lambda^* = \frac{1 + \omega}{2 + (p\omega/\sigma)^2} \tag{10}$$

The second order condition holds trivially given $\sigma^2 > 0$:

$$\frac{d^2}{d\lambda^2} E\mathcal{L} = 2(p^2\omega^2 + 2\sigma^2) > 0.$$

[1] Quadratic loss suggests that the committee believes that there is also a cost to offering a higher grade than anticipated. This can be seen as the protest of competing departments, a watchdog, or simply a budget constraint.

It is easy to see that when the committee is not sensitive to political pressure ($\omega = 0$) it puts equal weights on the rule and discretion and offers the optimal grade. When there is no pressure from the department, $p = 0$, the optimal weight put on the rule is

$$\lambda^* = \frac{1+\omega}{2}.$$

In this case, the grade offered still depends on the committee's sensitivity to pressure, but this effect originates from the loss function (7) rather than the discretion equation (2). Thus, the committee always puts more weight on the rule since $\lambda^* > \frac{1}{2}$. The department, whose aim is to reduce λ^* is worse off when it is incapable of exercising pressure on the committee.

When there is no noise (the quality of the department is perfectly observable) the committee puts full weight on discretion, $\lambda^* = 0$. As noise increases λ^* also increases:

$$\frac{\partial \lambda^*}{\partial \sigma^2} = \frac{(p\omega)^2(1+\omega)}{\left[2\sigma^2 + (p\omega)^2\right]^2} > 0$$

and

$$\lim_{\sigma^2 \to \infty} \lambda^* = \frac{1+\omega}{2}.$$

Thus, when the noise is very high, the more submissive the committee is to political pressure (the closer to unity ω is), the more weigth it puts on the rule. Conditional on p and ω, the expected grade offered, $E(\pi^o) = Q + (1-\lambda^*)p\omega$, decreases with noise increasing.

The optimal weight put on the rule is inversely related to the pressure from the department for given σ^2 and $\omega \neq 0$:

$$\frac{\partial \lambda^*}{\partial p} = -\frac{2p\omega^2\sigma^2(1+\omega)}{\left[2\sigma^2 + (p\omega)^2\right]^2} < 0.$$

The expected grade offered increases in p for $\omega \neq 0$. Therefore, if the department can choose the level of pressure prior to the committee's decision, it will try to increase it as much as possible (subject to some cost constraint).

The effect of ω on λ^* and π^o is not unambiguous and may change sign depending on p and σ^2. Differentiation of (10) with respect to ω gives

$$\frac{\partial \lambda^*}{\partial \omega} = -\frac{\sigma^2(p^2\omega^2 + 2p^2\omega - 2\sigma^2)}{(2\sigma^2 + p^2\omega^2)^2} \tag{11}$$

Therefore, λ^* is increasing in ω for $0 \leq \omega \leq \bar{\omega}$ and decreasing in ω for $\bar{\omega} \leq \omega \leq 1$, where $\bar{\omega} = \min\{(\sqrt{1 + 2\sigma^2/p^2} - 1); 1\}$. The threshold $\bar{\omega}$ depends on the noise to pressure ratio (σ/p). Thus, as long as the sensitivity of the committee is lower than the threshold, higher sensitivity means increased use of rule measure. A department aims at decreasing λ^* and thus requires the threshold to be as small as possible to reverse the effect of ω on λ^*. The department thus needs to put the highest pressure possible for fixed σ. As the signal becomes noisier, the department needs additional pressure in order to reduce λ^*. This reversal of the effect of committee's sensitivity is mainly due to its presence in both the loss function (7) and the discretion equation (2). For a given pressure, as the signal on the quality becomes noisier, the committee is relatively more concerned about the social pressure compared with the department's pressure. At an extreme, when the signal becomes totally unreliable ($\sigma^2 \to \infty$) a completely sensitive committee ($\omega = 1$) would not use discretion at all to avoid criticism even though the rule signal is equally unreliable.

2.1 Heteroscedastic Case.

These results can be easily generalized for an arbitrary covariance structure of the error terms in R and D. Assuming

$$\begin{pmatrix} e \\ v \end{pmatrix} \sim \left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma_e^2 & \sigma_{ev} \\ \sigma_{ev} & \sigma_v^2 \end{bmatrix} \right) \tag{12}$$

the optimal estimator for quality Q is the generalised least squares (GLS) estimator

$$\hat{Q} = \alpha R + (1-\alpha)(D - p\omega) \tag{13}$$

where

$$\alpha = \frac{\sigma_v^2 - \sigma_{ev}}{\Sigma^2}. \tag{14}$$

and

$$\Sigma^2 \equiv Var(v-e) = \sigma_e^2 + \sigma_v^2 - 2\sigma_{ev}. \tag{15}$$

Using (12)-(15) in the loss function, from the first order condition for optimal λ we get

$$\lambda^* = \frac{\alpha + (1-\alpha)\omega}{1 + (p\omega)^2/\Sigma^2} \tag{16}$$

or, equivalently,

$$\lambda^* = \frac{\sigma_v^2 - \sigma_{ev} + (\sigma_e^2 - \sigma_{ev})\omega}{\Sigma^2 + (p\omega)^2} \tag{17}$$

Conditional on the structure of the covariance matrix of the noise, the effect of p and ω on the optimal decision of the committee is the same as above. To analyze the effect of the the variances of random components in rule and discretion we assume, for simplicity, that the two random components are uncorrelated ($\sigma_{ev}^2 = 0$). Then

$$\alpha = \frac{\sigma_v^2}{\Sigma^2}$$

measures the proportion of the noise from discretion in the total noise. Differentiation of (16) with respect to α gives

$$\frac{\partial \lambda^*}{\partial \alpha} = \frac{1-\omega}{1 + (p\omega)^2/\Sigma^2} \tag{18}$$

which is non-negative. Intuitively, the noisier is the discretion signal, the more weight the committee puts on the rule, and vice versa.

It is plausible to assume that random disturbances to R and D are (positively) correlated. To analyze the effect of correlation between the two random components we differentiate (17) with respect to the covariance σ_{ev}:

$$\frac{\partial \lambda^*}{\partial \sigma_{ev}} = \frac{(1-\omega)(\sigma_v^2 - \sigma_e^2) - (1+\omega)(p\omega)^2}{\left[\Sigma^2 + (p\omega)^2\right]^2}.$$

A highly influenceable committee (ω close to unity) puts more weight on the rule and less on discretion, the lower is the covariance between two noises. In general, the optimal weight put on the rule increases with the covariance between e and v if

$$\sigma_v^2 - \sigma_e^2 > \frac{1+\omega}{1-\omega}(p\omega)^2.$$

In particular, if the committee is not influenceable at all ($\omega = 0$), the weight put on the rule increases with the increase in covariance if the rule is less noisy than discretion ($\sigma_e^2 < \sigma_v^2$), and vice versa.

