



Deposited via The University of Sheffield.

White Rose Research Online URL for this paper:

<https://eprints.whiterose.ac.uk/id/eprint/176109/>

Version: Accepted Version

Article:

McManus, R., Mumford, K. and Sechel, C. (2022) Measuring research excellence amongst economics lecturers in the UK. *Bulletin of Economic Research*, 74 (2). pp. 386-404. ISSN: 0307-3378

<https://doi.org/10.1111/boer.12299>

This is the peer reviewed version of the following article: McManus, R, Mumford, K, Sechel, C. Measuring research excellence amongst economics lecturers in the UK. *Bull Econ Res*. 2021, which has been published in final form at <https://doi.org/10.1111/boer.12299>. This article may be used for non-commercial purposes in accordance with Wiley Terms and Conditions for Use of Self-Archived Versions.

Reuse

Items deposited in White Rose Research Online are protected by copyright, with all rights reserved unless indicated otherwise. They may be downloaded and/or printed for private study, or other acts as permitted by national copyright laws. The publisher or other rights holders may allow further reproduction and re-use of the full text version. This is indicated by the licence information on the White Rose Research Online record for the item.

Takedown

If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing eprints@whiterose.ac.uk including the URL of the record and the reason for the withdrawal request.

Measuring research excellence amongst economics Lecturers in the UK.

Richard McManus¹, Karen Mumford^{2,3} and Cristina Sechel⁴

¹Christ Church Business School, Canterbury Christ Church University, UK.

²Management School, University of York, UK.

³IZA, Institute for the Study of Labour, Germany.

⁴Department of Economics, University of Sheffield, UK.

February 2021.

Abstract

Using a rich new data source, we explore the selection of economics Lecturers into the last UK Research Excellence Framework (REF) exercise. Only some one-in-two (54%) of these Lecturers were submitted to REF2014; 57% of men and 46% of women. The decision making of Institutions is found to be well approximated by a simplified selection approach; focusing on working papers and higher quality journal publications. Our results also reveal sizeable conditional differences in the probability of selection, especially so in departments with higher research rankings. More than half of the variance in selection probability remains unexplained, revealing considerable idiosyncrasies in the management of submissions and uncertainty across the discipline in this research assessment process.

Word count in paper: 9316 including authors, tables and references; 25 pages (plus an additional online appendix of 4 pages and 763 words).

Key words: REF; gender; selection; outputs; quality; inclusivity.

Corresponding author: Prof. Karen Mumford, karen.mumford@york.ac.uk Management School, University of York. Heslington York YO10 5DD, UK. (phone +44 904 323756).

1. Introduction

Several countries have introduced systems to evaluate the research produced by their university sector using peer review and/or metric based criteria (Regibeau and Rockett, 2016; Stern, 2016). For example, in 2014, more than two-thirds of the OECD countries were formally evaluating the performance and contribution of their Higher Education Institutions (HEIs), with most of these appraisal systems having been established since the turn of the century (OECD, 2018; page 189). No particular research assessment system has been commonly adopted, however, and a range of alternative criterion are instead used across nations (OECD, 2018; page 212).

The UK has the longest running of these appraisal exercises beginning with the Research Selectivity Exercise in 1986. In its current incarnation, the UK Research Excellence Framework (REF) incorporates a range of measures but focusses primarily on the quality of publications of individual staff members to judge the research quality of academic institutions. The REF ranking process is crucially important to the prestige of UK universities (Hicks, 2012). The great majority of government research funding in this sector is directly based on REF outcomes: some £1.6bn in 2017/18 (HEFCE, 2017). REF rankings also play a critical role in the public perception of departments, with university league tables believed to strongly influence under-graduate student enrolments, post-graduate recruitment and overall institution prestige (De Fraja et al., 2020). Increasing demands on public finance and the need for accountability suggest that this trend will continue in the future. Potential bias in research assessment exercises such as the REF is of major concern to both academics and research funding councils.¹

Up to, and including, the last REF exercise in 2014 (REF2014), institutions could choose which academic staff to submit for assessment. For individual academics, the inclusion, or not, in the institution's REF submission is a major indicator of success at producing high quality research outputs with implications for self-esteem and peer recognition. Who was included in REF2014, and the outputs that are submitted, is general knowledge recorded in the public domain (<http://results.ref.ac.uk/>). The longer-term career prospects associated with inclusion (or exclusion)

¹ Clerides et al. (2011) find no evidence of institutional bias in the 1996 and 2001 Research Assessment Exercise (RAE) in UK, and Bertocchi et al. (2015) find no statistical bias between the peer review process used in the Italian research assessment exercise and bibliometric evaluation of research quality.

are debateable. In the simplest scenario, for example, there may be no repercussion for the individual from inclusion (or exclusion); implying that their reputation, resource allocation and workload allocation all remain unaffected. Given the additional resources available to the institution from a successful across unit REF entry, other explanations are possible. One might expect academics with a strong REF research reputation to be allocated more resources, including time, to dedicate to future research projects.² Virtuous circles can arise when positive recognition for an individual motivates them to produce greater effort, further increasing the support of their peers, and leading to additional successes for the selected (the “Matthew” effect: Merton, 1968) even if the original assessment was itself spurious. Levin and Stephan (1991) show that these life-cycle effects are especially sizeable when recognition is granted to young academics. Although the internal selection processes followed by specific HEIs for deciding which academics to include in REF2014 are not published (and are often not fully known even within the institution), the incentives were structured to submit those Lecturers with the best research outputs.

It can be argued, however, that simply submitting staff members with the highest-ranking journal publications may not be the first best rule for institutional REF ranking success. Stern (2016, page 11) discusses ‘negative and perverse incentives’ for institutions in the REF2014 process. Gaming with respect to recruitment policies and the selection of which academics are entered may have critical implications for research design, institutional policies, and the academic labour market. Strategic employment for the purposes of boosting REF scores can negatively impact on investment and incentives for staff development and encourage rent-seeking behaviour among academics. This may encourage research output towards the measured dimensions that are included in the REF system. As some elements of research excellence may be difficult to measure, these types of changes may reduce overall research quality in the long term (Regibeau and Rockett, 2017). The potential linkage between formal recognition of research ability and the career paths of academics implies that it is important for the REF recognition scheme to accurately reflect research quality for individual academics rather than strategic decision making across institutions.

Differential recognition of research quality across genders is another concern with

² Gould (2002) further argues there is a positive reputation (and positive feedback loop) for those simply associating with those selected.

potentially negative consequences for current, and longer-term, gender equity in academic institutions. HEFCE (2015) found that females were less likely to be entered into REF2014 and that this was true, on average, across all disciplines. The issue of gender inequality in research grading exercises is further raised in Rees (2011) who discusses possible gender differences in selection, peer review and research design. Bayer and Rouse (2016) highlight that economics, in particular, is an academic discipline behind others with respect to equality and diversity concerns (see also Wu 2017; Lundberg and Stearns, 2019; Mumford and Sechel, 2019 and 2020).

If research outputs of women are less valued (Brown and Yang, 2015; Hengel, 2018) or the role of women in producing that research is undervalued (Sarsons, 2017), we might expect women to be less likely to be included in the REF implying potential negative resource allocation and feedback loops for female academics (Reskin and Hargens, 1979; Long and Fox, 1995). Furthermore, findings that social science students (male and female) rate female Professors lower in student evaluations suggest that women need to put more effort into their teaching to match male rankings (Boring, 2017; Mengel et al. 2018). The greater the proportion of teaching in the workloads of females, the larger is the potential relative disadvantage created by this rating difference.

