The Efficiency of Tenure Contracts in Academic Employment

Bruce Cater, Byron Lew and Marcus Pivato*

Abstract

Academic research is a public good whose production is supported by the tuition-paying students that a faculty's research accomplishments attract. A professor's spot contribution to the university's revenues thus depends not on her spot research production, but rather on her entire cumulative research record. We show that, under a broad range of education market conditions, a profit-maximizing university will apply a high minimum retention standard to the research production of a junior professor who has no record of past research, but a zero minimum standard to the spot production of a more senior professor whose background includes accomplishments sufficient to have cleared the high probationary hurdle. But if and when education market conditions change, tenure-based contracts may cease to be optimal.

I Introduction

For decades, the most widely-used employment arrangement between a university and a professor was the tenure-track contract. Under that contract, a professor who fails to meet some positive standard of research production during a finite probationary period is dismissed at that period's end. Yet, a professor who meets that initial standard is granted *tenure* and retained *regardless* of her research output thereafter.[1]

*Cater and Lew: Department of Economics, Trent University, Ontario, Canada; Pivato: Laboratoire THEMA et UFR Économie et Gestion, Université de Cergy-Pontoise, France. We are grateful to Barry Smith, Myrna Wooders, and anonymous referee for many helpful comments and suggestions, and to Duc Hien Nguyen and Stephen Swanson for excellent research assistance. Financial support from Trent University, NSERC grant #262620-2008, and Labex MME-DII (ANR11-LBX-0023-01) is gratefully acknowledged.

[1] Siow (1998) notes that, in the 1989 *Survey Among College and University Faculty* sponsored by the Carnegie Foundation Survey, 4.7 percent and 36.4 percent of tenured faculty in doctoral-granting and non-doctoral-granting schools, respectively, reported no publications in the previous two years and no current research. Yet, in their reviews of U.S. case law, legal scholars including Hendrickson (1988) and Morris (1992) cite *no* cases in which a tenured professor was dismissed primarily for low research productivity. The minimum research production standard for a tenured professor is, therefore, effectively 'zero'.

This contractual choice raises several questions:

1. The dismissal of those who are initially unproductive makes it clear that research is somehow important to a university. Why, then, would it not insist that a professor be productive in research at *every* stage of her career?

2. The research production of academics declines with age (Diamond, 1986; Levin and Stephan, 1991; Kenny and Studley, 1995; and Oster and Hamermesh, 1998). Does this pattern reflect some disincentive effect, and therefore a major drawback of tenure?

3. If the sort of leniency associated with tenure is somehow efficient, why do universities stand alone in extending that leniency?[2]

4. If universities derive some unique benefit from granting tenure, how is their recent and ongoing shift away from the use of tenure-eligible faculty to be explained?[3]

A number of efficiency-based explanations of tenure have been proposed.[4] Freeman (1977) suggests that risk averse professors are granted the security of tenure to compensate for the risk inherent in their research. Yet, non-academic employers manage to compensate workers who are risk averse and whose productivity is uncertain without ever having to forgive very poor performance of a core duty. So, while offering a plausible solution to Question 1, this theory cannot resolve Question 3, let alone Questions 2 and 4.

Carmichael (1988) suggests that a university faces a unique problem. Because the state of academic knowledge is vast and expanding, it is the incumbent occupants of its scarce faculty slots who can best judge the research potential of candidates. To maximize its research production, the university provides those incumbents with the security of tenure to ensure that they are willing to identify and hire candidates superior to themselves. This is an elegant solution to Questions 1 and 3. In assuming that a professor's research production is governed only by ability and not by effort, however, Carmichael does not seek to explore Question 2. Indeed, the analysis abstracts from that Question's very premise by assuming that expected research output is *constant*, rather than diminishing, over a professor's life cycle. This overstates the relative contribution of older, tenured professors

[2]Partnerships in legal, medical, and consulting practices also involve quasi-permanent appointments granted to those who succeed during a probationary period. But those arrangements are not characterized by the post-probationary leniency of tenure – a 'partner' who ceases to perform one of the tasks he was initially hired to perform *will* be terminated. Similarly, K-12 teachers are granted "tenure" following a probationary period. But teacher contracts explicitly provide for the dismissal of "incompetent" teachers, and while teacher unions understandably make the process of termination difficult and costly for a school board, terminations do occur. Indeed, the National Council of Teacher Quality actually grades U.S. states on the effectiveness of their policies regarding the dismissal of ineffective teachers.

[3]In 1975, 56.8 percent of U.S. faculty were tenured or on a tenure-track, while 13 percent were full-time non-tenure-track and 30.2 percent were part-time. By 2007, only 31.2 percent of faculty were tenured or tenure-eligible, while the full-time non-tenure-track and part-time contingent groups had increased to 18.5 and 50.3 percent, respectively (AAUP 2008-2009 Report on the Economic Status of the Profession).

[4]McKenzie (1996) and McPherson and Shapiro (1999) attempt to explain academic tenure on internal political grounds. Kahn and Huberman (1988) and Waldman (1990) offer economic explanations of the use of 'up-or-out' contracts, but do not address the issue of post-probationary minimum production standards.

and biases the analysis towards a finding that tenure is optimal for a research-maximizing university. Moreover, if tenure is necessary to solve an ongoing problem so critical to a university's mission, how is Question 4 to be answered?

On the assumption that research productivity falls with age, Siow (1998) argues that it becomes socially efficient for an older professor to spend less time on research and more time on teaching. Tenure solves an effort allocation problem by inducing older professors to do less research. This argument is a surprising and intriguingly plausible solution to Questions 2 and 3, but it cannot fully resolve Question 1. If universities wished to *eliminate* research production among older faculty, why are those faculty provided with a range of research incentives and support, including salary increases, teaching reductions, internal research grants, sabbaticals, and laboratory facilities? Of course, the provision of incentives and support may reflect universities' preference for older faculty to continue to produce at least some research. But how can that preference be reconciled with tenure's tolerance for no research production at all? Moreover, in arguing that tenure is needed to optimally allocate a professor's effort, this theory offers no answer to Question 4.

Despite their considerable insights, these prominent theories of academic tenure are thus clearly unable to fully resolve the contractual puzzle. The purpose of this paper is to offer a more complete and compelling explanation.

To that end, consider first a number of related questions regarding the nature of academic work. It is well known that the social value of certain kinds of ideas cannot be fully or even partially appropriated by their developers, and that governments provide subsidies to encourage this kind of work (Nelson (1959) and Arrow (1962)). But why are those subsidies directed to universities as opposed to private firms?

Aghion et al (2008) suggest that where research is characterized by a low (expected) appropriable value, it is important to economize on the wage and monitoring costs of researchers. It is thus efficient to grant those researchers creative freedom, for they value creative control and would require a wage premium to give it up. Because professors happen to be granted the freedom of tenure, a university is the place where subsidy-dependent research should be done. But what would preclude a firm from also realizing cost savings through granting creative freedom to *its* researchers? Is there something unique about a university — something that may also explain its unique contractual choice?

Suppose a government cares about the cost of developing of socially valuable but non-appropriable ideas. Other things equal, it would be indifferent between directing its subsidies to a firm and to a university. But other things are not equal. If the work were to be done in a firm, where a researcher engages in research alone, no revenues would result, and the full cost of the research would be borne by the subsidizing agency. But when done in a university, where professors also teach, revenues *do* result from research.

The reason is straightforward. Students value faculty knowledge, while universities differ in terms of the knowledge their faculties possess. Because knowledge is difficult to observe directly, faculty research may serve as a reliable proxy, either because faculty accumulate knowledge through research or because faculty who are more knowledgeable find it less costly to do research. Because observable research successes then serve to signal knowledge and attract students, research costs are at least partially defrayed by the resulting tuitions, reducing the funding required and giving us a simple explanation of why a university is the preferred target for government research subsidies.[5]

Consistent with that notion, the primary contribution of this paper is to show that, while the contractual choices of universities may well arise from several mechanisms working in confluence, a single mechanism — the way in which a professor's research output can be translated into tuition revenues — can, even in isolation, solve each of Questions 1 through 3, and offer a number of plausible and consistent answers to Question 4.

In Section II, we lay out a very simple model of an infinitely-lived university that seeks to maximize the profits earned from a faculty slot. Professors are hired from overlapping generations, each with a working lifetime of three periods. In each period, a professor's level of research effort, chosen to maximize her expected utility, stochastically determines her research output. For simplicity, professors are identical in terms of their research ability and effort disutility, and the tuition revenues resulting from any particular research accomplishment are assumed to not decay over time.

Section III describes the general structure of an employment contract, including the research effort incentives and expected profits that result from the university's choice of its minimum research output retention standards and its compensation structure.

In Sections IV and V, we derive our principal results. First, it is optimal for a university to induce a research effort profile that declines over a professor's working lifetime, for the opportunities to translate her research successes into subsequent instructional revenues decline as she approaches the end of her career. Second, a tenure-track contract is optimal. In our simple model, where ability, effort disutility, and the resulting effort profiles are homogenous, and where differences in future research draws are purely stochastic, the choice of each period's minimum research standard involves a simple tradeoff: Raising a current bar results in the retention of only those who have been *recently successful*, but it comes at the cost of foregoing the opportunity to continue to profit from the *prior successes* of those who have been *recently unsuccessful*. The resolution of this tradeoff then depends crucially on the stage of career at which a standard is being applied. At the end of the contract's first period, when a professor has had no earlier success, a high standard is best. But at the end of the contract's second period, when all those retained to that point have a record sufficient to have met the high first hurdle, the optimal minimum standard is zero.