3 Journal Quality and the Ranking of Departments.

Until now, the Research Assessment Exercise (RAE) has been the only formal source of department ranking, and largely determines how research funding is allocated in the UK's higher education institutions. On the other hand, a large number of informal, metric-based, department rankings can be found in the literature. Among these are Dusansky and Vernon (1998) in the US, Kalaitzidakis et al. (1999) in Europe, Lucas (1995) in Canada, and Harris (1990) in Australia. Many ranking exercises focus mainly on a limited number of leading journals (Conroy et al., 1995; Dusansky and Vernon, 1998, and Kalaitzidakis et al., 1999). Journal ranking is also increasingly being used to rank individual academics (Baltagi, 1999).

These informal rankings are predominantly based on some measure of citations. However, both perception-based and citation-based ranking methodologies are found in the literature (Harzing, 2001; Liner, 2002; Theoharakis and Hirst, 2002; Swidler and Goldreyer, 1998; Kalaitzidakis et al., 1999; Burton and Phimister, 1995).

There is an on-going debate over the reliability of citation-based rankings. While citations are an objective measure of journal quality, they have been criticized for being biased (Seglen, 1997) and for being a measure influence rather than excellence (Beed and Beed, 1996). However, perceptions can easily be criticized for being subjective. For example, Axarloglou and Theoharakis (2003) found that economists' perception of journal quality depends on their geographic origin, school of thought, journal affiliation, field of specialization and research orientation.

Almost all existing rankings are based on rule measures. These measures have two main elements, namely the choice of a set of research journals and the choice of functions of perceptive or objective quality variable, such as the number of citations. Only a handful of studies have attempted to match some rule based ranking of journals with the outcome of the RAE. One notable example is Burkitt and Baimbridge (1995), who use the number of papers published in the Economic Journal as an explanatory variable for the ratings awarded in the 1992 RAE, and find a significant positive relationship. Marston and Ayub (2000) repeated Burkitt and Baimbridge (1995) study for the UK accounting departments rating in the 1996 RAE. Their study is based on four UK accounting journals. Campbell et al. (1999) also studied the RAE outcome in UK law departments. In the US, Swidler and Goldreyer (1998) formally test the link between top finance journal publications and salaries of finance faculty at US research-oriented universities. The main limitation of these studies is the reliance on a very limited number of 'top-tier' journals.

4 The 2001 UK RAE.

The latest UK Research Assessment Exercise, and the recent Roberts' report that proposes its replacement (Roberts, 2003), have generated a great deal of debate among academics in the UK. The interest in the RAE and its future form is not confined to the UK, however. The UK RAE model has attracted a near-worldwide interest and is often cited as a benchmark. This is particularly true for Ireland, Australia, New Zealand and Hong Kong (von Tunzelmann and Mbula, 2003; Boston, 2002). The Roberts' report used information and views from various sources, including official documents, higher education institutions, researchers and other stakeholders. Among the issues raised in these studies are the use of metrics and expert judgement. The Roberts' Report (2003, Annex E) identified strong support among stakeholders in favour of the use of metrics alongside expert review in research assessment. The overwhelming view is that future assessments should give a greater role to metrics. However, the report recommended that the new system of research assessment must be based upon the judgement of experts. Assessment committees may or may not use performance indicators (metrics) to inform their judgement.

The main criterion for assessment in the UK RAE is research quality. Research academics were classified as international, national and sub-national levels, depending on the quality of the their research. Officially, RAE panels used their judgement to form a view on the quality of all submitting staff. Thus, it seems likely that the norm is discretion rather than rule. This is corroborated by the official line that panels were concerned only with the quality of the research output and not with the form it took (THES, 14 December 2001). Thus, whether a book, a chapter or a paper article, the RAE panels are supposed to have judged individual research outputs in their own merits. However, it is also acknowledged by the RAE that those outputs that have been through a rigorous process of peer review would not be examined. This reflects the view that the general reputation of a journal strongly affects how a research paper is assessed (Vick et al., 1998). It is therefore suspected that there has been at least some use of rule like measures of journal quality in the panel's assessment. This is especially likely at the level of top tier journals.

There has been much criticism directed at the way the RAE has been conducted. One of the most ardent criticisms was made by Universities UK (UUK), which argued that the same conclusions could have been reached with much less effort and that the RAE was hardly cost-effective (THES, 08 November 2002). Indeed, the financial cost of the RAE has been estimated at £36 million (Lewis, 2003).

In economics and econometrics, a total of 3255 research items were submitted in the RAE 2001. Of these, 2464 were refereed journal articles, and 791 other types of submission such as authored and edited books and working papers. The journals that had 10 or more articles submitted to the RAE are shown in Table 1. The highest number of submitted papers for a single journal was 118 articles for the Economic Journal. Econometrica, the Quarterly Journal of Economics and the American Economic Review had 40, 25 and 18 submissions respectively. For simplicity, no distinction is made between the various types of submissions, such as full papers, notes and proceedings.

[Table 1 about here]

5 Data Selection and Modelling Considerations.

The main variables of interest in this paper are rules, discretion and department pressure. Clearly, none of these are observed perfectly. The rules variable is the least problematic because information on rules, albeit noisy, is normally available to the public. We describe below how we construct a variable that can be used as a proxy for rules.

Measuring discretion is more problematic, since, by definition, it is not publicly observable. However, there are a few ways in which the presence of discretion in the committee's decision can be detected. The first is the use of political pressure by the department. Our model suggests that the optimal weight the committee puts on rules is inversely related to the level of pressure (see (10)). Empirically, significant pressure would indicate a positive weight on discretion in the committee's decision. The second is the presence of noise in the rules signal. When the rules signal does not deliver perfect information on the department quality, the committee uses both rules and discretion in its decision, with the optimal weight of the discretion being positively related to the noise level in the rules signal (see (18)). This can be assessed empirically by comparing the predictions of the fitted model, which is based partly on rules, with the actual outcome of the committee's decision. Finally, there could be some publicly available information on the committee's use of discretion.