Using a rich new data source, collected by the authors, our paper examines the relationship between research outputs and REF inclusion for economics Lecturers in the UK, paying particular attention to the quality and quantity of these outputs. Lecturers are typically the least experienced of those on standard academic contracts (combined administration, research and teaching) and they may be considered to have future career paths that are the most vulnerable to inclusion or exclusion from REF submission. Lecturers are also less likely to have their selection affected by having existing REF defined research impact³ on the wider non-academic world or having the ability to influence their institution's research environment. We focus our study on this academic grade

³ There is a small but growing literature using case study based analyses to consider how research impact is measured in the REF framework. In general, this literature tends to be critical of the framework adopted in REF2014, on the grounds of: monetary costs (Martin 2011); distorting incentives (Trevorrow and Volmer 2012; and Hicks 2012); potential threat to academic freedom and autonomy (Smith et al. 2011), consistency of individual reviewers on unit of assessment panels judging the quality of research (Tymms and Higgins 2018); difficulties with the efficiency and efficacy of measuring 'impact' (Ovseiko et al. 2012; Parker and van Teijlingen 2012); and implications from negative coverage in the media (Murphy and Sage 2014).

accordingly.⁴

Previous studies of research evaluation exercises largely focus on measuring the cost-effectiveness and/or validity of institutional rankings (Bertocchi et al., 2015; Geuna and Piloatto, 2016; Regibeau and Rockett, 2016; Tymms and Higgins, 2018). De Fraja et al. (2020) use Higher Education Statistical Agency (HESA) data to explore the relationship between department REF scores and the salaries of full Professors, finding higher REF ranked departments pay their Professors more on average and also have higher within department wage inequality. De Fraja et al. (2020) do not, however, have information on specific Professors including own research output or gender. At the individual level, HEFCE (2015) provides statistical analysis of inclusion in REF2014 that simultaneously controls for personal characteristics including gender, but does not have information on quality or quantity of outputs. To the best of our knowledge, no previous study explores the relationship between research outputs and inclusion in research evaluation exercises for individual academics.⁵

The remainder of the paper is structured as follows: the methodology and data are described in section 2 of the paper, including a detailed explanation of the REF process; section 3 considers the estimation results for the probability of being selected for REF2014; section 4 considers alternative assumptions and further discussion; and section 5 presents conclusions.

2. Methodology and data

We are interested in the probability that an individual Lecturer is selected for the REF2014 process conditional on a range of observable characteristics expected to explain that probability. More formally, a series of probit regressions are estimated, with the unobserved latent dependent variable (the propensity to be submitted to the REF2014 exercise, S_i) set equal to 1 if the individual Lecturer i was submitted to the REF and zero otherwise.

⁴ Some three-quarters of the economics Professors were entered into REF2014 (fractionally more of the female Professors than the male), a little over 60% of the Readers/Senior Lecturers were submitted, and only 54% of the Lecturers. The gender gap at the Lecturer level is the largest at 11 percentage points.

⁵ Our study can be related to Greenhalgh and Fahy (2015) who consider the assessment of Impact Case Studies for the Health Sciences discipline in the REF2014 process; and Hole (2017) who uses the REF2014 outcome rankings to retrospectively assess the quality of specific economics journals

$$Pr(S_i = 1) = \theta(\beta X_i) \tag{1}$$

where X_i is a vector of explanatory variables and θ is the standard normal distribution function (Maddala 1992; 327).⁶ We might expect this probability to be primarily dominated by the number and quality of research outputs the Lecturer has. It is, however, important to also allow for the features of the REF2014 rules.

2.1 The REF process

REF2014 was a process of peer review assessing the quality of research across the UK higher education institutions (HEIs). At the national level, 36 Units of Assessment (UoA) each represented a separate research discipline and had its own sub-panel carrying out assessment under the guidance of four main panels. Sub-panels could also refer submissions to other sub-panels for advice. HEIs were required to submit a “common set of data” to each relevant UoA for those research active staff in post on the census date (October 31st 2013) that the institution had selected for inclusion in the REF.

A typical Lecturer in the UK is on a combined teaching and research (or Category A) contract. For the purposes of REF2014, ‘Category A staff are defined as academic staff with a contract of employment of 0.2 FTE or greater and on the payroll of the submitting HEI on the census date [...] and whose primary employment function is to undertake either ‘research only’ or ‘teaching and research’ (REF 2012; paragraph 78). Academics that do not have a research requirement in their contracts, such as those on teaching only contracts, were not eligible for entry in the REF2014 process. There is no data available outside of institutions as to which Lecturers are on a teaching only contract, although there is evidence that these types of contracts have become more common in academia over the last decade (HESA 2020). Contractual status will be further explored in Section 4 below.

⁶ Using the Vuong (1989) test, we fail to reject the assumption of normality used in the probit model in favour of the logistic alternative employed by the logit model in the analysis below. We therefore present results using the probit model.

The “common set of data” required from HEIs included details of publications (and other forms of assessable output) submitted staff had produced between January 1st 2008 and December 31st 2013, up to a maximum of four outputs for each staff member. However, allowances were made for Lecturers with shorter tenure; Early Careers Researcher status (ECRs), as defined by REF2014, were those Lecturers who had taken a Category A contract on or after the 1st of August 2009. ECRs could be submitted to the REF with fewer than four outputs without penalty according to a scale linked with tenure as an ECR. The tapering of this scale is steep; basically each year of tenure requires an extra output for REF2014 inclusion up to the fourth year.

HEIs were told that the quality of submitted outputs would be assessed in terms of “originality, significance and rigour” with reference to international quality standards. We focus on those departments who primarily submitted to the Economics and Econometrics UoA (HEIs submitting to this UoA are listed in Table A1 of the Online Appendix).⁷

The need for HEIs to explicitly consider issues of gender equality, diversity and the transparent selection of staff for inclusion into REF2014 was stressed at all stages (REF 2012; Berry 2012). Each HEI was required to develop and document how they implemented a fair and transparent selection process across staff.

2.2 Sample selection

It is clearly a gain for a study such as ours to have information on all of the relevant Lecturers rather than only a subsample which may suffer from some form of selection bias. From 2008, the Royal Economics Society Women’s Committee Survey (RES-WCS) has harvested annual information from university department webpages on individual academic staff including grade of employment, gender, new hires and job leavers (Mitka et al., 2015). These survey entries are emailed biennially to respective Heads of Departments for verification⁸, in 2014 some 88% of the

⁷ Submitting to a different UoA was very rare amongst Lecturers in these departments: we found only one that had (to Business and Management). All of the other Lecturers in these departments were either submitted to the Economics and Econometrics UoA or they were not submitted at all.

⁸ Contact details for the Heads of Departments are obtained from CHUDE (the Conference of Heads of University Departments of Economics). CHUDE was established by the Royal Economic Society (RES) in 1987 in collaboration with the Association of University Teachers of Economics.

institutions responded. The RES-WCS does not, however, include information on research activity or outputs.