Section VI then considers several extensions of the basic model. The first of these is motivated by two questions. First, would initial success warrant the tolerance of tenure if the revenue impact of a research accomplishment were to decay over time? Second, in practice, a tenure decision is made not at the one-third point of a professor's career (as in our principal results), but much earlier, typically after six or seven years of a career that spans, say, thirty-five years or more. Allowing for more frequent retention decisions, is the solution a tenure-track contract characterized by a tenure decision occurring at a point that is consistent with the timing we see in practice? Numerical results for a six-period

[5] Bok (1986), James (1990) and Hearn (1992) first suggested that research attracts students, while Siow (1997) has found that schools with more successful researchers have larger shares of out-of-state and foreign students. Students need not observe and process research *directly* for them to be attracted to schools that have stronger researchers — choice may simply be based on university rankings (e.g., US News & World Report and Maclean's University Rankings) that are themselves based on measures of research quality and reputation built on research.

generational length model indicate that, for all but very low research discount factors, the optimal balance in the tradeoff described in Section IV is again a tenure-track contract, with the tenure decision occurring after one period.

Our second set of extensions introduce individual heterogeneity — a university, after one period of observation, resolves initial uncertainty regarding a professor's research ability. Because professors then differ in terms of their past, current, *and* expected future research output, the tradeoff the university faces becomes slightly more involved — in addition to the considerations identified above, a higher minimum standard makes it more likely that a professor whose ability is known will be dismissed and replaced with one whose ability is uncertain. The desirability of a higher bar then turns, in part, on whether a particular incumbent's ability has been observed to be high or low. For a parameterization consistent with observed differences in research output, and for all but very low research discount factors, the optimal contract is again tenure-track, with a lower (but still positive) tenure bar set for those assessed to be of higher ability. We then discuss the implications of the imperfect alignment between the interests of the university and the interests of faculty whose input is sought in a tenure decision.

A third set of extensions establish that our model's parameterization is consistent with large differences in research output, but small differences in compensation across professors.

Section VII discusses how our conditional results point to a number of possible explanations of the recent and ongoing shift away from the use of tenure-stream faculty, while Section VIII provides some concluding remarks.

This paper significantly extends the analysis in Cater, Lew and Smith (2008), who examine a related but much simpler model that abstracts from all questions related to effort, incentives, knowledge decay, and heterogeneity in ability, research output, and compensation. That paper offers only a limited solution to Question 1 that holds only under very strict and unrealistic conditions, and it provides no answers to Questions 2, 3 and 4.

II The model

A representative university

A representative university is created by a one-time physical capital endowment — say, a building — from a government or philanthropist that is sufficient to support the research and teaching activities of a fixed number of faculty 'slots'. Without loss of generality, we let that number be one. The university operates under the terms of a charter that directs it to produce and impart academic knowledge in perpetuity.

The results of faculty research are not appropriable.[6] The university's only revenues are its instruction tuitions, defined here to include the tuition paid by students as well as any government subsidies that are contingent on student enrolment. Its only costs are the wages of its faculty.

[6] We abstract from the possibility that a research discovery *can* be commercialized to ask why tenure is granted in academic fields, such as philosophy, public policy and pure mathematics, where research yields *no* appropriable results, and why tenure arose decades before the Bayh-Dole Act of 1980 first gave academic institutions the right to patent and commercialize the results of government-subsidized research.

The university is a price-taker in a competitive instruction market. Faculty are hired from an infinite pool of potential professors, drawn from overlapping generations, each with a working lifetime of three periods. Hiring and retention decisions can be made only at the beginning of a period. A professor's wages can be conditioned on her observable research output. Employment contracts are enforceable before the courts.

The university seeks to maximize its expected profits per period.[7] Its rate of discount is zero and it expects to live for infinitely many periods.

A representative professor

In each period of her working life, a representative professor will occupy either an academic or a nonacademic job. An academic career can only begin in the first period of a professor's working life and, as in Carmichael (1988), cannot be resumed after any period of nonacademic work because academic skills decay. The nonacademic option always exists; its per period maximized utility is a constant C_o that sets a floor on the utility that academic employment must provide.

A representative professor receives one offer of academic employment at the beginning of the first period of her working life. At the beginning of any period immediately following academic work, the university that employed her in the previous period may attempt to retain her, and competitor universities may also attempt to hire her.[8]

Our representative professor has a zero rate of discount and is, as in Carmichael (1988), risk averse.[9] Her constant marginal utility of money is normalized to 1. In choosing between alternative employment offers and various levels of effort, she will, therefore, attempt to maximize her expected lifetime income, less any research effort disutility.

During any period of academic employment, our representative professor also provides instruction, the disutility of which is a constant D. We normalize the professor's utility scale so that $C_o + D = 0$.

An initial academic offer that yields a lifetime expected income, net of effort disutility, of 0 is thus sufficient to initially attract a 'junior' professor. Similarly, because a professor faces infinitesimally small but positive job change costs, a university can prevent a 'middle-aged' or 'senior' incumbent from quitting to pursue nonacademic employment by offering her an expected income, net of the costs of any effort, of 0 at the beginning of the second

[7]Although there is precedent in the literature for an assumption of profit maximization (e.g., Rothschild and White (1995) and Siow (1998)), some additional justification is warranted. In the context of our model, we can think of a university as remaining viable only if accounting profits are non-negative, and as using any positive profits to advance the university's mission by expanding its physical capital (and therefore its number of supported faculty slots, its supply of instruction, and its production of research). Our representative university will then operate in a zero-accounting-profit equilibrium where it must maximize its expected profits per period.

[8]In our model, where a professor's research record is publicly observable, there is no meaningful distinction between an outside university raiding for a professor and a professor seeking employment with an outside university. It is, therefore, sufficient to consider only the implications of raiding.

[9]Tenure is, of course, surely appealing to professors who are risk-averse. But to allow an assumption of risk-aversion to drive an explanation of tenure would be unsatisfactory, for how would be explain why professors are extended the tolerance of tenure, but risk-averse workers in other professions are not?

or third period of her working life, while it can prevent itself from being 'raided' for a particular professor by a competitor university by matching any external offer.

Research effort and production

At the beginning of any period of academic employment, t $(= 1, 2, 3)$, a representative professor chooses an unobservable level of research effort, e_t (≥ 0), the quadratic utility cost of which is e_t^2. At the end of that period, the professor then realizes a publicly observable research output, described by a single quality-adjusted index, r_t, that is drawn randomly from the probability distribution ρ_{e_t} on $[0, \infty)$, assumed to come from either the uniform, exponential, or power-law families. These three probability distributions, each intuitively plausible and analytically tractable, are chosen to demonstrate that our results are robust across different models of intellectual creativity.

Tuition revenues

Other things equal, students prefer to be taught by more knowledgeable faculty. Faculty knowledge is unobservable. But we assume that knowledge is correlated with observable research output because the former is accumulated through the research process.[10] Students are, therefore, attracted to schools with faculty who have stronger research records, resulting in greater aggregate tuition revenue.[11] This attraction may result because students observe the research records of faculty directly or because they simply choose a university on the basis of independent school rankings that are themselves based, in part, on research output and its correlates.

Provided she remains in academic employment, the research output that a professor realizes at the end of any period will be, in effect, immediately considered by potential students who make their enrolment decisions and pay their tuition at the beginning of the next period. In any subsequent periods of academic employment, a professor retains all knowledge she has accumulated through her research. The revenue impact of the signalled knowledge may, however, decay over time, either as the knowledge signal somehow becomes less effective in reaching potential students or as the particular knowledge that is signalled becomes less valuable to them.

[10] Alternatively, we could allow professors to differ in unobservable, pre-existing knowledge. Because greater knowledge would reduce the cost of doing research, greater knowledge could be inferred from observations of greater research output.

[11] Even in a tuition-regulated environment, where the tuition an individual student pays is (almost) orthogonal to research production, a successful researcher can generate more aggregate tuition by attracting more students. Support for this idea can be found in, for example, the *Globe and Mail Canadian University Report 2013: Class Size*, in which undergraduate students give far poorer class size grades to elite research institutions than to institutions with more of a teaching focus where faculty have research profiles that are less strong. Where tuitions can vary across institutions, we may, where other things are equal, see stronger researchers generating more aggregate tuition *and* more tuition per student. But because some students may prefer smaller class sizes, we may also see a more accomplished researcher at a a research-focused school characterized by larger class sizes generating more aggregate tuition, but *less* tuition per student, than a less accomplished researcher at a teaching-focused school with smaller class sizes.

Formally, in any period of academic employment, any professor generates a 'base' level of tuition revenues, which we normalize to 0. Adding to that base, each unit of her research output realized at the end of a period will contribute tuition revenues of k_τ at the beginning of the τ^{th} period after that output is realized, where $k_\tau = \overline{k}\delta^{\tau-1}$, $\overline{k} \geq 0$, and the discount factor $\delta \in [0, 1]$. Research output realized at the end of the final period of a professor's working life has no impact on the university's revenues; the knowledge it signals leaves with her.[12] Knowledge accumulated through academic research is of no value in nonacademic employment, as in Carmichael (1988).

For the purposes of sections III, IV, and V, we will assume that the knowledge signalling value of any given research accomplishment does not diminish over time – that is, $\delta = 1$ (i.e., $k_1 = k_2 = k$).

III Contracts, incentives, and profits

Implications of Potential Raiding

Lazear (1986), Bernhardt and Scoones (1993) and Waldman (1990) each describe a situation where one firm initially employs a worker who might be raided by an outside firm. Each of those papers establish that raiding will occur only if the worker is a better match with the outside firm. Where match-quality is equal across the firms, the initial employer creates a contract to pre-empt raiding.