In the empirical part of this paper, we add two additional means of assessing discretion. Firstly, it might be useful to adopt a hybrid variable, based partly on rules and partly on discretion, as a measure of discretion. In using this variable, the intention is to at least capture part of the discretion used by the committee. A definition of this variable is provided in the empirical section. The second means is based on the argument that the committee's violation (or bending) of its own rules entails its use of discretion. We use the Gini coefficient to test for such a violation as explained below.

Political pressure is not observable either. In this paper, we proxy political pressure using department's size. Although it is hard to accept that departments actively pressurise RAE panels, it is possible that departments are endowed with certain characteristics that put pressure on the panels passively. A department's size is likely to encompass many of these influencing factors. Larger departments are more likely to have renowned academics and, thus, panels are likely to over-rate these departments. Larger departments are also more likely to have a higher absolute number of top quality research articles. Even though panels are supposed to look at a measure of relative performance, it is likely that it would be hard for them to ignore the absolute number of top quality research output. Larger departments are also influential politically because of their position in the educational system and their higher relative share of research output in the country. Some large departments also happen to enjoy historical and political prominence. Panels are likely to be reluctant to give Cambridge and Oxford, for instance, lower grades. The following quotation from the vice-chancellor of the University of Central England points both to the position of influential universities and to the general pressure the RAE panels are likely to feel:

> "Any new system that has the effect of removing research money from Oxbridge et al is a nonstarter. It is not even worth thinking about the row that would result. Equally, any system that achieved a greater concentration of research funding would be deeply unpopular and would risk leading to the exclusion of the originality and occasional good fortune that leads to the development of new ideas and concepts." (Knight, 2002).

5.1 Proxies for Rules.

To derive a proxy for rules we need to derive a metric for measuring journal quality. As both perception-based and citation-based measures may have their particular merits and limitations, we use both measures in ranking journal articles.

In this paper we use three citation-based indices: impact factor, immediacy index and cited half-life. All three are available from the Journal Citation Report (JCR). The impact factor measures a journal's relative importance in terms of relative citation. The Immediacy Index is a measure of how quickly the "average article" in a journal is cited, while the cited half-life index reflects the age of the majority of cited articles published in a journal. We also use two perception-based indices: the Harzing (2001) classification and the Axarloglou and Theoharakis (2003) 'worldwide' quality index (WWQI). The latter study is based on a survey of opinions of 2,103 economists worldwide. The Harzing (2001) classification is based on various surveys based in England (Lancaster, Nottingham, Bradford, and Aston) and overseas (The Netherlands and Hong Kong). These surveys were carried out between 1994 and 2000.

For each department, the sum of immediacy, impact, half-life and WWQI scores were accumulated. These reflect the overall absolute departmental performance with respect to each criterion. Similarly, we use the total number of papers classified as international by the Harzing compilation, and the total number of working papers. The total number of 'other research output' items, which mainly includes unclassified refereed journal articles,[2] authored books, chapters in books and edited books, are also used.

We mimic the RAE panel assessment of individual researchers by taking the ratio of the total of each index to total submissions. In this way we have, for instance, the average performance in terms of immediacy index per submitted research output. Thus, this average performance can increase in two ways: one is by increasing the index itself (higher quality journal), and the other is by increasing the number of journal papers. For example a department that submits only books and chapters in books will have zero immediacy or impact scores.

Due to the potentially large number of explanatory variables and the limited number of observations, we explored the possibility of data reduction. Factor analysis was used to obtain the latent factor(s) generating the five average quality scores (impact, immediacy, half-life, WWQI, and Harzing International). The results suggest a single significant principal component, which accounts for almost 84% of the variability of the 5 average quality scores. This standardised factor represents the per-submitted-item quality performance of a given department. We label this 'AQI' for average quality index. This is our main proxy for rules.

The average number of working papers and 'other research output' items are also potential proxies for rules. Moreover, besides the quality of published research outputs, the 2001 RAE also defined international and national excellence in terms of postgraduate research activity and external research income. Thus, to assess whether the RAE panel took postgraduate activity and external research income into account, we consider two additional variables, namely the total number of PhD degrees awarded and the total external funds obtained by each department for the period covered by the RAE. These are four additional proxies for rules that can be tested empirically.

5.2 Proxy for Pressure.

The criteria used for the 2001 RAE ranking clearly suggest a relative comparison rather than an absolute one. Thus, as a rule, department size should not be influential in the RAE grading. In order to account for the suspected size effect, we include the total number of submissions as a proxy for size and, hence, pressure. This is preferred to the number of submitted staff because many academics have less than 4 submissions.

5.3 Proxies for Discretion.

In the construction of a hybrid variable, we explore the possibility that the RAE panel might give top tier journals more value than the rules measure suggests. We therefore counted the number of 'top tier' journal articles in each department. Top tier is defined here as journals that have scores greater than

[2] These include some well-known journals, such as Econometrics Journal, Journal of Time Series Analysis and Oxford Development Studies.

or equal to 1 for the impact factor, 0.1 for immediacy index, 10 for half-life, and 20 for WWQI. The Rand Journal of Economics has the lowest scores with these criteria. We call this the 'seniority effect' in the sense that the committee may use discretion to correct for the perceived understatement of top tier journals.

Another proxy, discussed earlier, is the Gini coefficient. At least officially, the average performance of departments is not the main focus of the RAE panels. Instead, the appraisal of individual academics is supposed to be the criterion for overall department rating. Thus, the overall rating of a department may depend crucially on the distribution of papers among academic staff. Specifically, for the same average quality level, a department that has a more even distribution of papers among its staff has a better chance of obtaining a higher grade. For example, a department that has one academic with four international papers and another with four national papers will be inferior to the one that has two academics with two international and two national papers each, even though on average they are identical.[3] Thus, a department that has more equality in terms of research quality output should be expected to score higher in the eyes of the RAE panel. To test for the presence of this effect a measure of asymmetry or inequality is required. We adopt the Gini coefficient as a proxy for the extent of the RAE panel's appraisal of individual academics.

To calculate the Gini coefficient we used the impact factor. The total impact score was computed for each submitted staff, and the cumulative contribution of each staff to the total departmental impact score was used to calculate the Gini coefficient using the formula:

$$GINI = 1 - \frac{1}{n}\sum_{i=1}^{n}(\Phi_i - \Phi_{i-1})$$

where Φ_i is the department's total impact score share of individual i and $\Phi_0 = 0$. The Gini coefficient takes a value between zero for perfect equality and 1 for perfect inequality.