The number of outputs, type of output (journal publications, books, book chapters, working papers or conference proceedings), and journal publication information for Lecturers submitted to the REF2014 was taken from the REF submission records. For those Lecturers not submitted, the authors collected detailed assessable output and journal publication data including the date of their first publication, the number of outputs in the period eligible for the REF2014, and the journals of their publications; where the maximum number of outputs is capped at 4 with inclusion preference given to their highest quality journal publications.

The quality of each individual journal publication was measured using the Keele journal ranking list (Hudson 2013). There is no official list of journal quality used across economics departments, nor is there a commonly accepted list.⁹ Examples of alternative lists are provided in Hudson (2013) and the Association of Business Schools ('ABS'), we considered these alternatives and found the choice of list made little difference to the results in preliminary analysis (results available upon request). We choose to focus on the Keele list as it was widely available during the planning phases of the REF2014 exercise and was well known in economics departments. The Keele list provided a four point ranking scale for journals ranging from 1 for nationally recognised, 2 for internationally recognised, 3 for internationally excellent, to 4 for world leading.

The full dataset contains 322 Lecturers but 17 individuals, 9 male, were excluded from the analysis because they were submitted to the REF2014 with fewer than the required outputs relevant for their researcher status. These 17 Lecturers may have had compassionate allowances, for example, for periods of ill health or maternity leave. However, we do not have this additional information to adjust REF output requirements accordingly and we exclude them from the sample. There are also 5 Lecturers (all male) in our data who were not submitted to REF2014 and who show no evidence of being research active prior to the REF entry cut-off and are therefore expected to be on teaching only contracts. We further exclude these 5 and, consequently, a total of 22

⁹ The Keele list has subsequently been amalgamated with the ABS list to form the Chartered Association of Business Schools Academic Journal Guide 2018.

individuals are left out of the analysis, leaving an estimation sample of interest of 300 Lecturers (83 females and 217 males).

Considering the sample of interest in more detail (see Table 1), on average, women make up 28% of these Lecturers, this is consistent with the historic under representation of women in UK academic economics (Mitka et al. 2015; Mumford and Sechel, 2020). A little over half (54%) of the Lecturers were submitted to the REF; 57% of the males but only 46% of the females. This is a substantial unconditional gender gap in the probability of being submitted to REF2014, without allowing for any conditional factors women are 11 percentage points (pp) less likely to be selected into the REF process than are males.

[Table 1 around here]

The RAE grade is the department's awarded score in the 2008 Research Assessment Exercise (the immediately prior equivalent to the REF) and is included as a measure of the department's long term commitment to the research quality assessment programme. It may be the case that high ranked RAE departments place more priority on hiring Lecturers who are likely to qualify for future research excellence exercises, such as REF2014. These departments may also provide a more conducive environment to develop outputs suitable for submission. There were 4 points on the RAE 2008 outcome scale, ranging from 1 (quality that is recognised nationally) to 4 (quality that is world-leading) in terms of "originality, significance and rigour" (RAE, 2008). There was also the very rarely used "unclassified". The departments in our sample range between an average RAE2008 grade of 2.65 (Brunel University and City University) and 3.55 (LSE), with an overall department average of 3. There is very little difference between the average RAE department score for the male and female Lecturers.

The relationship between RAE rank and REF2014 submission is far from one-to-one for these departments. We can approximate each department's total REF submission rate (for all academic staff, not just Lecturers) by using their actual numbers submitted to REF2014 for the numerator and approximating total staff with the departmental totals provided by the RES-WCS data for the denominator. These department submission rates vary from 79% to 33% for

submissions only to the Economics and Econometrics Unit of Assessment (or 81% to 33% for submission including any Unit of Assessment).¹⁰ The correlation between the Economics and Econometrics REF2014 staff submission rate and RAE grade is 0.2906.¹¹ A positive relationship between RAE grade and the submission of Lecturers in REF2014 is expected.

The women tend to have more journal publications and fewer working papers. Females are also found to have a very slightly higher average quality of journal publications.¹² A positive relationship is expected between journal publication quality and REF submission. Assessable outputs other than journal publications have no explicit quality measure that is observable to the researcher, so an average journal quality measure for non-journal publications is assumed. For example, if a Lecturer has one journal publication rated at 3, one rated at 2, and two book chapters (i.e., with missing quality), the average output quality for this person would be 2.5. But if a Lecturer has all outputs as working papers, books, book chapters or conference proceedings their average output quality is assumed to be zero (we will return to further consider the implications of this assumption in section 4 below.)

The regional identifiers are for London, England excluding London, and Scotland. It is noteworthy that despite the REF being a national exercise, virtually all of the departments submitted in the Economics and Econometrics Unit of Assessment were from English universities. There were only four departments submitted from Scotland and none submitted from either Northern Ireland or Wales.

The RES-WCS data allow us to establish early career researcher (ECR) status according to the REF2014 criteria for all of the Lecturers working in the relevant departments at the REF2014

¹⁰ Highly varying submission rates across institutions is not an economics only phenomenon (HEFCE, 2015; De Fraja et al., 2020).

¹¹ In departments with a RAE grade of 3 or more, the staff submission rate to the Economics and Econometrics Unit of Assessment ranges from 38% to 79%; mean of 55. In departments with a RAE grade below 3, the staff submission rate to the Economics and Econometrics Unit of Assessment ranges from 33% to 77%; mean of 57.

¹² There is some evidence that females in our sample are more likely to have their publications in the 3 and above ranked journals, although there is no significant difference in the overall quality ranking between males and females. De Fraja et al. (2020) compare using a simple department grade point average for output quality (taken from REF records) with the more complex funding score formula that applies different weights to alternative quality grades, they find GPA is a better measure of department prestige and determinant of Professorial pay.

census date. We identify four groups accordingly; not ECR (NECR), late ECR (LECR), middle ECR (MECR), and early ECR (EECR).¹³ Whilst the LECR category has the highest proportion female (41%, see Table A2), the EECR has the greatest number of the women in it (48% of all of the women), perhaps not surprising when the numbers in this category make up more than half of the ECRs.¹⁴

We include within department co-author information for those Lecturers submitted, and those not submitted. Co-authorship in the same institution is not common in these departments (almost 90% of the Lecturers don't engage in it, see Table 1). Hamermesh (2015) argues that co-authorship is the positive outcome of a successful search process; he shows tenure is positively associated with a broader range of co-authors but not necessarily more co-authors per article. The relationship between tenure and co-authorship is not obvious in our sample of Lecturers. Co-authorship within the institution is almost twice as common amongst the LECR group (33.3% have at least one such paper) as the NECR (17.9%), and is rarer for the MECR (6.7%) or the EECR (5.7%). Women are more than twice as likely to have a co-author in the same department as males (19% and 9% respectively) but there do not appear to be obvious differences in the gender or rank of their co-authors¹⁵. Whilst there are many potential research enhancing aspects to co-authorship, it was also recognised that an easy way of gaming the REF would be to simply add the names of colleagues to publications. Sarsons (2017) further argues that published co-authored papers send a weaker signal of quality when considering promotions compared with single-authored work, and that this weakening in signal is felt much more by female co-authors than male. The guidance on how credit would be allocated for output co-authored within an institution is not clear in the REF 2014 guidelines (REF 2012: page 15, paragraph 76), nor in the previous RAE 2008 guidelines (RAE 2005: page 31, paragraph 102). It was consequently widely believed that output submitted

¹³ Detailed definitions and the REF2014 specified output requirements for each ECR category are provided in the Online Appendix Tables A2 and A3, detailed output information are provided in Tables A4 and A5.