In our model, there is no match-quality heterogeneity — a professor has no preference for one university over another, and both her research potential and the revenues her past research accomplishments can generate are independent of the university that employs her. Equilibrium raiding is, therefore, inefficient, reducing the profits that a professor can contribute to her academic employer(s) over the course of her career.

It is straightforward to show that if our representative university must choose between a strategy of hiring 'junior' professors under an arrangement that front-loads profits (by deferring compensation), hiring 'junior' professors under an arrangement that 'back-loads' profits, or raiding other universities for 'middle-aged' or 'senior' professors those other universities wish to retain, the only equilibrium is one where all universities hire 'junior' professors under arrangements that defer compensation and that are 'raid-proof'. We will thus proceed by solving for the university's profit-maximizing, raid-proof contract.[13]

[12] We abstract from rare cases where a university may continue to realize revenues from its association with a particularly accomplished professor even after her retirement.

[13] University faculty experience fewer job changes than workers in general (see, the estimates in Barbezat and Hughes (2001) versus those in Polachek and Siebert (1993), respectively). There may be many other reasons for this — there is, after all, essentially one type of job in academia (excluding administration), and universities are typically long-lived, while firms may exit. But our raid-proof feature is generally consistent with the observed turnover differential. The raid-proof feature also appears to be consistent with the observation that when academics do transition between jobs, that move is, if anything, associated with a small salary reduction (Barbezat and Hughes, 2001). Cases of mobility between tenured positions that result in wage increases may result from match-quality issues that are beyond the scope of this paper.

Academic contracts

When attempting to hire a 'junior', a university considers two things: (1) the conditions, if any, under which it wishes to retain an incumbent professor into subsequent periods of her working life, and (2) the wage structure necessary to recruit her initially, to induce her 'optimal' effort, and to successfully retain her when her retention is sought.

The retention conditions are assumed to take the form of minimum research production standards that a professor's *most recent* research realization must equal or exceed for the university to wish to retain her into (at least) the following period.

Compensation is a period-specific base wage and bonus structure that is linear in research output. This 'linear' incentive structure gives us a tractable model. Macleod and Malcomson (1989), Pearce and Stacchetti (1998), and Hogan (2002) consider similar payment schemes, where only the base wage is part of the explicit contract. The bonus for unobservable effort is promised only 'implicitly', but, in repeated interaction, it is in the best interest of the firm to honor the implicit component. In our model, the bonus is tied to observable research output, so both the base wage and bonus components are explicit.[14]

Because the 'raid-proof' equilibrium involves the front-loading of profits, a potential time-inconsistency problem will arise. That is, once a profit-maximizing university 'underpays' a professor at the beginning of her career and realizes higher profits than it would if it were to retain her and pay deferred compensation in subsequent periods, it will, in the absence of any contractual obligation, *not* retain the professor and thus not pay her deferred compensation in a subsequent period(s). This will, of course, undermine the incentive for a 'junior' professor to exert any costly research effort.

To solve that time inconsistency problem, our representative university commits to the *academic contract*, $\mathbf{C} := (w_1,\, w_2,\, w_3;\, b_1,\, b_2;\, b_{21},\, b_{31},\, b_{32};\, s_1,\, s_2)$, comprised of base wages $(w_1,\, w_2,\, w_3)$, bonus multipliers $(b_1,\, b_2;\, b_{21},\, b_{31},\, b_{32})$, and retention standards $(s_1,\, s_2)$.

A professor who accepts $\mathbf{C}$ will receive a salary of

$$S_1(r_1) \quad := \quad w_1 + b_1 r_1 \tag{1}$$

at the end of her first period of employment. In the event that her first research draw $r_1 \geq s_1$, she then has the option of remaining with the university through her second period. If she chooses to remain, she receives a salary of

$$S_2(r_1, r_2) \quad := \quad w_2 + b_2 r_2 + b_{21} r_1 \tag{2}$$

at the end of that period. Similarly, if her $r_2 \geq s_2$, she is given the option of remaining with the university through the third and final period of her working life. Taking that option results in a salary of

$$S_3(r_1, r_2) \quad := \quad w_3 + b_{31} r_1 + b_{32} r_2. \tag{3}$$

[14] In a very general principal-agent model, Holmstrom and Milgrom (1987) have shown the optimality of a linear compensation scheme based on aggregate performance. Strictly speaking, their result (which assumes a Gaussian distribution for the agent's output, and no possibility of termination) is not applicable to our model (which assumes non-Gaussian distributions and does allow for termination); however, it makes our linear incentive structure a plausible simplifying assumption.

received at the period's end. Because r_3 can generate no revenues for the university, the contract contains no bonus for r_3.

The contract's general structure places no restrictions on the timing of research bonuses tied to r_1 and r_2, allowing them to be paid, if at all, immediately upon the research realization and/or in any subsequent period of retention. Moreover, although an academic employment contract cannot enumerate every possible combination of research quantity and quality that would warrant retention, we model retention standards as being explicit to focus on the question of why a positive or a zero standard would ever be chosen.[15]

Research incentives and expected profits

For any level of research effort, e (≥ 0), let $\overline{r}(e) := \int_0^\infty r\ d\rho_e[r]$ be the expected value of r, where ρ_e is the probability distribution of research output. For any $s \geq 0$, let

$$P(e,s) \ := \ \int_s^\infty d\rho_e[r] \quad \text{and} \quad \overline{R}(e,s) \ := \ \frac{1}{P(e,s)} \int_s^\infty r\ d\rho_e[r] \tag{4}$$

be, respectively, the probability that $r \geq s$ and the expected value of r given that $r \geq s$.

To maximize her lifetime expected net benefit under **C**, the professor must choose optimal effort levels for each period. To do this, she solves a dynamic programming problem, starting with period 3 of her working life and working backward.

Period 3. – Suppose the professor has been retained under **C** through the first two periods of her career, and that her $r_2 \geq s_2$. If she were to remain with the university, then, in the absence of any bonus for third period research production, her optimal e_3 would be 0, so the net benefit of remaining with this university for the third period of her career would be

$$\text{NB}_3(r_1, r_2) \ \underset{(3)}{=} \ w_3 + b_{31}r_1 + b_{32}r_2. \tag{5}$$

To ensure that the professor would not instead choose to pursue nonacademic employment, it is necessary that

$$\text{NB}_3(r_1, r_2) \quad \geq \quad 0, \quad \forall\ r_1, r_2 \geq 0. \tag{6}$$

Period 2. – Now suppose that a professor has completed the first period of **C**, that her $r_1 \geq s_1$, and that (6) is satisfied. The expected net benefit of choosing to remain with the university for (at least) the second period of her working life would be

$$\text{NB}_{2,3}(r_1, e_2) \quad = \quad \text{NB}_2(r_1, e_2) + P(e_2, s_2)\,\text{NB}_3(r_1, e_2), \tag{7}$$

where

$$\text{NB}_2(r_1, e_2) \quad := \quad w_2 + b_2\overline{r}(e_2) + b_{21}r_1 - e_2^2 \tag{8}$$

is the net benefit of period 2 employment alone, and

$$\text{NB}_3(r_1, e_2) \quad := \quad w_3 + b_{31}r_1 + b_{32}\overline{R}(e_2, s_2).$$

[15] While an interesting and important issue to address, the incentive effects of the incompleteness of an academic contract is beyond the scope of this paper.

is the expected value of (5), given period-2 effort e_2. If the professor were to choose to remain with the university, she would then choose her optimal level of period-2 research effort, e_2^*, so as to maximize (7). For the university to ensure that a professor will not pursue her nonacademic option at this stage, the contract must satisfy

$$\mathrm{NB}_{2,3}(r_1, e_2^*) \quad \geq \quad 0, \quad \forall\, r_1 \geq 0. \tag{9}$$

Period 1. – Suppose (6) and (9) are satisfied. For a 'junior' professor, knowing that her period-2 research effort will be e_2^*, the expected net benefit of $\mathbf{C}$, given period-1 effort e_1, is

$$\mathrm{NB}_{1,2,3}(e_1, e_2^*) \quad = \quad \mathrm{NB}_1(e_1) + P(e_1, s_1)\,\mathrm{NB}_{2,3}(e_1, e_2^*). \tag{10}$$

Here,

$$\mathrm{NB}_1(e_1) \quad := \quad w_1 + b_1\overline{r}(e_1) - e_1^2 \tag{11}$$

is the net benefit of period 1 alone, while

$$\mathrm{NB}_{2,3}(e_1, e_2^*) \;:=\; w_2 + b_2\overline{r}(e_2^*) + b_{21}\overline{R}(e_1, s_1) - (e_2^*)^2 + P(e_2^*, s_2)\left(w_3 + b_{31}\overline{R}(e_1, s_1) + b_{32}\overline{R}(e_2^*, s_2)\right)$$

is the expected value of (7), given period-1 effort e_1 and anticipating optimal period-2 effort e_2^*. If the 'junior' professor were to accept $\mathbf{C}$, she would choose her optimal level of period-1 research effort, e_1^*, so as to maximize (10). It will be rational for the potential 'junior' professor to accept the academic offer if and only if $\mathrm{NB}_{1,2,3}(e_1^*, e_2^*) \;\geq\; 0$.