5.4 The Probit Model.

The 41 RAE grades for economics departments are used as the dependent variable in a probit model using the above proxies as regressors. Because of data limitation we only consider three grades, namely 3, 4 and 5. Thus, no distinction is made between 3a and 3b grades, and 5 and 5* grades.

An ordered probit model is used to model the relationship between RAE ratings and selected inputs or explanatory variables. The probit model consists of estimating a model based on a utility index given by $I_i = \beta' x_i$ where x_i is a vector of exogenous variables and β is a vector of parameters. In an ordered probit model, there are additional threshold parameters. For instance, if there are three states: RAE3 ($y_i = 0$), RAE4 ($y_i = 1$) and RAE5 ($y_i = 2$), the probit model involves one additional parameter, μ, which enters the likelihood function. The probabilities for each state are given by

$$\begin{aligned} P(y_i &= 0) = 1 - F(I_i) \\ P(y_i &= 1) = F(\mu - I_i) - F(-I_i) \\ P(y_i &= 2) = 1 - F(\mu - I_i) \end{aligned}$$

where F is the Normal cumulative distribution function. The model can be readily generalised to include more states.

6 A Data-Driven Model.

We start our empirical exercise with a simple agnostic model. The aim for now is not to test any theory but to simply consider the RAE results as an output and attempt to link such an output with a number of inputs. The flexibility of this model makes it possible to see the significance as well as the marginal impact of each input.

[3] Assuming that 2 international and 2 national papers (or better) earns an academic the 'international' status.

The initial model includes size, seniority level, average quality index (AQI), the ratio of total working papers to total submitted items, the ratio of other refereed and non-refereed research outputs to total submitted items, the ratio of total PhD candidates to total submitted staff, the ratio of total funds to total submitted staff, and the Gini score of inequality. Table 2 provides summary statistics for these variables.

[Table 2 about here]

The unrestricted model, shown in the first part of Table 3, strongly suggests the insignificance of working papers, funding and the Gini score. All these variables were found to be insignificant by both individual and joint tests for removal from the initial model. Size also appears to be insignificant in the unrestricted model but subsequent restricted models suggest strong significance. The selected model, presented in the second part of Table 3, shows strong significance of all parameters except the intercept, which is still significant at the 10% level. The likelihood ratio statistic for the three restrictions was 2.69, which is insignificant at the 1% level.

As suggested earlier, both size and seniority level have a positive impact on the probability of a higher RAE rating. Higher average quality of research output is highly significant in determining RAE ratings, while the ratios of the other outputs have a negative impact. Note that the insignificance of the coefficient of working papers suggests a neutral effect, in the sense that it neither increases nor decreases the probability of a higher rating. The likely reason for this is that many 'new' academics are given some sort of exemption, and it is generally observed that new academics have a higher proportion of working papers for obvious reasons.

[Table 3 about here]

The estimated model gives rise to two major questions. The first is how well would our model predict the RAE outcome? The other question is what is the individual effect of each of the significant variables on the probability of the various ratings?

Using the restricted model, we generate probabilities for each department given size and the other metrics. The results are shown in Table 4. Prediction in probit models is not straightforward, since what we obtain is predicted probabilities of outcomes rather than outcomes per se. However, if we apply a convention whereby the outcome is selected on the basis of the highest probability our model would suggest seven wrong predictions out of the 41 cases. Three of these misses are borderline cases with probabilities that are almost equal. These are London Guildhall (under-rated), Leicester (over-rated), and Newcastle (under-rated). The other four wrong predictions are more clear-cut as well as interesting. The first two universities are a 5-4 rating mismatch. While Queen Mary obtained a five our model suggests a 4 rating with probability 0.586. On the other hand, Bristol obtained a four but our model suggests a 5 rating with probability 0.619. The other pair of universities is a 4-3 rating mismatch. These are Surrey, which was given a 3 but predicted to be a 4 with probability 0.747, and Wales Swansea, which obtained a 4 but was predicted to be a 3 with a similar high probability.

Notice that the probabilities of 5* rated departments are virtually 1 in most cases. However, the probabilities for Cambridge, Nottingham and Oxford are also close or equal to one. This is probably due to the size effect as the three departments are among the largest.

[Table 4 about here]

Although the estimated model is helpful in telling which variables are significant in RAE rating, it tells us little about the individual impact of each variable. In Table 5 we provide the approximate marginal effects on the probabilities of RAE rating. The estimated effects are based on the mean values of the five variables (Green, 2000, p.879). The probability for a 3, 4 and 5 rating of an 'average' department (i.e. based on mean values) is 0.0029, 0.9491 and 0.0480 respectively. An average department has about 80 submitted items, seven top tier papers, zero AQI score, 39% of its submitted items in the 'other output' category, and about one PhD student per academic staff.

The fifth, sixth and seventh columns provide the marginal effect of a one unit increase of each variable (relative to the mean value) on the probability of three, four and five rating respectively. Clearly, all but ‘other output’ have a positive impact on the probability of a 5 rating. For example, a department that has one additional top tier paper relative to an average department sees its probability of a five increase by 0.0223, and a department that has two students per academic staff would see the probability of a 5 increase by 0.131. On the other hand, every 10% increase in ‘other output’ sees the probability of a 5 reduced by about 0.0368.

A better insight of the marginal impact of each variable is given in the last column of Table 5. The figures show the change in the probability of a 5 rating when a typical average department is compared with the best (worst) department with respect to the variable in question. Surprisingly, seniority level is by far the most influential variable. The difference in the probability of a 5 rating between an average department and a similar department with the highest number of top tier journal articles is a staggering 0.6233. Meanwhile, what should have been expected to be the most influential variable – the average quality index – has a maximum increase of 0.1932. More surprisingly, size is almost as equally important as the average quality index. The other variables are similar in magnitude with the highest proportion of PhD students increases the probability of a 5 by about 0.14. On the other hand, the worst department in terms of having the highest proportion of other output sees its probability of a 5 rating reduced by about 0.12.

[Table 5 about here]

7 Estimating the Discretion Model.

The data driven model estimated in the previous section appears to mimick the RAE panel satisfactorily, with a significant fit and a good prediction of the RAE outcome. However, the mechanism in which rules, discretion and pressure operate is not captured properly by the data driven model. Even though data driven models enjoy flexibility and are often regarded as more poweful tools for forecasting, in our case the data driven model offers little help in confirming or rejecting the discretion model proposed in the previous sections. In this section, therefore, we provide a direct assessment of our discretion model.