¹⁴ ECRs represent 75% of all Lecturers submitted, although given the submission requirements, this represents 50% of all outputs submitted by lecturers in REF2014. As a share of all academic submitted to the REF, Lecturers represent less than one-fifth all those academics submitted with those are a higher seniority being submitted at a higher frequency (over two-thirds of Readers and Professors were submitted).

¹⁵ Both the female and male Lecturers are more likely to co-author with a male colleague within their department than a female colleague. Of those females who co-author with others in their school, 77% of these co-authors are male; for male Lecturers 78% of those who co-author internally do so with a male colleague. More than half of these co-authors of Lecturers are Professors, with 53% of female Lecturers co-authoring with Professors and 59% of males.

to the Economics and Econometrics Unit of Assessment would be penalised if it was co-authored by more than one person in that department's submission. A negative relationship between within department co-authorship and REF2014 submission is expected.

3. Estimation results

To reiterate, we are interested in the probability that an individual Lecturer is selected for the REF2014 process conditional on a range of observable characteristics expected to explain that probability. Interpreting the coefficient estimates from probit models is not straight forward, the more intuitive marginal effects at the means of the explanatory variables are instead reported in Table 2 along with differential effects for binary variables (the coefficient estimates are available from the authors upon request). Column 1 of Table 2 presents the unconditional gender gap in REF selection; women are 11.4% less likely to be included. Unsurprisingly, the measure of fit is very low with only some 8% of variation being explained. The additional explanatory variables we include in our model are number of publications, number of working papers, average quality of outputs, if co-authors are present in the same institution, early career status, department regional location, and 2008 department RAE grade. All of these variables are defined and discussed in section 2 above.

[Table 2 around here]

3.1 *Gender specific estimation*

Columns 2 and 3 of Table 2 present the estimated marginal effects for the model for females and males, respectively. In aggregate, the results from these gender specific estimates are statistically similar, which may be due to the relatively small samples included in the analysis, especially of females. There are, however, some noteworthy qualitative and quantitative findings. Reading down the columns it is observable that the number of publications is not relevant for either gender. In contrast, having an extra working paper increases the chance of a male being included by 15 percentage points (pp); substantially more so than for females. Higher average quality of outputs is equally important for both genders. For both males and females, having a co-author in the same department lowers the propensity to be submitted to REF2014. It is noteworthy that this relationship is actually qualitatively stronger for males than for females. All ECR categories are more likely to be submitted than the longer tenured Lecturers (the omitted not early career, NECR,

category). This is especially true for female LECR. We find no regional differences. Finally, the RAE grade of the department has a very strong and statistically significant relationship with selection probability for males but not for females. Working in a department with a RAE grade one unit higher than the average in the sample is associated with a substantial 60% greater chance of selection into REF2014 for males.

3.2 High and low ranked RAE departments

Results for separate analysis of HEIs graded below a 3 in the Economics and Econometrics UoA in the 2008 RAE and those graded 3 and above are presented in columns 4 and 5 of Table 2. There are very close to half the Lecturers in each of these groupings (154 in the 15 departments ranked below 3, and 146 in the 13 departments ranked 3 and above). In general, the high ranked research institutions place considerably more importance on the average quality of outputs and the number of working papers (indeed, the latter is insubstantial and insignificant in the low ranked departments). Furthermore, even within the low and high RAE groups, there is a substantial and significant difference in the probability of Lecturers being submitted to REF; *ceteris paribus*, high ranked RAE departments are considerably more likely to do so. We also find substantial, but statistically insignificant, gender differentials with women being some 11 pp less likely to be selected in the REF in the lower, and 9 pp less likely in the higher, RAE ranked departments.

To better understand the relationship between gender, RAE grade, and submission probability we next estimate a pooled model including an interaction term between being female and RAE 2008 grade (see column 1 of Table 3).¹⁶ This model allows for more flexibility in modelling gender and RAE grade and allows for less constraint from the small gender specific sample sizes; this pooled model is our benchmark model.

[Table 3 around here]

Considering the findings for the benchmark model in more detail (column 1 of Table 3),

¹⁶ We considered a fully gender interactive pooled model and found no statistically significant relationship between submission and terms interacting gender and other explanatory dependent variables, including the department RAE grade (results available upon request).

the number of journal publications is again not found to be related to submission, whilst the number of working papers is. An additional working paper is associated with 11 pp greater chance of inclusion (we will return to consider the link with working papers and submission more fully below). The average quality of outputs is found to be positively related to submission; a whole point increase in the average output quality implies the Lecturer is 14 pp more likely to be submitted. Whereas, having a co-author in the same institution reduces the probability of being included in the REF submission by some 28 pp, at the mean. Shorter tenure as a Lecturer is also substantially and significantly associated with a higher selection probability (compared to the omitted NECR, not early career researcher, category). These results are strongly associated with the REF2014 tapering rule for each of the ECR categories. Working in a department with higher RAE results in the 2008 research assessment exercise is strongly associated with a higher probability (42 pp) of a Lecturer being submitted in the REF in our benchmark model.

The gender differences reported in Table 3 are not statistically significant at standard confidence levels, nevertheless, the marginal effects related to gender and RAE grade are essentially unchanged across specifications of the model and implications may be of some interest. In a non-linear model, such as the probit model used here, the marginal effect of the RAE and gender interaction term is not a single value calculated at the mean and so interpretation is problematic. A more useful approach is to consider the predicted probabilities, predicted submission probabilities suggests a complex selection decision. For example, going from the department mean value of the RAE to the highest reported department RAE value is only associated with an increased probability of being submitted to the REF of 5 pp for females. In contrast, going from the mean value of the RAE to the highest value is associated with an increase in the probability of being submitted to the REF by 26 pp for males. Whilst based on limited statistical significance, these findings would suggest higher ranked RAE 2008 departments submit males with a higher probability, whereas females are submitted by the higher RAE ranked departments with similarly low probabilities as they are in the lower ranked departments. Finally, there are no quantitative nor qualitative regional differences in the probability of Lecturers being submitted to REF; the regional measures are included as controls in all of the models presented in Table 3.

The overall fit of our benchmark model is reasonable for cross-sectional models of this

type; the pseudo R-squared measure suggests that the model is explaining some 40% of the REF2014 submission probability for the Lecturers. Given the nature of the selection exercise and the explanatory variables we are including in the analysis, we might reasonably expect this goodness of fit measure to be higher. This may suggest we do not have the right measure of research outputs, that departments are using other selection criteria, and perhaps even that departments are engaging in gaming processes. We next consider some of these possibilities.

4. Further considerations.

4.1 *Excluding other assessable outputs*

The analysis presented so far allows for books, book chapters and conference proceedings to be included with working papers and journal publications as assessable outputs. It may be that this definition is too broad and the institutions only considered working papers and published journal articles when making their submission decisions (see column 2 of Table 3). Using a simpler metric is consistent with the findings of Clerides et al (2011) with the earlier RAE in the UK, see also Betocchi et al (2015) for Italy.¹⁷ Comparing columns 1 and 2 of Table 3, reveals no qualitative or quantitative differences in the results, suggesting that the simpler list of assessable outputs (working papers and journal publications) is a good approximation of the decision-making information used by institutions when submitting to the Economics and Econometrics UoA. The proportion of variation explained by the models using the simpler list of assessable outputs (column 2) or the fuller range (column 1) are also essentially the same (see the final panel of Table 3).