To the university, w_1 represents a cost that has no influence on the professor's choice of effort profile, (e_1^*, e_2^*). To minimize its costs, the university will set $w_1 := -b_1\overline{r}(e_1^*) + (e_1^*)^2 - P(e_1^*, s_1)\,\mathrm{NB}_{2,3}(e_1^*, e_2^*)$, so that the contract satisfies the *minimal recruitment* condition:

$$\mathrm{NB}_{1,2,3}(e_1^*, e_2^*) \quad = \quad 0. \tag{12}$$

We will say that $\mathbf{C}$ is *admissible* if it satisfies (6), (9) and (12). Period-specific expected profits, as of the beginning of each of the contract's three periods, are then given by:

$$\overline{\Pi}_1 \quad = \quad -w_1 - b_1\overline{r}(e_1^*), \tag{13}$$

$$\overline{\Pi}_2(r_1) \quad = \quad -w_2 - b_2\overline{r}(e_2^*) + (k_1 - b_{21})r_1 \qquad \text{and} \tag{14}$$

$$\overline{\Pi}_3(r_1, r_2) \quad = \quad -w_3 + (k_2 - b_{31})r_1 + (k_1 - b_{32})r_2. \tag{15}$$

The contract $\mathbf{C}$ is said to be *raid-proof* if, for all $r_1, r_2 \geq 0$, we have $\overline{\Pi}_2(r_1) \leq \overline{\Pi}_1$ and $\overline{\Pi}_3(r_1, r_2) \leq \overline{\Pi}_1$ (where these quantities are as defined in equations (13-15)).

IV Analysis

We say that a contract is *tenure-track* if $s_1 > 0$ and $s_2 = 0$ (or, equivalently, $0 < P(e_1, s_1) < 1$ and $P(e_2, s_2) = 1$ for any $e_1, e_2 \geq 0$). We say that the contract induces a *declining effort profile* if $e_1^* > e_2^*$. We now come to our main result.

Theorem 1 *Suppose $\delta = 1$. Let $\{\rho_e\}_{e\in\mathbb{R}_{\neq}}$ be a family of probability distributions on $[0,\infty)$, and let $\mathbf{C}$ be an admissible, raid-proof contract that maximizes expected profits per period.*

- **(a)** *For all $e \geq 0$, suppose ρ_e is the uniform probability distribution on $[0,e]$.* (That is, $d\rho_e(r) = 1/e$ if $r \in [0,e]$ and $d\rho_e(r) = 0$ if $r > e$.) *Then $\mathbf{C}$ is tenure-track, with a declining effort profile.*
- **(b)** *For all $e \geq 0$, suppose ρ_e is the exponential probability distribution $d\rho_e(r) = \frac{1}{e}\exp(-r/e)$. Then $\mathbf{C}$ is tenure-track, with a declining effort profile.*
- **(c)** *For any $\alpha > 1$ and $e \geq 0$, let ρ_e^α be the power law distribution $d\rho_e^\alpha(r) = \frac{e^\alpha\,\alpha}{(e+x)^{\alpha+1}}$. There exist $\underline{\alpha}, \overline{\alpha} \in (1,\infty)$ such that, if $\alpha \in (1,\underline{\alpha})$ or $\alpha \in (\overline{\alpha},\infty)$, then $\mathbf{C}$ is tenure-track, with a declining effort profile. In particular, this holds if $\alpha = 2$.*

In Theorem 1, we see that our representative university will both adopt a tenure-track contract and induce declining effort that results in research production declining, on average, over the life cycle. Intuitively, in choosing the retention standards to apply at the beginning of the second and third periods, the university faces a trade-off: raising the minimum standard applied to the most recent research draw raises that draw's conditional mean contribution to future revenues, but it comes at the cost of foregoing any additional, future revenues from the past research accomplishments of those who fail to meet the current standard. The resolution of this tradeoff then depends crucially on the professor's cumulative research record. For a professor entering her second working period, the application of a high minimum standard to her first period research draw is optimal, for she has no past research accomplishments to forego. But for a professor who is entering her third working period, the application of a zero minimum standard to her second period research draw is optimal, for it allows the university to continue to profit from the first period accomplishments that enabled her to clear the high first standard.[16]

Moreover, in each of the first and second periods of the contract, the university will induce a professor's research effort up to the point where the resulting marginal revenue is equal to her (increasing) marginal cost. The marginal cost of a given unit of effort is independent of the period in which that effort is exerted. But because any resulting research output can be translated into tuition revenues only in the period(s) *subsequent* to its realization, the marginal revenue from a unit of second period research output will be lower than that which results from a unit of research output produced in the first period. The profit-maximizing level of induced research effort will, therefore, be lower in the second period than in the first, and research output will be observed to decline, on average, with age.[17]

[16] If tenure is granted because early research success allows an older professor to continue to generate revenues, one might ask: why did U.S. universities practice mandatory retirement prior to its abolition? A similar question could be asked of *any* employer who chose to retire *any* productive worker. Lazear (1979) shows that where there is an incentive for an employer to defer compensation, as there is in our model, the practice of mandatory retirement is efficient.

[17] An alternative, but much more involved, model specification would see a professor making decisions about allocating time between exploiting and investing in human capital. A declining research output profile could then result, for as she ages, a professor would have less incentive to invest and her human capital would become more and more dated.

Evidence consistent with the notion that declining observed output profiles result from an optimally declining induced research effort, rather than from a disincentive effect of tenure, is presented in Goodwin and Sauer (1995), Hutchinson and Zivney (1995), and Beckman and Schneider (2012), who find a near-continuous pattern of research output around the time tenure is granted. Further evidence consistent with our idea that a university will induce declining effort by rewarding early research successes more generously than later ones is presented in Siow (1991), who finds that the long-run increase in salary from an additional publication declines with the age at which the article is published.

V Outline of Theorem 1 proof

The proof of Theorem 1 is long and appears in the appendix. This section, however, describes the basis for that proof and outlines the major steps involved.

Recall that our representative university operates in a raid-proof equilibrium. In any period, the university will find itself in one of three 'states': its single faculty 'slot' will be occupied by a 'junior' professor (state 1), a 'middle-aged' professor (state 2), or a 'senior' professor (state 3). Whenever a 'junior' ('middle-aged') incumbent is retained into the following period, the university will transition from state 1 (2) to state 2 (3). If the university cannot 'raid' from other universities, then it can only hire junior professors; thus, whenever any incumbent is not retained into the following period, the university returns to state 1. If other universities will not 'raid' from our representative university, then the probability of retaining a professor is exactly the probability that her research exceeds the minimum standards s_1 and s_2 specified by the contract. Thus, the retention probabilities are $p_1 := P(e_1^*, s_1)$ and $p_2 := P(e_2^*, s_2)$. This system defines a 3-state Markov process with transition probability matrix

$$\begin{bmatrix} 1-p_1 & p_1 & 0 \\ 1-p_2 & 0 & p_2 \\ 1 & 0 & 0 \end{bmatrix}. \tag{16}$$

This process has stationary probability distribution (π_1, π_2, π_3) given by

$$\pi_1 = \frac{1}{1+p_1+p_1p_2}, \quad \pi_2 = \frac{p_1}{1+p_1+p_1p_2}, \quad \text{and} \quad \pi_3 = \frac{p_1p_2}{1+p_1+p_1p_2}. \tag{17}$$

Equation (14) gave the expected period-2 profit at the start of period 2 — that is, once the realization of r_1 is already *known*. Similarly, (15) gave the expected period-3 profit at the start of period 3, when the realizations of r_1 and r_2 are both known. However, at the start of period 1, the future values of r_1 and r_2 are both unknown. At that moment, assuming $k_1 = k_2 = k$, the expected profits which **C** will generate in each of three periods of a professor's career are

$$\begin{aligned} \overline{\Pi}_1 &\underset{(13)}{=\!=} -w_1 - b_1\overline{r}(e_1^*); \\ \overline{\Pi}_2 &\underset{(14)}{=\!=} -w_2 - b_2\overline{r}(e_2^*) + (k - b_{21})\overline{R}(e_1^*, s_1); \\ \text{and}\ \overline{\Pi}_3 &\underset{(15)}{=\!=} -w_3 + (k - b_{31})\overline{R}(e_1^*, s_1) + (k - b_{32})\overline{R}(e_2^*, s_2). \end{aligned} \tag{18}$$

Combining (18) and (17), the expected profit per period of the university is given by

$$\overline{\Pi}(\mathbf{C}) \quad := \quad \pi_1\overline{\Pi}_1 + \pi_2\overline{\Pi}_2 + \pi_3\overline{\Pi}_3. \tag{19}$$

The university must find the (raid-proof) contract which maximizes the value of $\overline{\Pi}$. The proof of Theorem 1 will proceed in three steps:

1. We relax the need to optimize over raid-proof contracts, by showing that a non-raidproof contract can be 'retroactively raidproofed' without affecting its optimality.

2. We show that it suffices to solve the optimization problem over a particularly desirable class of contracts we call MNQ ('minimal no-quitting').

3. We establish Theorem 1 for the class of MNQ contracts.

Steps 1 and 2 both use the concept of *contract equivalence*. Let $\mathbf{C}$ and $\widetilde{\mathbf{C}}$ be two academic contracts. We say that $\mathbf{C}$ and $\widetilde{\mathbf{C}}$ are ***equivalent*** if:

(Eq1) In both contracts, the professor's optimal effort profile (e_1^*, e_2^*) is the same.

(Eq2) Both contracts have the same research standards (s_1, s_2).

(Eq3) Both contracts yield the same expected lifetime net benefit $\mathrm{NB}_{1,2,3}$ for the professor.

In particular, (Eq2) implies that $\mathbf{C}$ is tenure-track if and only if $\widetilde{\mathbf{C}}$ is also tenure-track. (Eq3) implies that $\mathbf{C}$ satisfies minimal recruitment condition (12) if and only if $\widetilde{\mathbf{C}}$ does.