To keep things simple, we assume the setting (1)-(3). However, we dispose of the assumption that the committee uses OLS (or GLS) to compute the optimal estimate $\hat{Q}$ of quality Q. Instead, we assume that the committee uses specific weights, γ and $(1-\gamma)$ on rule and discretion, respectively, and *estimate* these weights within the model. One reason for this choice is flexibility: we are trying to “guess” what estimator the committee used, rather than impose a particular estimator (OLS or GLS) on the committee. Another, and perhaps more important, reason is that the committee might be asked by a controlling authority to use specific weights on rules and discretion in its decision. For example, the controlling authority may want the committee to ignore the rules completely (and, thus, sets $\gamma = 0$). In the context of the RAE exercise, the funding bodies requires the panels to use their ‘expert judgement’ in evaluating the quality of research. Even though the panel might be informed by some bibliometric measure, we do not expect the optimal estimator $\hat{Q}$ to be based on the OLS or GLS estimators. Thus the weight γ put on discretion may be different from 0.5 as in the case of OLS. If, for example, the panels are supposed to use their expert judgement, we would expect γ to be less than 0.5.

Under this new assumption equation (4) becomes

$$\hat{Q} = \gamma R + (1-\gamma)(D - \omega p) \tag{19}$$

It is easy to show that the optimal rules weight λ_i of department i in this case is given by

$$\lambda_i = \frac{\gamma + (1-\gamma)\omega}{1 + (\omega p_i)^2/2\sigma^2} \tag{20}$$

Each department is offered a grade

$$\pi_i^o = \lambda_i R_i + (1-\lambda_i) D_i \tag{21}$$

where λ_i is given by (20). Note that $Var(\pi_i^o) = (\lambda_i^2 + (1-\lambda_i)^2)\sigma^2$. If we observe noisy signals $\tilde{R} + e$ and $\tilde{D} + v$ of R and D respectively, then we can write (21) as

$$\pi_i^o = \lambda_i \tilde{R}_i + (1-\lambda_i)\tilde{D}_i + \epsilon_i \quad (22)$$

where ϵ if a function of e and v and $\tilde{R}$ and $\tilde{D}$ are exogenous.

To motivate the probit model for the RAE exercise we note that π_i^o is not observable. The RAE committee only discloses a limited scale of grades. For illustration purposes we assume for the time being that the RAE committee only declares high or low quality. Thus, in a probit setting, the unobserved grade offered can be written as

$$\pi_i^o = \beta_0 + \lambda_i \tilde{R}_i + (1-\lambda_i)\tilde{D}_i + \epsilon_i \quad (23)$$

The intercept β_0 is needed for estimation purposes (Green, 2000, p.819). The disturbance term is heteroscedastic since $Var(\epsilon_i) = (\lambda_i^2 + (1-\lambda_i)^2)\sigma^2$, but this can be easily incorporated in the probit model by simply dividing (23) by $\sqrt{Var(\epsilon_i)}$.

In this simple binary setting, we do not observe π_i^o. Instead, what we do observe is the outcome 'high' (say $y = 1$) if $\pi_i^o > 0$ and 'low' if $\pi_i^o \leq 0$ (say $y = 0$).[4] The model can be easily generalised to multinomial outcomes.

To complete the probit model we need to specify the rule and discretion equations. Because of the limited number of observations, we have to keep the number of regressors to a minimum. As discussed earlier, we proxy the rule measure by AQI, pressure by Size$(= p)$, and discretion measure by Seniority. Thus, the complete model consists of (23) and (20) and the equivalent of (1) and (2)

$$\begin{aligned} \tilde{R}_i &= \beta_1 \text{AQI}_i \\ \tilde{D}_i &= \beta_2 \text{Seniority}_i + \beta_3 \omega \text{Size}_i \end{aligned}$$

The parameters of interest are γ (the optimality parameter), ω (the social/political sensitivity of the committee), σ^2 (the noise of the signals R and D), and β_i, $i = 0, 3$. To preserve compatibility with the theoretical model we standardised the three regressors in order to have identical scale.

The estimated model is shown in Table 6. All parameters are significant and have the expect signs. The suggested model fits the data satisfactorily, and with a log-likelihood value very similar to the data driven model, despite using three regressors only.

The estimated model indicates that the RAE panel is not highly sensitive to political or social pressure, but the significance of ω clearly indicates some level of sensitivity of the RAE committee to social pressure. The significance of β_3 also points to the political sensitivity of the RAE panel.

Contrary to expectations, however, the RAE panel appears to have used a slightly higher than expected γ to form its optimal estimator $\hat{Q}$. However, even though the estimate of 0.674 for γ is highly significant, the 95% confidence interval is [0.192,1.156] and would therefore not reject the null hypothesis $\gamma = 0.5$ at the 5% level of significance. A more convenient way of testing the null hypothesis $\gamma = 0.5$ is to use $\gamma = \exp(\eta)/[1+\exp(\eta)]$ and estimate η instead. This way, the null $\gamma = 0.5$ is expressed as $\eta = 0$. The estimation yielded an estimate $\hat{\eta} = 0.828$, which implies $\hat{\gamma} = 0.696$, which is close to the value estimated directly. However, the t-statistic of $\hat{\eta}$ was 1.83, suggesting insignificance of $\hat{\eta}$. This was also supported by a likelihood ratio statistic close to zero. Thus, the RAE panel is likely to have put equal weights on both rule and discretion measures. The parameters associated with rules, discretion and pressure are all significant and imply a positive impact of these regressors on the probability of a higher grade.

[Table 6 about here]

The level of discretion used varies greatly from one department to another. The estimated levels of discretion are given in Table 7. The lowest values for λ were 0.04 (Oxford), 0.07 (LSE) and 0.08 (Cambridge). As expected, these universities appear to enjoy the highest levels of discretion. The highest values for λ were around 0.71 for Birkbeck, Loughborough and Queen Mary. The average value is 0.433, indicating a higher use of discretion on average.

The predictive performance of this theory based model suggests its superiority over the data driven model. First, it has six incorrect predictions compared with seven in the data driven model. More

[4] We may think of this as the RAE panel knowing the grade but choosing to give a discrete grading of low and high.

importantly, five of these are 3 to 4 rating mismatch. Only Bristol is (again) predicted a 5 but was given a 4 rating. The detailed predicted probabilities are given in Table 7.