4.2 *On assumed quality, and subsequent publication, of working papers*

Individuals with working papers in the REF2014 submission required institutions to make selection decisions in the face of greater uncertainty (without the quality signal journal of publication provides). Our results may be influenced by the assumption that the quality of working papers is the same as the average quality of the individual's journal publications. The working papers may actually have been higher (or lower) quality than the existing journal publications for an individual. Institutions may have recognised this and judged the quality of the working papers themselves via

¹⁷ Various combinations of assessable outputs were also considered (results are available from the authors upon request), in none of these cases were the results found to be qualitatively or quantitatively different to those presented in Table 3.

peer review, making more informed submission decisions accordingly.

We address this possibility by changing the status for those working papers on the output census date (December 31st 2013) that were subsequently published in journals (by December 31st 2019).¹⁸ Interestingly, subsequent publication of these working papers is still relatively rare some six years on from the REF2014 census date. Of the 143 working papers in the sample at the REF cut-off date only 60 (42%) had subsequently been published; the average quality of these subsequent publications was 3.3 for both males and females, and the average *ex post* publication rate was 39% for males and 55% for females. This difference in *ex post* publication rate is not statistically significant in our sample, however it is consistent with the findings of Hengel (2018) that female economists take longer in the publication process than do males.

Results including this *ex post* publication information are provided in column 3 of Table 3, including only working papers and journal articles as outputs and coding the remaining unpublished working papers with a zero quality measure. Comparing the results in columns 2 and 3 reveals little change. Unsurprisingly, there is an increase in the importance of the number of publications, and a small decrease in the importance of the number of working papers. The relationship between the average quality of outputs and selection lowers suggesting that working paper quality may have been judged higher than other publications in the actual submission process. The gender gap also increases, from 9 to 12%. However, none of the changes in the results are statistically significant. Nevertheless, the findings are consistent with institutions making a reasonable attempt to assess working paper quality when making their REF2014 inclusion decisions (Tymms and Higgins, 2018), more so for men than for women.

As discussed in section 2 above, the highest quality performance rank for outputs in REF2014 was a 4 which was awarded for ‘quality that is world-leading in terms of originality,

¹⁸ Three searches for subsequent publication of working papers were carried out: in January 2016; April 2018; and December 2019 (final check on December 31st, 2019). In each iteration of this search process, the working paper was considered published if it had the same title, or, if it had a title similar enough to the original to be picked up through research platforms including Google Scholar, ResearchGate and IDEAS. Checks of staff, institutional and/or working paper webpages were also conducted to further track working papers into published output. It is possible, however, that a working paper may not be recognised as subsequently published if a sufficient title change was made, and there was no linkage information available on the staff members or working paper webpages.

significance and rigour'¹⁹. However, it is may also be possible that for HEIs having a publication in one of the top 5 ranked economics journals²⁰ (a so called '5*' paper) is weighted more highly for REF inclusion than having one of the other level 4 publications. There are 18 of these 5* publications in our data set from 17 Lecturers; 4 female Lecturers and 13 male. Most (15) were published by the REF cut-off with the remaining 3 published shortly after (and so would have been in late revision stage).

It might also be argued that departments 'hired for the REF' meaning that they restricted hiring just before the REF to those who already had a high quality publication 'in the bag'. A little over half of the eighteen 5* papers submitted by Lecturers in REF2014 were published by EECRs (10 out of 18, or 56%), 2 are published by MECR (11%), 3 by LECR (17%), and 3 by NECR (17%). These proportions are roughly in line with the distribution of early career status across the Lecturers: 53% of the Lecturers are EECR, 10% are MECR, 9% LECR, and 28% NECR (see Table A3). Analogously, 7% of the EECR Lecturers had a 5* publication, 7% of the MECRs, 8% of the LECR, and 4% of the NECR. Whilst 5* publications are less common amongst NECR Lecturers, our results do not suggest that they are more common amongst the EECRs.

Reclassifying outputs that were published in a 5* journal as a 5 rather than a 4 in our output quality measure has negligible impact on the results (see column 4 of Table 3). Comparing columns 3 and 4, we can see no qualitative or quantitative differences in the results, suggesting that for Lecturers having a 4 as opposed to a 5* publication is not relevant for REF submission.

4.3 *Research inactive.*

We originally excluded 5 Lecturers from our sample of interest because they were not research active, were believed to be on teaching only contracts, and were thus not eligible for REF submission. It is possible that others were also on a teaching only contract despite having some early research activity. There are 32 Lecturers (23 male) in our sample who were not submitted to

¹⁹ <https://www.ref.ac.uk/2014/panels/assessmentcriteriaandleveldefinitions/>

²⁰ The American Economic Review; Econometrica; Journal of Political Economy; Review of Economic Studies; and the Quarterly Journal of Economics. We also considered the Annals of Statistics but found none for this journal in our data set.

REF2014 and do not present evidence²¹ of any additional research output (either a publication or an additional working paper, conference paper, book or book chapter) between the REF cut-off date and December 31st 2019, four of these Lecturers (all male) appear to have left the profession. Reclassifying these Lecturers as research inactive and removing them from the analysis, leaves 268 in the sample of interest. In other words, we are treating these Lecturers as if they were on teaching only contracts on the REF cut-off date. In reality, it may be the case that many of these 32 simply did not further develop their academic research profiles over the subsequent 6 years.²²

Results for the restricted sample of 268 Lecturers, including their *ex post* publication information and allowing for 5* publications in the average quality measure, are provided in column 5 of Table 3. By comparing columns 4 and 5 of Table 3, we can see that the reclassification of those without additional research activity post-REF has little impact on the results. There is some evidence that they are more likely to be amongst those NECR with less submission likelihood.²³ There is also some growth in the relationship between 2008 RAE grade and submission, indicating that higher ranked RAE departments may have been more inclined to have Lecturers on teaching-only contracts at the REF cut-off date, although this change in the results is not very strong.

5. Discussion and conclusions

We explore the relationship between research outputs and REF2014 inclusion for economics Lecturers in the UK, using a rich new data source collected by the authors and paying particular attention to the quality and quantity of these outputs. We find that only one-in-two (54%) of these Lecturers were submitted to REF2014; 57% of men and 46% of women.

We find that HEIs using a simpler list of assessable outputs, focussing on working papers and journal publications papers, is a good approximation of the decision-making information used by institutions when submitting to the Economics and Econometrics UoA. Evidence from subsequent publications of working papers in journals, further implies that institutions were using

²¹ As verified from searching their institutional and personal webpages, and through research platforms including Google Scholar, ResearchGate and IDEAS.

²² Of the 32 lecturers which we consider to be research inactive since the last REF, 20 (63%) of them had no research outputs in our original dataset, which represents 65% of all those with no outputs (see Table A4).

²³ 73% were NECR, 24% were either EECR or MECR.

a peer review process and making quality predictions when submitting working papers to the REF2014 process. Controlling for ECR status according to REF2014 rules is also found to be important. Institutions are considerably more likely to submit Lecturers with shorter tenure in their REF selection process. This appears to be related to the steep tapering of output requirements with years of experience for the different REF categories of ECRs. The potential negative impact on longer-tenured Lecturers being submitted may be more extreme in economics, where the journal publication process is comparatively lengthy, than in other disciplines.