Lemma 2 *If contracts* $\mathbf{C}$ *and* $\widetilde{\mathbf{C}}$ *are equivalent, then both contracts yield the same value of* $\overline{\Pi}$ *in equation* (19). (Thus, $\mathbf{C}$ is $\overline{\Pi}$-maximizing if and only if $\widetilde{\mathbf{C}}$ is.) □

The next proposition accomplishes Step 1 in our proof strategy. Recall that $\overline{r}(e) := \int_0^\infty r \; d\rho_e[r]$.

Proposition 3 *Assume* $\overline{r}'(e) \neq 0$ *for all* $e \geq 0$. *Let* $\mathbf{C}$ *be any admissible, tenure-track contract which is not raid-proof. There exists an admissible, raid-proof contract* $\widetilde{\mathbf{C}}$ *which is equivalent to* $\mathbf{C}$ (and hence, is also tenure-track). □

Proposition 3 says that, to demonstrate that the raid-proof $\overline{\Pi}$-maximizing contract is tenure-track, it suffices to first find a *non*-raid-proof contract which maximizes $\overline{\Pi}$ by being tenure-track, because we can always 'retroactively raidproof' it later.

We will focus on a class of contracts which are especially easy to optimize. We say that $\mathbf{C}$ is a ***minimal no-quitting*** (MNQ) contract if the conditions (6) and (9) are satisfied with *equalities* —that is,

$$\mathrm{NB}_{2,3}(r_1, e_2^*) = 0, \quad \text{and} \quad \mathrm{NB}_3(r_1, r_2) = 0, \quad \forall\, r_1, r_2 \geq 0. \tag{MNQ}$$

If $\mathbf{C}$ satisfies (MNQ), then $\mathrm{NB}_{2,3} = \mathrm{NB}_2$ and $\mathrm{NB}_{1,2,3} = \mathrm{NB}_1$; this will make it much easier to characterize (and control) the professor's utility-maximizing effort profile (e_1^*, e_2^*).

Define $\beta : (0,\infty) \longrightarrow (0,\infty)$ by $\beta(e) := 2e/\overline{r}'(e)$ for all $e > 0$. We will require the family of distributions $\{\rho_e\}_{e \in \mathbb{R}_{\neq}}$ to satisfy the following assumption:

$$\beta \textit{ is a bijection from } (0,\infty) \textit{ to } (0,\infty). \tag{B}$$

One way to satisfy (B) is for β to be strictly increasing, with $\lim_{e \searrow 0} \beta(e) = 0$, and $\lim_{e \nearrow \infty} \beta(e) = \infty$. This just means that there are *not* strongly increasing returns to effort —a very weak assumption. It is easy to check that all the distribution families in Theorem 1 satisfy (B). The next proposition accomplishes Step 2 in our strategy.

Proposition 4 *Suppose* $\{\rho_e\}_{e \in \mathbb{R}_{\neq}}$ *satisfies* (B).

- **(a)** *Let* $\mathbf{C}$ *be any contract satisfying minimal recruitment condition* (12)*. There is a MNQ contract* $\widetilde{\mathbf{C}}$ *equivalent to* $\mathbf{C}$.
- **(b)** *Let* $\mathbf{C}$ *be a profit-maximizing contract in the space of all admissible contracts. Let* $\widetilde{\mathbf{C}}$ *be a profit-maximizing contract in the space of all admissible MNQ contracts. Then* $\widetilde{\mathbf{C}}$ *provides the same expected profit per period as* $\mathbf{C}$. □

If hypothesis (B) holds, then Proposition 4(b) implies that, to find the $\overline{\Pi}$-maximizing contract, it suffices to maximize $\overline{\Pi}$ over the set of admissible MNQ contracts. For any MNQ contract, it can be shown that $b_{21} = b_{31} = b_{32} = w_3 = 0$, while the values of w_1 and w_2 are entirely determined by b_1 and b_2 (see Lemma A in the Appendix). Thus, an MNQ contract has only four free parameters: b_1, b_2, s_1, and s_2. Furthermore, we can achieve any desired effort profile (e_1, e_2) and retention probabilities (p_1, p_2) with a suitable choice of parameters $(b_1, b_2; s_1, s_2)$ (see Lemma B in the Appendix). Thus, the space of MNQ contracts can be parameterized by the set of all 4-tuples $(e_1, e_2; p_1, p_2)$. When an MNQ contract is expressed in this form, $\overline{\Pi}$ can be expressed as a function $\overline{\Pi}(e_1, e_2; p_1, p_2)$. With a mild technical assumption, we can then define functions $e_1^* : [0,1]^2 \longrightarrow \mathbb{R}_{\neq}$ and $e_2^* : [0,1]^2 \longrightarrow \mathbb{R}_{\neq}$ such that, for any fixed (p_1, p_2), the values of the parameters (e_1, e_2) which maximize $\overline{\Pi}(e_1, e_2; p_1, p_2)$ are $e_1^*(p_1, p_2)$ and $e_2^*(p_2)$ (see Lemma C). At this point, the $\overline{\Pi}$-maximization problem is reduced to finding the values of p_1^* and p_2^* in $[0,1]$ which maximize the function $\hat{\Pi}(p_1, p_2) := \overline{\Pi}[e_1^*(p_1, p_2), e_2^*(p_1, p_2); p_1, p_2]$. If the family of probability distributions $\{\rho_e\}_{e \in \mathbb{R}_{\neq}}$ and the derivative $\partial_2 \hat{\Pi}$ satisfy certain technical conditions, then the $\hat{\Pi}$-maximizing value of p_2 is $p_2^* = 1$ —in other words, the $\overline{\Pi}$-maximizing MNQ contract is tenure track (see Lemma E(a)). Furthermore, if p_1^* and p_2^* then satisfy certain conditions (in particular, if $p_1^* > 1/2$) then the $\overline{\Pi}$-maximizing MNQ contract induces a declining effort profile (see Lemma E(b)).

In particular, the uniform, exponential, and power-law families of distributions all satisfy the technical conditions required by Lemma E; thus, for all three families of distributions, the $\overline{\Pi}$-maximizing element in the space of MNQ contract is tenure-track, and induces a declining profile of effort (see Lemmas F, G, and H). In other words, the conclusions of Theorem 1 hold for the space of MNQ contracts. Then Proposition 4(b) implies that the conclusions of Theorem 1 hold for the space of *all* contracts. Finally, Proposition 3 implies that the conclusions of Theorem 1 hold for the restricted space of raid-proof contracts; this establishes Theorem 1.

VI Extensions

We now turn to a number of extensions to our basic model. To conserve on space, we present only brief summaries of the numerical analyses; the full results are expounded in Cater, Lew and Pivato (2015).

Generational length and decaying knowledge signals

Thus far, we have assumed that professors live for three periods and that the revenue impact of a research accomplishment does not diminish over time. We now modify these assumptions to allow for more frequent assessments of performance and for the tuition-generating value of a knowledge signal to decay over time.

To embed these modifications, we introduce a six-period generational length into the model presented in Sections II and III, and we allow the discount factor, δ, to be smaller than 1. It turns out that for each of our distributional cases (i.e., uniform, exponential, and power-law), and for δ values down to 0.4 (i.e., where the remaining tuition-generating value of signalled knowledge effectively decays at a rate of 60 percent per period), tenure is optimally granted to those who meet a positive standard of research production at the end of one period. The robustness of this result is importantly consistent with the broad use of tenure-based contracts across fields that differ in terms of their rate of knowledge decay (McDowell, 1982), and the timing of the tenure decision more closely corresponds to what we see in practice. Our results further show that a declining effort profile is optimal, induced by declining research bonuses that are consistent with Siow (1991).

For discount factors below 0.4, the ongoing tuition-generating value of research success falls off sharply, so the university sets positive minimum standards throughout a professor's career, effectively insisting that she produce 'fresh' knowledge on an ongoing basis.

Filtering on ability

Heterogenous ability and the optimal contract

To this point, we have also modelled expected research output to be a stochastic function of effort only. This has meant that, for the incentives implicit in any particular contract, professors of a given generation can be distinguished in terms of their past, but not their expected future research draws. Retention rules can thus only be 'backward looking'.

However, faculty do surely differ in terms of their research ability, and, when making a retention decision, a university can consider a professor's observed research output to date *as well as* some assessment of her ability and her expected future research production. How, then, does a university balance the tension between these backward- and forward-looking considerations when initial research output may be a poor indicator of future research ability? What will it do with, say, an incumbent who is judged to be highly able but unlucky to date or whose record is thought to be misleadingly strong?[18]

[18] We are grateful to an anonymous referee for suggesting this line of inquiry.

To consider these questions, we suppose that all professors, upon their initial hiring, are equally able — during the first periods of their working lives, it is common knowledge that they each possess an 'average' research ability. After one period of employment, both the university and professor then realize whether the professor's research ability for the balance of her working life will be 'high' or 'low', with probabilities h and $1-h$, respectively, that are independent of her first period research draw. Research effort is as described in section II. But high (low) levels of research ability translate into higher (lower) expected research output from any given level of effort. For a professor of 'average' ability, research output follows a stochastic process as described in section II. But for professors of 'high' and 'low' ability, the expected value of research output that results from a given level of effort is scaled up and down by factors of $1 + a$ and $1 - a$, respectively, where $0 < a < 1$.

Setting $h = .5$ and $a = .1$, we find that for δ values from 1.0 down to 0.6, and for each of our three distributions, the tenure-track contact is optimal. But the university now effectively filters on ability by applying a lower (higher) tenure bar to a professor found to be of high (low) ability. Because she has a greater (lesser) expected research output going forward, it takes a lower (higher) stock of first period research production to make her more profitable, in an expected sense, than a potential replacement of unknown ability.