[Table 7 about here]

8 Conclusion.

This paper presents a model of committee decision-making when both rules and discretion are available. Some empirical evidence on the performance of economics departments in the RAE 2001 is offered. Among the predictions of our model are that higher pressure always increases the level of discretion, while increased noise of observed signals increases the level of rules. The sensitivity of committees to political and social pressure can have both negative and positive impact on the level of discretion, depending on the sensitivity of the committee and the noise to pressure ratio. The RAE 2001 result was related to a combination of proxies for rules, pressure and discretion using two probit models. The results suggest significance of both rules and discretion. Pressure was found to matter in the determination of department's assessment of research quality.

Although we did not obtain perfect predictions most grades were correctly predicted by both the data driven and the discretion models. However, despite using only a fraction of explanatory variables, the discretion model produced a more satisfactory prediction than the data-driven model. This lends more credence to our model.

In the data driven model, even though the official line was to consider both funding and PhD students only the ratio of PhD to staff was found to be significant. Working papers appear to have a neutral effect, but other forms of research output, such as books have a detrimental impact on the grading of departments.

The results show that the flexibility of a discretionary system may come at a price. In our case, the RAE panel was clearly biased in favour of larger departments. In fact, based on the marginal impact, size was almost as influential as the average quality metrics. However, the biggest marginal impact was by far due to our hybrid proxy for discretion. The presence of discretion also appears to have been exercised so as to go against the official standing that rankings are based on the proportions of staff. The insignificance of the Gini coefficient indicates that it is more likely that the overall department average performances influenced the panel's decision, even though it should not have been so.

The discretion model was tested empirically and was found to fit well with real life data. The estimation confirms the expected impact of rules, discretion and pressure. The results also suggest that the panel was not highly sensitive to political and social pressure. Finally, contrary to expectation, the panel appears to have used OLS rule as the optimal estimator for quality.

The report by Sir Gareth Roberts (2003) proposed the introduction of quality metrics, but their use was left to the 'discretion' of the RAE panel. However, the funding bodies have recently decided to continue with expert judgement as the official method of evaluation of departments. In this respect, it is perhaps more appropriate to be surprised that the 2001 RAE panels could have used rules at all as suggested by the significance of our proxies for rules. Still, there is always the possibility that the panel's discretionary view coincided with certain rule measures. However, our model of discretion and pressure suggests that this would be unlikely.

References

Axarloglou, K. and Theoharakis, V. (2003) "Diversity in economics: An analysis of journal quality perceptions," http://ssrn.com/abstract_id=379120.

Baltagi, B. (1999) "Applied Econometrics Rankings: 1989-1995," *Journal of Applied Econometrics*, 14, 423-441.

Beed, K. and C. Beed. (1996) "Measuring the Quality of Academic Journals: The Case of Economics," *Journal of Post Keynesian Economics*, 18, 3, 369-396.

Beyer, J. J., and Trice, H. M. (1978) *Implementing change: alcoholism policies in work organizations.* New York: Free Press.

Blackburn, K. and Christensen, M. (1989) "Monetary Policy and Policy Credibility," *Journal of Economic Literature*, 27, 1-45.

Boston, J. (2002) "The Purpose of the Research Assessment Exercises in Britain and Hong Kong," Briefing Note for the PBRF Technical Working Group. URL: www.minedu.govt.nz/web/downloadable/dl7503_v1/purpose-of-rae-in-uk-and - hk.doc.

Bourgeois, L. J., & Brodwin, D. R. (1984) "Strategic implementation: five approaches to an elusive phenomenon," *Strategic Management Journal*, 5, 241-264.

Burkitt, B. and Baimbridge, M. (1995) "Publication performance in The Economic Journal and UFC research rankings," *Applied Economics*, 27, 397–402.

Burton, M. and Phimister, E. (1995) "Core Journals: A Reappraisal of the Diamond List," *The Economic Journal*, 105, 361-373.

Campbell, K., Vick, D.W., Murray, A.D., and Little, G.F. (1999) "Journal publishing, journal reputation, and the United Kingdom's Research Assessment Exercise," *Journal of Law and Society*, 26, 470-501.

Clarida, R., Gali, J. and Gertler, M. (1999) "The Science of Monetary Policy", *Journal of Economic Literature*, 37 (4), 1661-1707.

Conroy, M.E., Dusaksky, R., Drukker, D., and A. Kildegaard (1995) "The productivity of Economics Departments in the U.S.: Publications in the Core Journals," *Journal of Economic Literature*, 33, 1966-1971.

Dusansky, R. and Vernon, C. J. (1998) "Rankings of U.S. economics departments," *Journal of Economic Perspectives*, 12, 1 (Winter), 157-70.

Eisenhardt, K.M. (1989) "Agency theory: an assessment and review," *Academy of Management Review*, 14(1), 57-74.

Fischer, S. (1990) "Rules Versus Discretion in Monetary Policy," in B. Friedman and F. Hahn (eds.), Handbook of Monetary Economics, Amsterdam: North-Holland.

Green, W.H. (2000), *Econometric Analysis*, Prentice Hall: New Jersey.

Harris, G. (1990) "Research Output in Australian Universities Economics Departments: an Update for 1984-1988," *Australian Economic Papers*, 249-259.

Harzing, Anne-Wil (2001) "Journal Quality List", http://www.harzing.com .

Kalaitzidakis, P., Mamuneas,. T. and Stengos, T. (1999) "European Economics: An Analysis Based in Publications in the Core Journals," *European Economic Review*, 43, 1150-1168.

Knight, P., (2002) "RAE review: Gladiatorial contest is the victim of its own success," THES, 6 December.

Kydland, F.E. and Prescott, E.C. (1977) "Rules Rather than Discretion: The Inconsistency of Optimal Plans," *Journal of Political Economy*, 85, 473-492.

Lewis, J. (2003) "Why I think the RAE review needs to go back to basics," Times Higher Educational Supplement, 10 January.

Liner, G.H. (2002) “Core Journals in Economics,” *Economic Inquiry*, 40, 1, 138-145.

Lucas, R. (1995) “Contributions to Economics Journals by the Canadian Economics Profession, 1981-1990,” *Canadian Journal of Economics*, 28, 4, 949-960.

Majumdar, S. K. (1997) “Incentive regulation and productive efficiency in the U.S. telecommunications industry,” *Journal of Business*, 70, 4, 547-576.