A strong negative association between submission and co-authorship within the institution is also established in the results. Women are found to co-author with colleagues in the same department more commonly than males, but the penalty for males from co-authoring is higher. Discouraging co-authorship may lead to diminished within-department collegiality with detrimental consequences for research quality and the development of junior academics. Furthermore, whilst there are no regional differences in the results, there are also no Welsh or Northern Irish HEIs choosing to submit to the Economics and Econometrics Unit of Assessment, raising serious issues as to the lack of a National context for the REF exercise in this discipline.

Our study has implications for the upcoming REF2021 exercise and we can postulate some early implications. A co-authorship rule has been announced for REF2021 allowing for up to two co-authors per output in a submission. By lowering uncertainty over the grading of co-authorship, we would expect this ruling to lessen the negative association between co-authorship and REF selection. The tapering requirement for ECRs has also been removed which should avoid the distinction in selection probabilities by type of ECR and potential preference for shorter tenure Lecturers. A primary recommendation of the Stern (2016) report is that all research active staff be returned in the next REF in 2021 (Stern, 2016; page 19); this was clearly not the case for economics Lecturers in REF2014. Under the REF2011 guidelines (REF, 2019; page 28) institutions could choose to reclassify individuals from research active to teaching focussed contracts before the REF2021 census date (July the 31st, 2020) to avoid inclusion without penalty. Staff who left the institution (retired, resigned or made redundant²⁴) before the census date can also have their output

²⁴ Completing a fixed-term employment contract of more than two years is often counted as a redundancy by HR departments in the UK. We had 16 Lecturers on fixed term contracts in our sample, only four were chosen for the REF

submitted without penalty to the institution. Subject to Human Relations (HR) arrangements within institutions, such contract changes are sometimes complex but manageable. Being moved into a research inactive employment role would imply a considerably higher teaching load and lower status as a research economist; arguably imposing substantially higher costs than those incurred by individuals not included in the REF2014 process

Lecturers are typically the least experienced of those on standard academic contracts and they may be considered to have future career paths that are the most vulnerable to inclusion or exclusion from research excellence activities. After controlling for research output quantity and quality, co-authorship, early career status, department research ranking and location, and gender; more than half of the variance in selection probability is still left unexplained. This indicates considerable idiosyncrasies in the management of submissions by HEIs in the REF process and uncertainty for individual academics across the discipline in the research assessment process.

Declaration of interests. The authors have no conflicts of interests to declare.

Acknowledgments. We would like to thank Victoria Bateman, Steve Machin, Peter Neary, Denise Osborn, Peter Smith; participants of the 2018 RES Conference for helpful comments and advice; and to Craig Fox for excellent research assistance.

REFERENCES

- Bayer, A., Rouse, C.E. 2016. 'Diversity in the economics profession: A new attack on an old problem.' *The Journal of Economic Perspectives*, 30(4), 221-242.
- Berry, D. 2012. REF Codes of Practice for the selection of staff: A report on good practice by the Equality and Diversity Advisory Panel (EDAP).
<http://www.ref.ac.uk/pubs/refcodesofpracticegoodpracticereport/>
- Bertocchi, G., Gambardella, A., Jappelli, T., Nappi, C.A., Peracchi, F. 2015. Bibliometric evaluation vs. informed peer review: Evidence from Italy. *Research Policy*, 44(2), 451-466.
- Boring, A. 2017. 'Gender biases in student evaluations of teaching.' *Journal of Public Economics*, 145, 27-41.
- Brown, A., Yang, F. 2015. 'Does society underestimate women? Evidence from the performance of female jockeys in horse racing.' *Journal of Economic Behavior & Organization*, 111, 106-118.

all of who were male.

- Clerides, S., Pashardes, P., Polycarpou, A. 2011. Peer review vs metric-based assessment: Testing for bias in the RAE ratings of UK Economics Departments. *Economica*, 78(311), 565-683.
- De Fraja, G., Facchini, G., Gathergood, J. (2020) 'Academic salaries and public evaluation of university research: Evidence from the UK Research Excellence Framework.' *Economic Policy*, forthcoming.
- Geuna, A., Piolatto, M. 2016. 'Research assessment in the UK and Italy: Costly and difficult, but probably worth it (at least for a while).' *Research Policy*, 45, 260-271.
- Gould R. 2002. 'The origins of status hierarchies: A formal theory and empirical test.' *American Journal of Sociology*, 107, 1143-1178.
- Greenhalgh, T., Fahy N. 2015. 'Research impact in the community-based health sciences: An analysis of 162 case studies from the 2014 UK Research Excellence Framework.' *BMC Medicine*, 13(1), 232.
- Hamermesh, D.S. 2015. 'Age, cohort and coauthorship.' IZA DP 8828.
- HEFCE. 2015. The Selection of Staff for Inclusion in the REF 2014. <http://www.hefce.ac.uk/pubs/year/2015/201517/>
- HEFCE. 2017. Annual funding allocations 2017-18. <http://www.hefce.ac.uk/funding/annalocns/1718/>
- HESA. 2020. Higher Education Staff Statistics: UK, 2018/19. <https://www.hesa.ac.uk/news/23-01-2020/sb256-higher-education-staff-statistics>
- Hengel, E. 2018. 'Publishing while female. Are women held to higher standards? Evidence from peer review.' Paper presented at the AEA Conference Philadelphia 5/1/2018. <https://www.aeaweb.org/conference/2018/preliminary.2095>.
- Hicks, D. 2012. 'Performance-based university research funding systems.' *Research Policy*, 41(2), 251-261.
- Hole, A.R. 2017. 'Ranking economics journals using data from a national research evaluation exercise.' *Oxford Bulletin of Economics and Statistics*, 79(5), 621-636.
- Hudson, J. 2013. 'Ranking Journals.' *The Economic Journal*, 123(570), F202-F222.
- Levin, S., Stephan, P. 1991. 'Research productivity over the life cycle: Evidence for academic scientists.' *American Economic Review*, 81(1), 114-132.
- Long, J., Fox, M. 1995. 'Scientific careers: Universalism and particularism.' *Annual Review of Sociology*, 21, 45-71.
- Lundberg, S., Stearns, J. 2019. 'Women in Economics: Stalled progress.' *Journal of Economic Perspectives*, 33(1), 3-22. .
- Maddala, G.S. 1992. *Introduction to Econometrics*. Second Edition. Prentice Hall: New Jersey.
- Martin, B.R. 2011. 'The Research Excellence Framework and the 'impact agenda': Are we creating a Frankenstein monster?' *Research Evaluation*, 20(3), 247-254.
- Mengel, F., Sauermann, J., Zolitz, U. 2018. 'Gender bias in teaching evaluations.' *Journal of the European Economic Association*, DOI:10.1093/jeea/jvx057.
- Merton, R. 1968. 'The Matthew effect in science.' *Science*, 159, 55-63.
- Mitka, M., Mumford, K., Sechel, C. 2015. The 10th Royal Economic Society Women's Committee Survey: The Gender Balance of Academic Economics in the UK 2014. <http://www.res.org.uk/view/publicationsWomensComm.html>
- Mumford, K., Sechel, C. 2019. 'Job satisfaction amongst academic economists in the UK.' *Economic Letters* 182: 55-58.
- Mumford, K., Sechel, C. 2020. 'Pay and job rank amongst academic economists in the UK: Is gender relevant?' *British Journal of Industrial Relations* 58(1): 82-113.