Imperfect alignment of university and faculty interests

Although a university makes the ultimate decision in a tenure case, it may rely on the recommendation of a professor's departmental colleagues who are better able to assess the value of both her first period research and her research ability.

If colleagues care strictly about a professor's ability and the expectation of future output, our analysis suggests that the university will not strictly abide by their recommendations to grant (deny) tenure to anyone assessed to be of high (low) ability. The university will deny tenure to a low ability professor, except where she initially produces a very high level of research output. In this way, the university effectively allows colleagues considerable 'downside' discretion in a case where a professor's file looks misleadingly strong. By contrast, because the university will insist that even a high ability professor initially meet a positive standard, it will, in effect, limit the 'upside' discretion of colleagues who see her as simply having been unlucky to date.

Research output and compensation differences

Our choice of $a = .1$ assumes a difference of roughly 20 percent between the mean output of low and high ability professors for a given level of effort. To determine whether that parameterization can be reconciled with observations that the research output gap between a department's "bottom" and "top" producers may be well in excess of 100 percent, note first that high ability professors not only produce more on average from a given level of effort, but also exert more *unobservable* effort. So, for example, for the case of an exponential distribution where $\delta = 0.8$, the combination of both *greater* and *more effective* effort mean that the expected research output of high ability faculty in, say, period 2 is almost 50 percent greater than is the expected output of low ability faculty.

What is equally important is that observed differences between the "top" and "bottom" producers are more about the tails of the pooled output distribution of all faculty, than they are about the difference between the means of the respective output distributions of those of high and low ability. For the probability-weighted pool of low and high ability faculty who meet their respective tenure standards, first (Q_1) and third (Q_3) quartile values of the current and cumulative research output distributions show, for all distributional cases and rates of decay for which the tenure-track contract is optimal, dramatic differences of far more than 100 percent between the "bottom" and "top" producers. This is true if we compare those within a given generation, and all the more so if we compare, say, the group of the most productive junior with the group of the least productive senior faculty. These results validate our parameterization as being consistent with observed large research output gaps.[19] Moreover, they suggest that the large gaps in output should not be interpreted as indicating that "mistakes" were made in granting tenure to some — such disparities are to be expected within an optimally-tenured pool.

To address the question of whether, in spite of these very large differences in research output, our model can generate compensation differences that are much smaller, we also simulated the income distribution for each output distribution and decay rate case where a tenure-track is optimal. In each case, the contract is raid-proofed by taking the bonus for each period's research output, and dividing it equally across the remaining periods of the contract. Our model expresses compensation both in unspecified units of money and relative to what is essentially the value of alternative employment. Within this context, it is trivial to then assign a dollar value to 1 unit of money and to the value of alternative employment that yields $Q_3/\,Q_1$ income ratios that are small relative to the $Q_3/\,Q_1$ research output ratios. Indeed, because our model places no restrictions on the size of the marginal compensation associated with inducing research effort relative to the value of alternative employment, it is consistent with salary differences that are proportionally smaller *or* larger than the underlying research gaps.

VII The changing faculty mix

We now turn to the question of how our conclusions regarding the efficiency of the tenure-track contract might be reconciled with universities' recent and ongoing shift away from that arrangement, towards the use of 'contingent' faculty — a group that includes full-time faculty who have both teaching and research duties but who are not tenure-eligible, and full- and part-time faculty who hold teaching-only positions.

Our results, while robust, are not unconditional. A university in our model can profit from inducing non-appropriable but costly research only where research successes signal the sort of knowledge students seek, and the tenure-track sequence of minimum standards is optimal only where a university expects that signal to remain effective long after the research successes are realized.

This suggests a number of possible explanations for the declining use of tenure-eligible

[19] For values of a closer to 0, tenure-track contracts remain optimal, with narrower research distributions, while for larger values of a, the tenure-track is optimal, but tenure is granted only to high ability faculty.

faculty. First, universities may now expect that current research will not signal the sort of knowledge students will be seeking decades down the road. This would cause universities to employ faculty who do research, but who are expected to produce 'fresh' research at every stage of their careers.

Second, in focusing on research because of its central role in the tenure puzzle, we have abstracted from the possibility that more and more of the curricula taught at universities — particularly in lower-level undergraduate courses — is becoming routine, standardized, 'canonical' material. Knowledge sufficient to deliver that curricula may be signalled by, say, a Ph.D. alone, without the need for a research record, creating an increasing role for teaching-only positions.

Third, the well-documented shift towards 'applied' programs (i.e., those that are relatively vocational in nature and that teach skill sets relevant to identifiable occupations), and away from the humanities, social sciences and pure sciences (Bertelson, 1998; Currie and Newson, 1998; Gumport, 1993 and 2000; Kerr, 1994; Nussbaum, 2010; Rhoades, 1998; Slaughter, 1993; and Slaughter and Rhoades, 1996) may have played a role. Our analysis suggests that the use of teaching-only faculty in 'applied' programs would be a profit-maximizing response to a situation where relevant 'applied' knowledge is more efficiently acquired through, and signalled by, 'field' or practical experience than it is by research. This is consistent with empirical evidence: In 'applied' program areas such as business and education, part-timers represent 46 and 48.7 percent of all faculty, respectively. By contrast, in the humanities, social sciences, and pure natural sciences — programs that focus more on the development of general skills including analytical reasoning — part-time (teaching-only) faculty represent only 34.6, 29.7, and 23.5 percent of all faculty, respectively. (See the 2004 National Study of Postsecondary Faculty (NSOPF: 04) Report on Faculty and Instructional Staff in Fall 2003.)

Fourth, where long term tuition revenues are in question — say, because forecasts of demographic shifts or increasing supply-side competition result in declining expected enrolments, or where shifts in public policy are expected to reduce future enrolment-based subsidies — our analysis suggests that universities will increase their use of faculty to whom only shorter-term commitments are made.

VIII Concluding remarks and future research

This paper has provided an explanation of the use of the tenure-track contract in academia that is rooted in the unique nature of academic work. We have argued that because a professor's research signals knowledge that attracts tuition-paying students, a university can profit from retaining a professor who initially establishes a strong research record, regardless of her research output thereafter.

Among other things, our analysis has offered a counter-argument to the view that observations of declining research output over the lifecycle are a result of the disincentive effects of tenure. Our argument is that because the opportunities to realize tuition revenues from a professor's spot research accomplishments diminish as she approaches the end of her career, the pattern results from the university optimally inducing a declining effort profile.

Tenure does not amount to absolute job security — while tenured professors are not dismissed for poor research productivity, Lovain (1983/84), Hendrickson (1988) and Morris (1992) note that they *are* dismissed for failing to perform their teaching duties. Our theory offers a simple explanation: the past research accomplishments of a tenured professor can be translated into the tuition revenues necessary to make her profitable only if she continues to teach.

Perhaps the most important implication of our analysis is that, where the conditions assumed in our model are met, the tolerance for research failure that characterizes tenure *is* consistent with a university's interest in advancing knowledge through research production. Although it might seem that a university could produce more research simply by replacing *any* unproductive scholar, or by providing older professors with greater research incentives, our analysis suggests that, by deviating from its profit-maximizing rule, either the university's long-term viability would be undermined or greater levels of ongoing subsidies would be required.

Our analysis also serves to correct the misperception that tenure's declining use is proof of its inherent inefficiency. A number of plausible explanations of the trend have been suggested, each consistent with our explanation of the use of tenure.

Finally, it is worth reiterating that our analysis has considered the behaviour of universities and the behaviour of professors under a set of reasonable, simplifying assumptions. Future work may shed greater light on these behaviours by examining the implications of risk aversion, non-linear compensation, contract incompleteness, and match-quality heterogeneity.

Appendix: Proofs

Proof of Lemma 2. Let $\overline{\Pi}$ be the expected profit per period under $\mathbf{C}$, as defined in eqn.(19). Let $\widetilde{\Pi}$ be the expected profit per period under $\widetilde{\mathbf{C}}$. Then clearly

$$\overline{\Pi} \;=\; \overline{R} - \overline{C} \quad \text{and} \quad \widetilde{\Pi} \;=\; \widetilde{R} - \widetilde{C}, \tag{2.1}$$

where $\overline{R}$ and $\widetilde{R}$ represent the university's expected *revenue* per period under the two contracts, while $\overline{C}$ and $\widetilde{C}$ represent the university's expected *costs* per period.

(Eq1) implies that the professor will exhibit the same probability distribution of research outputs; in particular she will have the same expected values $R_1^* := \overline{R}(e_1^*, s_1)$ and $R_2^* := \overline{R}(e_2^*, s_2)$. Then (Eq2) implies she will have the same retention probabilities (p_1, p_2) in both contracts. Thus equation (17) says both contracts have the same stationary probability distribution (π_1, π_2, π_3) over the three periods. Thus, assuming $k_1 = k_2 = k$, both contracts generate the same expected revenue per period, namely

$$\widetilde{R} \quad = \quad \pi_1 \cdot 0 + \pi_2 \cdot k\, R_1^* + \pi_3 \cdot k\,(R_1^* + R_2^*) \quad = \quad \overline{R}. \tag{2.2}$$

Let $\overline{S}_1, \overline{S}_2, \overline{S}_3$ denote the professor's expected salaries in the three periods, under $\mathbf{C}$.