Majumdar, S. K. and Marcus A. (2001), “Rules Versus Discretion: The Productivity Consequences of Flexible Regulation”, *Academy of Management Journal*, 44, 1, 170-179.

Marcus, A. (1988) “Implementing externally induced innovations: a comparison of rule-bound and autonomous approaches,” *Academy of Management Journal*, 31, 2, 235-256.

Roberts, G. (2003) “Review of research assessment: final report”, http://www.ra-review.ac.uk.

Seglen, P.O. (1997) “Why the Impact Factor of Journals Should not be Used for Evaluating Research,” *British Medical Journal*, February, 314, 497.

Sitkin, S. B., & Bies, R. J. (1994) *The legalistic organization.* Thousand Oaks, CA: Sage Publications.

Strebel, P. (1987) “Organizing for innovation over an industry cycle,” *Strategic Management Journal*, 8, 117-124.

Taylor, J.B. (1993), “Discretion versus policy rules in practice,” Carnegie - Rochester Conference Series on Public Policy, 39, 195-214.

Theoharakis, V. and Hirst, A. (2002) “Perceptual Differences of Marketing Journals: A Worldwide Perspective,” *Marketing Letters*, 13, 4, 389-402.

Vick, D.W., Murray, A.D., Little, G.F., and Campbell, K. (1998) “The perception of academic lawyers concerning the effects of the United Kingdom’s Research Assessment Exercise,” *Journal of Law and Society*, 25, 554-560.

von Tunzelmann, N. and Mbula, E.K. (2003) “Changes in research assessment practices in other countries since 1999,” Annex G, Roberts’ Report (2003).

Table 1. Journals with 10 or more submissions in RAE 2001.

Title	Total Papers	Title	Total Papers
Economic Journal	118	Review of Economics and Statistics	17
Manchester School	77	Weltwirtschaftliches Archiv	17
Economics Letters	74	Applied Economics Letters	16
Oxford Economic Papers	64	Bulletin of Economic Research	16
Economica	55	Journal of Comparative Economics	16
European Economic Review	55	Journal of Development Studies	16
Journal of Public Economics	55	British Journal of Industrial Relations.	15
Review of Economic Studies	53	Economic Modelling	15
Journal of Econometrics	51	Journal of Industrial Economics	14
Oxford Bulletin of Economics and Statistics	51	Journal of Post Keynesian Economics	14
Journal of Economic Theory	48	Oxford Review of Economic Policy	14
Scottish Journal of Political Economy	47	International Journal of Industrial Organization	13
Applied Economics	44	Journal of Agricultural Economics	13
Econometrica	40	Journal of International Development	13
Games and Economic Behavior	32	Journal of Labor Economics	13
World Development	31	Journal of Population Economics	13
Journal of Political Economy	30	RAND Journal of Economics	13
Journal of Development Economics	27	European Journal of Political Economy	12
Quarterly Journal of Economics	25	International Journal of Finance and Economics	12
International Economic Review	24	Journal of Economic Behavior and Organization	12
Journal of Health Economics	24	Journal of Monetary Economics	11
Journal of Applied Econometrics	23	Public Choice	11
Journal of Economic Dynamics and Control	21	Scandinavian Journal of Statistics	11
Journal of Business and Economic Statistics	20	Applied Financial Economics	10
Journal of International Economics	20	Canadian Journal of Economics	10
Cambridge Journal of Economics	19	Health Economics	10
Econometric Theory	19	History of Political Economy	10
American Economic Review	18	Journal of Macroeconomics	10
Economic Theory	18	Labour Economics	10
Journal of Mathematical Economics	18	Scandinavian Journal of Economics	10
Regional Studies	18		

Table 2. Summary Statistics of Main Variables.

Series	Mean	Std Error	Minimum	Maximum
Size	79.39	48.02	26	236
Seniority	7.049	7.906	0	35
AQI	0	1	-1.958	1.771
Working papers	0.121	0.091	0	0.388
Other output	0.392	0.182	0.096	0.714
Funding	48.595	55.449	0	254.048
PhD	0.938	0.562	0	2.004
Gini coefficient	0.466	0.1	0.271	0.668

Table 3. Estimation Results for the Data-Driven Model.

	Unrestricted Model			**Selected Model**		
	Parameter Value	t-statistic	p-value	Parameter Value	t-statistic	p-value
Intercept	0.986	3.024	0.002	0.550	1.720	0.085
Threshold	4.401	7.893	0.000	4.428	8.775	0.000
Size	0.007	1.483	0.138	0.011	2.271	0.001
Seniority Level	0.265	3.738	0.000	0.224	3.396	0.002
AQI	1.063	2.743	0.006	1.093	3.135	0.002
Working papers	-2.029	-0.918	0.358			
Other (Ref. & Non-ref.)	-3.909	-5.048	0.000	-3.688	-4.656	0.000
PhD	1.424	4.199	0.000	1.312	3.938	0.000
Funding	-0.004	-0.509	0.610			
GINI	-0.028	-0.037	0.970			
	Log Likelihood: -10.902			Log Likelihood: -12.247		

Table 4. Predicted Probabilities for RAE Grades (Data-Driven Model).