- Murphy, T., Sage, D. 2014. 'Perceptions of the UK's Research Excellence Framework 2014: A media analysis.' *Journal of Higher Education Policy and Management*, 36(6), 603-615.
- OECD. 2018. *Science, Technology and Innovation Outlook 2018. Adapting to Technological and Societal*. OECD Publishing, Paris. https://doi.org/10.1787/sti_in_outlook-2018-en
- Ovseiko, P.V., Oancea, A., Buchan, A.M. 2012. 'Assessing research impact in academic clinical medicine: A study using Research Excellence Framework pilot impact indicators.' *BMC Health Services Research*, 12(1), 478.
- Parker, J., van Teijlingen, E. 2012. 'The Research Excellence Framework (REF): Assessing the impact of social work research on society.' *Practice*, 24 (1), 41-52.
- RAE. 2005. rae2008 Research Assessment Exercise. Guidance on Submission. June 2005. Ref RAE 03/2005. <https://www.rae.ac.uk/pubs/2005/03/#exec>
- RAE. 2008. RAE 'Confirms the UK's dominant position in international research'. <http://www.rae.ac.uk/news/2008/results.asp>
- Rees, T. 2011. 'The gendered construction of scientific excellence.' *Interdisciplinary Science Reviews*, 36(2), 133-145.
- REF. 2012. Assessment framework and guidance on submissions. <http://www.ref.ac.uk/media/ref/content/pub/assessmentframeworkandguidanceonsubmissions/GOS%20including%20addendum.pdf>
- REF. 2019. REF 2019/01 Guidance on submissions. January 2019. <https://www.ref.ac.uk/publications/guidance-on-submissions-201901/>
- Regibeau, P., Rockett, K. 2016. 'Research assessment and recognized excellence: Simple bibliometrics for more efficient academic research evaluations.' *Economic Policy*, 31(88), 611-652.
- Reskin, B., Hargens, L. 1979. 'Scientific advancement of male and female chemists.' In *Discrimination in Organizations*, eds. R. Alverez and K. Lutterman, pages 100-122. (San Francisco: Jossey-Bass).
- Sarsons, H. 2017. 'Recognition for group work: Gender differences in academia,' *American Economic Review: Papers & Proceedings*, 107(5), 141-145.
- Smith, S., Ward, V., House, A. 2011. "'Impact' in the proposals for the UK's Research Excellence Framework: Shifting the boundaries of academic autonomy.' *Research Policy*, 40(10), 1369-1379.
- Stern, N. 2016. *Building on Success and Learning from Experience: An Independent Review of the Research Excellence Framework*. Department for Business, Energy & Industrial Strategy.
- Trevorrow, P., Volmer, D.A. 2012. 'Dispelling the myths surrounding the Research Excellence Framework.' *Rapid Communications in Mass Spectrometry* 26(4), 399-402.
- Tymms, P., Higgins S. 2018. 'Judging research papers for research excellence.' *Studies in Higher Education*, 43(9), 1548-1560, DOI: 10.1080/03075079.2016.1266609
- Vuong, Q.H. 1989. 'Likelihood ratio tests for model selection and non-nested hypothesis.' *Econometrica* 57(2), 307-333.
- Wu, A. 2017. 'Gender stereotyping in academia: Evidence from Economics job market rumours forum', AEA Conference presentation Philadelphia, 2018. Retrieved from <https://www.aeaweb.org/conference/2018/preliminary/2095> on April 20, 2018.

Table 1. Summary statistics

	Females		Males		Both genders	
	mean	st. dev.	mean	st. dev.	mean	st. dev.
female	–	–	–	–	0.28	0.45
REF submission status	0.46	0.50	0.57	0.50	0.54	0.50
2008 RAE grade	2.97	0.25	3.01	0.27	3.00	0.26
number of publications	1.93	1.54	1.55	1.37	1.65	1.43
number of working papers	0.31	0.68	0.44	0.76	0.40	0.74
average quality of publications	2.71	1.22	2.67	1.18	2.68	1.19
average quality of outputs	1.80	1.62	1.83	1.58	1.82	1.59
England (excluding London)	0.67	0.47	0.71	0.45	0.70	0.46
London	0.19	0.40	0.17	0.37	0.17	0.38
Scotland	0.13	0.34	0.12	0.33	0.13	0.33
NECR	0.28	0.45	0.28	0.45	0.28	0.45
LECR	0.13	0.34	0.07	0.26	0.09	0.29
MECR	0.11	0.31	0.10	0.30	0.10	0.30
EECR	0.48	0.50	0.55	0.50	0.53	0.50
co-authors in same institution	0.19	0.40	0.09	0.28	0.12	0.32
Observations	83		217		300	

Number pairs in bold are significantly different at the 90% confidence level; in italics and bold they are significantly different at the 95% confidence level.

Table 2. Probit regression (marginal effects at means)

Dependent variable: REF = 1 if individual entered in REF2014, 0 if not entered

	(1) full sample	(2) females	(3) males	(4) RAE<3	(5) RAE>=3
number of publications		-0.029 (0.059)	0.035 (0.046)	<i>0.094*</i> <i>(0.054)</i>	<i>-0.059</i> <i>(0.055)</i>
number of working papers		0.119 (0.129)	0.148** (0.060)	<i>0.036</i> <i>(0.074)</i>	<i>0.275***</i> <i>(0.095)</i>
average quality of outputs		0.177*** (0.056)	0.179*** (0.037)	0.134*** (0.042)	0.221*** (0.048)
co-authors in same institution		-0.240* (0.144)	-0.325** (0.153)	-0.377*** (0.114)	-0.187 (0.187)
Researcher status (baseline is NECR)					
<i>LECR</i>		<i>0.597***</i> <i>(0.129)</i>	<i>0.302***</i> <i>(0.111)</i>	0.559*** (0.086)	0.285** (0.145)
<i>MECR</i>		0.329 (0.216)	0.445*** (0.061)	0.487*** (0.115)	0.425*** (0.075)
<i>EECR</i>		0.713*** (0.119)	0.713*** (0.075)	0.772*** (0.077)	0.745*** (0.100)
Location (baseline is England excluding London)					
<i>London university</i>		-0.006 (0.175)	0.030 (0.129)	-0.165 (0.153)	-0.036 (0.165)
<i>Scottish university</i>		0.076 (0.216)	0.127 (0.119)	0.179 (0.137)	0.092 (0.212)
2008 RAE grade		0.208 (0.292)	0.595*** (0.178)	<i>0.174</i> <i>(0.583)</i>	<i>1.461***</i> <i>(0.384)</i>
female	-0.114* (0.064)			-0.112 (0.117)	-0.085 (0.127)
Observations	300	83	217	154	146
Pseudo R-squared	0.008	0.353	0.402	0.399	0.489
AIC	414.853	96.014	199.201	152.407	125.494
BIC	422.261	122.621	236.380	188.851	161.297

Standard errors in parentheses. * p<0.10, ** p<0.05 *** p<0.01. Coefficient pairs in bold are significantly different at the 90% confidence level; in italics and bold are significantly different at the 95% confidence level.