Then $\overline{C}$ is simply the professor's expected salary per period, namely:

$$\overline{C} \quad = \quad \pi_1\overline{S}_1 + \pi_2\overline{S}_2 + \pi_3\overline{S}_3 \underset{(17)}{=\!=} \frac{\overline{S}_1 + p_1(\overline{S}_2 + p_2\overline{S}_3)}{1 + p_1 + p_1 p_2} \quad = \quad \frac{\overline{S}}{1 + p_1 + p_1 p_2},$$

where $\overline{S} := \overline{S}_1 + p_1(\overline{S}_2 + p_2\overline{S}_3)$ is the professor's expected lifetime salary in $\mathbf{C}$. Likewise, $\widetilde{C} := \widetilde{S}/(1 + p_1 + p_1 p_2)$, where $\widetilde{S}$ is the professor's lifetime salary in $\widetilde{\mathbf{C}}$. The professor's expected lifetime net benefit under the two contracts can be expressed by

$$\mathrm{NB}_{1,2,3} \quad = \quad \overline{S} - (e_1^*)^2 - p_1 \cdot (e_2^*)^2 \quad \text{and} \quad \widetilde{\mathrm{NB}}_{1,2,3} \quad = \quad \widetilde{S} - (e_1^*)^2 - p_1 \cdot (e_2^*)^2.$$

But (Eq3) says $\widetilde{\mathrm{NB}}_{1,2,3} = \mathrm{NB}_{1,2,3}$; hence $\widetilde{S} = \overline{S}$; hence $\widetilde{C} = \overline{C}$. Combining this with equations (2.1) and (2.2), we get $\widetilde{\Pi} = \overline{\Pi}$. □

Proof of Proposition 3. Let (e_1^*, e_2^*) be the utility-maximizing effort profile for $\mathbf{C}$. Let $r_1^* := \overline{r}(e_1^*)$ and $r_2^* := \overline{r}(e_2^*)$. If $\widetilde{\mathbf{C}}$ is equivalent to $\mathbf{C}$, then (e_1^*, e_2^*) will also be the utility-maximizing effort profile for $\widetilde{\mathbf{C}}$ (we will ensure this later). In that case, the expected profit of $\widetilde{\mathbf{C}}$ *before* each period will be given by:

$$\begin{aligned}
\overline{\Pi}_1 \quad &\underset{(13)}{=\!=} \quad -w_1 - b_1 r_1^*; \\
\overline{\Pi}_2(r_1) \quad &\underset{(14)}{=\!=} \quad -w_2 - b_2 r_2^* + (k - b_{21}) r_1; \\
\text{and} \quad \overline{\Pi}_3(r_1, r_2) \quad &\underset{(15)}{=\!=} \quad -w_3 + (k - b_{31}) r_1 + (k - b_{32}) r_2.
\end{aligned}$$

To make $\widetilde{\mathbf{C}}$ raid-proof, it suffices to ensure that $\overline{\Pi}_3(r_1, r_2) = \overline{\Pi}_2(r_1) = \overline{\Pi}_1$ for all $r_1, r_2 \geq 0$. To do this, we must set

$$b_{21} \quad := \quad b_{31} \quad := \quad b_{32} \quad := \quad k; \tag{3.1}$$
$$w_3 \quad := \quad w_1 + b_1 r_1^*; \qquad \text{and} \tag{3.2}$$
$$w_2 \quad := \quad w_1 + b_1 r_1^* - b_2 r_2^*. \tag{3.3}$$

The net benefit of contract $\widetilde{\mathbf{C}}$ for the professor during period 3 is then

$$\widetilde{\mathrm{NB}}_3(r_1, r_2) \quad = \quad w_3 + k r_1 + k r_2, \qquad \text{by (5) and (3.1).} \tag{3.4}$$

At the beginning of period 2, the value of r_1 is known, and the expected future value of $\widetilde{\mathrm{NB}}_3$, as a function of e_2, is given:

$$\widetilde{\mathrm{NB}}_3(r_1, e_2) \underset{(3.4)}{=\!=} w_3 + k r_1 + k\overline{r}(e_2). \tag{3.5}$$

Let $\widetilde{\mathrm{NB}}_{2,3}$ be the net benefit of $\widetilde{\mathbf{C}}$ at the start of period 2 (including the anticipated future benefit of period 3). By hypothesis, $\mathbf{C}$ is tenure-track (i.e. $p_2 = 1$); hence, to be equivalent, $\widetilde{\mathbf{C}}$ must also be tenure-track. In this case, the expected value of $\widetilde{\mathrm{NB}}_{2,3}$ at the beginning of period 2, as a function of e_2, is given:

$$\begin{aligned}
\widetilde{\mathrm{NB}}_{2,3}(r_1, e_2) \quad &\underset{(7)}{=\!=} \quad \widetilde{\mathrm{NB}}_2(r_1, e_2) + \widetilde{\mathrm{NB}}_3(r_1, e_2) \\
&\underset{(8,3.1)}{=\!=\!=} \quad w_2 + b_2\overline{r}(e_2) + kr_1 + \widetilde{\mathrm{NB}}_3(r_1, e_2) - e_2^2 \\
&\underset{(3.5)}{=\!=} \quad (w_2 + w_3) + 2k\, r_1 + (k + b_2)\overline{r}(e_2) - e_2^2 \\
&\underset{(3.2,3.3)}{=\!=\!=\!=} \quad 2w_1 + 2b_1 r_1^* - b_2 r_2^* + 2k\, r_1 + (k + b_2)\overline{r}(e_2) - e_2^2. \qquad (3.6)
\end{aligned}$$

Let s_1 be the period 1 standard of $\mathbf{C}$ (and hence, of $\widetilde{\mathbf{C}}$). If the professor exerts effort e_1 during period 1, and is retained during period 2, then the conditionally expected value of r_1, given this information, is $\overline{R}(e_1) := \overline{R}(e, s_1)$ [see eqn.(4)]. Thus, the expected future value of $\widetilde{\mathrm{NB}}_{2,3}$ at the beginning of period 1, as a function of e_1 and e_2, is given:

$$\widetilde{\mathrm{NB}}_{2,3}(e_1, e_2) \quad \underset{(3.6)}{=\!=} \quad 2w_1 + 2b_1 r_1^* - b_2 r_2^* + 2k\, \overline{R}(e_1) + (k + b_2)\overline{r}(e_2) - e_2^2. \qquad (3.7)$$

Let $\widetilde{\mathrm{NB}}_{1,2,3}$ be the lifetime net benefit of $\widetilde{\mathbf{C}}$ at the start of period 1 (including the anticipated potential future benefits in periods 2 and 3). For any $e \geq 0$, let $P(e) := p(e, s_1)$ [see eqn.(4)]. Thus, the expected value of $\widetilde{\mathrm{NB}}_{1,2,3}$, as a function of e_1 and e_2, is

$$\begin{aligned}
\widetilde{\mathrm{NB}}_{1,2,3}(e_1, e_2) \quad &\underset{(10,11)}{=\!=\!=} \quad w_1 + b_1\overline{r}(e_1) + P(e_1) \cdot \widetilde{\mathrm{NB}}_{2,3}(e_1, e_2) - e_1^2 \\
&\underset{(3.7)}{=\!=} \quad w_1 + b_1\overline{r}(e_1) + P(e_1)\left(2w_1 + 2b_1 r_1^* - b_2 r_2^* + 2k\, \overline{R}(e_1) + (k + b_2)\overline{r}(e_2) - e_2^2\right) - e_1^2 \\
&= \quad \Big(1 + 2P(e_1)\Big)\, w_1 + b_1\overline{r}(e_1) - e_1^2 \\
&\qquad + \; P(e_1)\left(2b_1 r_1^* - b_2 r_2^* + 2k\, \overline{R}(e_1) + (k + b_2)\overline{r}(e_2) - e_2^2\right). \qquad (3.8)
\end{aligned}$$

Let $p_1 := P(e_1^*, s_1)$ and let $\overline{R}_1^* := \overline{R}(e_1^*)$. If the professor exerted effort profile (e_1^*, e_2^*), then the expected lifetime net benefit of $\widetilde{\mathbf{C}}$ would be

$$\begin{aligned}
&\widetilde{\mathrm{NB}}_{1,2,3}(e_1^*, e_2^*) \\
&\quad \underset{(3.8)}{=\!=} \quad \Big(1 + 2P(e_1^*)\Big)\, w_1 + b_1\overline{r}(e_1^*) - (e_1^*)^2 \\
&\qquad\qquad + \; P(e_1^*)\left(2b_1 r_1^* - b_2 r_2^* + 2k\, \overline{R}(e_1^*) + (k + b_2)\overline{r}(e_2^*) - (e_2^*)^2\right) \\
&\quad = \quad (1 + 2p_1)w_1 + b_1 r_1^* - (e_1^*)^2 + p_1\left(2b_1 r_1^* + 2k\, \overline{R}_1^* + k r_2^* - (e_2^*)^2\right). \qquad (3.9)
\end{aligned}$$

The expected lifetime net benefit offered by contract $\mathbf{C}$ is $\mathrm{NB}_{1,2,3} = 0$, because $\mathbf{C}$ is admissible by hypothesis. We must also make $\widetilde{\mathrm{NB}}_{1,2,3} = 0$. For any values of b_1 and b_2, we can achieve this by setting

$$w_1 \quad = \quad w_1(b_1) \quad := \quad \frac{-b_1 r_1^* - p_1\left(2b_1 r_1^* + 2k\, \overline{R}_1^* + k r_2^* - (e_2^*)^2\right) + (e_1^*)^2}{1 + 2p_1}. \qquad (3.10)$$