University	P(3)	P(4)	P(5)	Predicted Rating	Actual Rating
1. Birkbeck College	0.0000	0.0479	0.9521	5	5
2. Brunel University	0.0001	0.7505	0.2494	4	4
3. City University	0.9637	0.0364	0.0000	3	3
4. Keele University	0.6497	0.3503	0.0000	3	3
5. London Guildhall University	0.4887	0.5113	0.0000	4	3
6. LSE	0.0000	0.0000	1.0000	5	5*
7. Loughborough University	0.8031	0.1969	0.0000	3	3
8. Manchester Metropolitan University	0.9947	0.0053	0.0000	3	3
9. Queen Mary, University of London	0.0000	0.5857	0.4143	4	5
10. Royal Holloway, University of London	0.0024	0.9435	0.0541	4	4
11. University College London	0.0000	0.0000	1.0000	5	5*
12. University of Aberdeen	0.8068	0.1932	0.0000	3	3
13. University of Birmingham	0.0566	0.9412	0.0022	4	4
14. University of Bristol	0.0000	0.3811	0.6189	5	4
15. University of Cambridge	0.0000	0.0073	0.9928	5	5
16. University of Dundee	0.7906	0.2094	0.0000	3	3
17. University of Durham	0.0031	0.9511	0.0459	4	4
18. University of East Anglia	0.0169	0.9725	0.0106	4	4
19. University of East London	0.9859	0.0142	0.0000	3	3
20. University of Edinburgh	0.0031	0.9516	0.0453	4	4
21. University of Essex	0.0000	0.0010	0.9990	5	5*
22. University of Exeter	0.0000	0.0931	0.9069	5	5
23. University of Glasgow	0.0757	0.9229	0.0014	4	4
24. University of Kent at Canterbury	0.2652	0.7347	0.0001	4	4
25. University of Leicester	0.0000	0.5198	0.4802	4	5
26. University of Liverpool	0.2935	0.7064	0.0001	4	4
27. University of Manchester	0.0001	0.7235	0.2765	4	4
28. University of Newcastle	0.0000	0.4703	0.5297	5	4
29. University of Northumbria at Newcastle	0.9999	0.0001	0.0000	3	3
30. University of Nottingham	0.0000	0.0293	0.9708	5	5
31. University of Oxford	0.0000	0.0000	1.0000	5	5
32. University of Sheffield	0.9441	0.0559	0.0000	3	3
33. University of Southampton	0.0000	0.1911	0.8089	5	5
34. University of St Andrews	0.0107	0.9726	0.0168	4	4
35. University of Stirling	0.4393	0.5607	0.0000	4	4
36. University of Strathclyde	0.0143	0.9731	0.0126	4	4
37. University of Surrey	0.2531	0.7468	0.0001	4	3
38. University of Sussex	0.4504	0.5496	0.0000	4	4
39. University of Wales, Swansea	0.7145	0.2856	0.0000	3	4
40. University of Warwick	0.0000	0.1062	0.8938	5	5*
41. University of York	0.0000	0.0213	0.9787	5	5

Table 5.Marginal effects on probabilities of RAE rating.

	Mean	Min.	Max.	Change in P(RAE=3)	Change in P(RAE=4)	Change in P(RAE=5)	Max. Impact P(RAE=5)
Size	79.390	26	236	-0.0001	-0.0010	0.0011	0.1723
Seniority Level	7.049	0	35	-0.0020	-0.0204	0.0223	0.6233
AQI	0	-1.958	1.771	-0.0096	-0.0995	0.1091	0.1932
Other Output	0.392	0.096	0.714	0.0323	0.3360	-0.3683	-0.1186
PhD	0.938	0.000	2.004	-0.0115	-0.1195	0.1310	0.1396

Table 6. Estimation results (Discretion model).

	Parameter Value	t-statistic	p-value
Intercept (β_0)	0.213	13.365	0.000
Threshold	5.560	8.808	0.000
AQI (β_1)	0.227	8.600	0.000
Seniority Level (β_2)	0.262	4.811	0.000
Size (β_3)	0.827	2.247	0.025
σ^2	0.006	4.478	0.000
ω	0.138	4.049	0.000
γ	0.674	2.742	0.006
	Log Likelihood: -12.905		

Table 7. Predicted Probabilities for RAE Grades (Discretion Model).

	Lambda	P(3)	P(4)	P(5)	Predicted Rating	Actual Rating
1. Birkbeck College	0.71776	0.0000	0.0636	0.9364	5	5
2. Brunel	0.47360	0.0000	0.8054	0.1946	4	4
3. City	0.28182	0.9462	0.0538	0.0000	3	3
4. Keele	0.43898	0.7277	0.2723	0.0000	3	3
5. London Guildhall	0.40620	0.3797	0.6203	0.0000	4	3
6. LSE	0.07608	0.0000	0.0000	1.0000	5	5*
7. Loughborough	0.71886	0.8803	0.1197	0.0000	3	3
8. Manchester Metropolitan	0.42783	0.9582	0.0418	0.0000	3	3
9. Queen Mary	0.71568	0.0000	0.4580	0.5421	5	5
10. Royal Holloway	0.48548	0.0000	0.8669	0.1331	4	4
11. University College London	0.30279	0.0000	0.0000	1.0000	5	5*
12. Aberdeen	0.63980	0.8271	0.1730	0.0000	3	3
13. Birmingham	0.63980	0.1161	0.8839	0.0000	4	4
14. Bristol	0.61545	0.0000	0.4267	0.5734	5	4
15. Cambridge	0.08996	0.0000	0.0008	0.9992	5	5
16. Dundee	0.45033	0.5732	0.4268	0.0000	3	3
17. Durham	0.42783	0.0000	0.9097	0.0903	4	4
18. East Anglia	0.45033	0.0024	0.9945	0.0031	4	4
19. East London	0.25458	0.9012	0.0988	0.0000	3	3
20. Edinburgh	0.39573	0.0002	0.9791	0.0207	4	4
21. Essex	0.39346	0.0000	0.0000	1.0000	5	5*
22. Exeter	0.47360	0.0000	0.3537	0.6463	5	5
23. Glasgow	0.60649	0.0004	0.9857	0.0139	4	4
24. Kent at Canterbury	0.54636	0.0297	0.9703	0.0000	4	4
25. Leicester	0.59479	0.0000	0.3614	0.6386	5	5
26. Liverpool	0.25458	0.6203	0.3797	0.0000	3	4
27. Manchester	0.25964	0.0000	0.8637	0.1363	4	4
28. Newcastle	0.33806	0.0000	0.8403	0.1597	4	4
29. Northumbria at N.	0.24217	0.9969	0.0032	0.0000	3	3
30. Nottingham	0.19392	0.0000	0.0022	0.9978	5	5
31. Oxford	0.04007	0.0000	0.0000	1.0000	5	5
32. Sheffield	0.59479	0.8298	0.1702	0.0000	3	3
33. Southampton	0.33612	0.0000	0.0085	0.9915	5	5
34. St Andrews	0.61794	0.0000	0.9792	0.0208	4	4
35. Stirling	0.43898	0.6184	0.3816	0.0000	3	4
36. Strathclyde	0.67776	0.2521	0.7480	0.0000	4	4
37. Surrey	0.61794	0.1900	0.8100	0.0000	4	3
38. Sussex	0.63980	0.9322	0.0678	0.0000	3	4
39. Wales, Swansea	0.50962	0.1859	0.8141	0.0000	4	4
40. Warwick	0.21318	0.0000	0.1449	0.8552	5	5*
41. York	0.15490	0.0000	0.0001	0.9999	5	5