Table 3. Probit regression (marginal effects at means); varying output measures.

Dependent variable: REF = 1 if individual entered in REF2014, 0 if not entered

	(1) Ave journal rank assumed for wp Benchmark model	(2) Only wp and journal considered	(3) Later publication info used for wp	(4) top5 publications w/ post-ref info for wp	(5) top5 publications w/ post-ref infor for wp excluding research inactive staff
number of publications	0.023 (0.035)	0.042 (0.036)	0.061* (0.035)	0.058* (0.035)	0.055 (0.036)
number of working papers	0.143*** (0.054)	0.151*** (0.054)	0.125** (0.052)	0.123** (0.052)	0.102* (0.053)
average quality of outputs	0.174*** (0.029)	0.167*** (0.030)	0.152*** (0.028)	0.155*** (0.028)	0.141*** (0.027)
co-authors in same institution	-0.283*** (0.107)	-0.289*** (0.106)	-0.325*** (0.098)	-0.325*** (0.098)	-0.357*** (0.109)
Researcher status (baseline is NECR)					
<i>LECR</i>	0.398*** (0.080)	0.403*** (0.080)	0.397*** (0.080)	0.395*** (0.080)	0.316*** (0.074)
<i>MECR</i>	0.433*** (0.069)	0.438*** (0.069)	0.422*** (0.071)	0.421*** (0.071)	0.349*** (0.062)
<i>EECR</i>	0.712*** (0.062)	0.724*** (0.061)	0.676*** (0.066)	0.674*** (0.066)	0.633*** (0.080)
2008 RAE grade	0.424*** (0.148)	0.434*** (0.148)	0.388*** (0.146)	0.363** (0.147)	0.405*** (0.148)
female	-0.092 (0.081)	-0.093 (0.080)	-0.121 (0.078)	-0.120 (0.079)	-0.119 (0.081)
female*RAE	-0.265 (0.207)	-0.259 (0.207)	-0.270 (0.208)	-0.274 (0.209)	-0.298 (0.227)
Region controls	yes	yes	yes	yes	yes
Observations	300	300	300	300	268
Pseudo R-squared	0.378	0.380	0.368	0.373	0.330
AIC	283.352	282.460	287.546	285.391	266.885
BIC	331.501	330.609	335.695	333.540	313.568

Standard errors in parentheses. * p<0.10, ** p<0.05 *** p<0.01. All models include regional controls. Female*RAE is the interaction effect computed using `-inteff-` command in Stata.

ONLINE APPENDIX – intended for online provision only.

Table A1. List of Universities with CHUDE economics departments submitting to Economics and Econometrics Unit of Assessment

Brunel University; City University London; London School of Economics and Political Science; Queen Mary University of London; Royal Holloway University of London; University College London; University of Aberdeen; University of Birmingham; University of Bristol; University of Cambridge; University of East Anglia; University of Edinburgh; University of Essex; University of Exeter; University of Glasgow; University of Leicester; University of London Birkbeck; University of Manchester; University of Nottingham; University of Oxford; University of Sheffield; University of Southampton; University of St Andrews; University of Surrey; University of Sussex; University of Warwick; University of York

Table A2. Researcher status and required outputs

Date academic began independent research (1)	Required output for REF14 (2)	ECR status (3)	Female (4)	Male (5)	%Female (6)	Total (7)
On or before July 31, 2009	4	Not early career researcher (NECR)	23	61	27%	84
Between August 1, 2009 and July 31, 2010	3	Late early career researcher (LECR)	11	16	41%	27
Between August 1, 2010 and July 31, 2011	2	Middle early career researcher (MECR)	9	21	30%	30
Between August 1, 2011 and October 31, 2013	1	Early early career researcher (EECR)	40	119	25%	159
After October 31, 2013	N/A	Not eligible for REF				
Total			83	217	28%	300

Table A3. ECR status, by REF submission status

	not in REF	in REF	% in REF	Total
<i>All lecturers</i>				
NECR: 4 outputs required	70	14	17%	84
LECR: 3 outputs required	16	11	41%	27
MECR: 2 outputs required	14	16	53%	30
EECR: 1 output required	38	121	76%	159
Total	138	162	54%	300
<i>Female lecturers</i>				
NECR: 4 outputs required	21	2	9%	23
LECR: 3 outputs required	6	5	45%	11
MECR: 2 outputs required	6	3	33%	9
EECR: 1 output required	12	28	70%	40
Total	45	38	46%	83
<i>Male lecturers</i>				
NECR: 4 outputs required	49	12	20%	61
LECR: 3 outputs required	10	6	38%	16
MECR: 2 outputs required	8	13	62%	21
EECR: 1 output required	26	93	78%	119
Total	93	124	57%	217

Note: output requirements may differ due factors such as mitigating circumstances or leaves of absence.

Table A4. ECR status, by number of outputs (including all submitted and not submitted to REF 2014)

	number of outputs (% of row category total in parenthesis)										
	0	1	2	3	4	ave					
<i>All lecturers</i>											
NECR: 4 outputs required	9	11%	15	18%	8	10%	8	10%	44	52%	2.75
LECR: 3 outputs required	3	11%	3	11%	1	4%	10	37%	10	37%	2.78
MECR: 2 outputs required	3	10%	4	13%	7	23%	10	33%	6	20%	2.40
EECR: 1 output required	16	10%	85	53%	27	17%	12	8%	19	12%	1.58
Total	31	10%	107	36%	43	14%	40	13%	79	26%	2.10
<i>Female lecturers</i>											
NECR: 4 outputs required	3	13%	4	17%	2	9%	1	4%	13	57%	2.74
LECR: 3 outputs required	1	9%	0	0	1	9%	3	27%	6	55%	3.18
MECR: 2 outputs required	1	11%	2	22%	1	11%	4	44%	1	11%	2.22
EECR: 1 output required	5	13%	18	45%	8	20%	3	8%	6	15%	1.68
Total	10	12%	24	29%	12	14%	11	13%	26	31%	2.23
<i>Male lecturers</i>											
NECR: 4 outputs required	6	10%	11	18%	6	10%	7	11%	31	51%	2.75
LECR: 3 outputs required	2	13%	3	19%	0	0%	7	44%	4	25%	2.50
MECR: 2 outputs required	2	10%	2	10%	6	29%	6	29%	5	24%	2.48
EECR: 1 output required	11	9%	67	56%	19	16%	9	8%	13	11%	1.55
Total	21	10%	83	38%	31	14%	29	13%	53	24%	2.05

Note: output requirements may differ due factors such as mitigating circumstances or leaves of absence.

Table A5. Number of outputs, by REF submission status

number of outputs	not in REF	in REF	% in REF	Total
<i>All lecturers</i>				
0	27	0	0%	27
1	29	89	75%	118
2	18	23	56%	41
3	20	19	49%	39
4	44	31	41%	75
Total	138	162	54%	300
<i>Female lecturers</i>				
0	7	0	0%	7
1	7	22	76%	29
2	7	3	30%	10
3	7	4	36%	11
4	17	9	35%	26
Total	45	38	46%	83
<i>Male lecturers</i>				
0	20	0	0%	20
1	22	67	75%	89
2	11	20	65%	31
3	13	15	54%	28
4	27	22	45%	49
Total	93	124	57%	217