At this point, $\widetilde{\mathbf{C}}$ has only two free parameters: b_1 and b_2. Substituting eqn.(3.10) into (3.7) and (3.8), we define, for all $b_1, b_2 \in \mathbb{R}$, the functions

$$\begin{aligned}
&\widetilde{\mathrm{NB}}_{2,3}(b_1, b_2; e_1, e_2) \\
&\qquad := \ 2w_1(b_1) + 2b_1 r_1^* - b_2 r_2^* + 2k\,\overline{R}(e_1) + (k+b_2)\overline{r}(e_2) - e_2^2, \quad \text{and} \qquad (3.11) \\
&\widetilde{\mathrm{NB}}_{1,2,3}(b_1, b_2; e_1, e_2) \quad := \quad \Big(1 + 2P_1(e_1)\Big)\, w_1(b_1) + b_1 \overline{r}(e_1) - e_1^2 \\
&\qquad + P(e_1)\left(2b_1 r_1^* - b_2 r_2^* + 2k\,\overline{R}(e_1) + (k+b_2)\overline{r}(e_2) - e_2^2\right). \qquad (3.12)
\end{aligned}$$

Now we must choose b_1, b_2 so that the effort profile (e_1^*, e_2^*) is still optimal for the professor under contract $\widetilde{\mathbf{C}}$. That is, we must ensure that

$$\partial_{e_2}\, \widetilde{\mathrm{NB}}_{2,3}(b_1, b_2; e_1^*, e_2^*) \ = \ 0 \quad \text{and} \quad \partial_{e_1}\, \widetilde{\mathrm{NB}}_{1,2,3}(b_1, b_2; e_1^*, e_2^*) \ = \ 0; \qquad (3.13)$$

Differentiating eqn.(3.11) we get $\partial_{e_2}\, \widetilde{\mathrm{NB}}_{2,3}(b_1, b_2; e_1^*, e_2^*) = (k+b_2)\overline{r}'(e_2^*) - 2e_2^*$. Thus, we have $\partial_{e_2}\, \widetilde{\mathrm{NB}}_{2,3}(b_1, b_2; e_1^*, e_2^*) = 0$ if and only if

$$b_2 = \frac{2e_2^*}{\overline{r}'(e_2^*)} - k. \qquad (3.14)$$

Differentating eqn.(3.12), we get a (complicated) expression for $\partial_{e_1}\, \widetilde{\mathrm{NB}}_{2,3}(b_1, b_2; e_1^*, e_2^*)$. Solving for b_1 to satisfy eqn.(3.13), we get

$$b_1 \ = \ \frac{B}{\overline{r}'(e_1^*)\,(2\,p_1 + 1)}, \qquad (3.15)$$

where $B := 4\,P'(e_1^*)p_1\,k\overline{R}_1^* + 2\,P'(e_1^*)p_1\,kr_2^* - 2\,P'(e_1^*)p_1\,(e_2^*)^2 - 2\,P'(e_1^*)(e_1^*)^2 + 2\,e_1^* + 4\,e_1^* p_1 + P'(e_1^*)\,b_2\,r_2^* + 2\,P'(e_1^*)\,b_2\,r_2^*\,p_1 - 2\,P'(e_1^*)k\overline{R}(e_1^*) - 4\,P'(e_1^*)k\overline{R}(e_1^*)p_1 - P'(e_1^*)r(e_2^*)k - 2\,P'(e_1^*)r(e_2^*)kp_1 - P'(e_1^*)r(e_2^*)\,b_2 - 2\,P'(e_1^*)r(e_2^*)\,b_2\,p_1 + P'(e_1^*)(e_2^*)^2 + 2\,P'(e_1^*)(e_2^*)^2 p_1 - 2\,P(e_1^*)k\overline{R}'(e_1^*) - 4\,P(e_1^*)k\overline{R}'(e_1^*)p_1$.

Proof of contract equivalence. The expressions (3.14) and (3.15) are well-defined because $\overline{r}'(e_2^*) \neq 0$ and $\overline{r}'(e_1^*) \neq 0$ by hypothesis. If we define b_1 and b_2 as in (3.14) and (3.15), then the equations (3.13) hold, so the professor's optimal effort profile is (e_1^*, e_2^*), as desired. Thus, condition (Eq1) is satisfied. If we then substitute the value of $w_1(b_1)$ from eqn.(3.10) into expression (3.9), we will get $\widetilde{\mathrm{NB}}_{1,2,3} = 0 = \mathrm{NB}_{1,2,3}$; thus, condition (Eq3) is satisfied. Condition (Eq2) is satisfied automatically because we have assumed that both $\mathbf{C}$ and $\widetilde{\mathbf{C}}$ have the same value for s_1, and set $s_2 = 0$.

Proof that $\widetilde{\mathbf{C}}$ is admissible. $\widetilde{\mathbf{C}}$ satisfies (12) because $\mathbf{C}$ does, by condition (Eq3). Now, $\mathbf{C}$ also satisfies the 'no quitting' constraints (6) and (9), so $\mathrm{NB}_{2,3} \geq 0$ and $\mathrm{NB}_3 \geq 0$; thus, it suffices to show that $\widetilde{\mathrm{NB}}_{2,3} \geq \mathrm{NB}_{2,3}$ and $\widetilde{\mathrm{NB}}_3 \geq \mathrm{NB}_3$. To do this, first note that (5) implies

$$\widetilde{\mathrm{NB}}_3 - \mathrm{NB}_3 \quad = \quad \widetilde{S}_3 - \overline{S}_3. \qquad (3.16)$$

Also, $\widetilde{\mathbf{C}}$ and $\mathbf{C}$ induce the same effort profile (e_1^*, e_2^*); thus, the professor experiences the same disutility of effort $(e_2^*)^2$ in period 2 of both contracts; thus, equation (8) implies that $\widetilde{\mathrm{NB}}_2 - \mathrm{NB}_2 = \widetilde{S}_2 - \overline{S}_2$. Furthermore, $p_2 = 1$ in both contracts; thus, equation (7) implies that

$$\widetilde{\mathrm{NB}}_{2,3} - \mathrm{NB}_{2,3} \quad = \quad (\widetilde{\mathrm{NB}}_2 - \mathrm{NB}_2) + (\widetilde{\mathrm{NB}}_3 - \mathrm{NB}_3) \underset{(3.16)}{=\!=} (\widetilde{S}_2 - \overline{S}_2) + (\widetilde{S}_3 - \overline{S}_3). \quad (3.17)$$

Lemma 2 says $\overline{\Pi} = \widetilde{\Pi}$. But $\widetilde{\mathbf{C}}$ is raid-proof, while $\mathbf{C}$ was not. This means we must have $\widetilde{\Pi}_1 \geq \overline{\Pi}_1$, while $\widetilde{\Pi}_2 \leq \overline{\Pi}_2$ and $\widetilde{\Pi}_3 \leq \overline{\Pi}_3$. Since both contracts yield the same expected revenue (2.2) in each period, this can only mean that $\widetilde{S}_2 \geq \overline{S}_2$ and $\widetilde{S}_3 \geq \overline{S}_3$. Substituting this into equations (3.16) and (3.17) yields $\widetilde{\mathrm{NB}}_3 - \mathrm{NB}_3 \geq 0$ and $\widetilde{\mathrm{NB}}_{2,3} - \mathrm{NB}_{2,3} \geq 0$; hence $\widetilde{\mathbf{C}}$ satisfies (6) and (9). $\square$

To prove Proposition 4, we need the following lemma.

Lemma A *Suppose contract* $\mathbf{C}$ *satisfies minimal recruitment condition* (12) *and constraint* (MNQ)*, and suppose* $\{\rho_e\}_{e\in\mathbb{R}_{\neq}}$ *satisfies* (B)*. Define* $\epsilon := \beta^{-1} : (0,\infty)\longrightarrow(0,\infty)$*.*

(a) *The professor's optimal effort profile is given by* $e_1^* = \epsilon(b_1)$ *and* $e_2^* = \epsilon(b_2)$*.*

(b) *Let* $\omega(b) := \epsilon(b)^2 - b\,\overline{r}[\epsilon(b)]$*. Then* $\mathbf{C}$ *must have* $b_{12} = b_{13} = b_{23} = w_3 = 0$*,* $w_2 = \omega(b_2)$*, and* $w_1 = \omega(b_1)$*.*

Proof: Hypothesis (B) implies β is invertible. Examining eqn.(5) reveals that, to make $\mathrm{NB}_3 = 0$ for all $r_1, r_2 \geq 0$, we must set $b_{13} := b_{23} := w_3 := 0$. We then have

$$\mathrm{NB}_{2,3}(r_1, e_2) \underset{(7)}{=} \mathrm{NB}_2(r_1, e_2) \underset{(8)}{=} w_2 + b_2\overline{r}(e_2) + b_{21}r_1 - e_2^2.$$

Thus, the optimal effort e_2^* is the solution to the equation $b_2\overline{r}'(e_2) \;=\; 2e_2$. It is easy to check that $e_2^* := \epsilon(b_2)$ is the unique solution to this equation. To ensure that $\mathrm{NB}_2 = 0$ for all $r_1 \geq 0$, we must then set $b_{21} := 0$ and set $w_2 = \omega(b_2)$. We then have

$$\mathrm{NB}_{1,2,3}(e_1, e_2^*) \underset{(10)}{=\!=} \mathrm{NB}_1(e_1) \underset{(11)}{=\!=} w_1 + b_1\overline{r}(e_1) - e_1^2.$$

Thus, e_1^* is the solution to the equation $b_1\overline{r}'(e_1) \;=\; 2e_1$; again, the unique solution is $e_1^* := \epsilon(b_1)$. If we finally set $w_1 = \omega(b_1)$, then we satisfy (12). $\square$

Proof of Proposition 4. (a) Suppose $\mathbf{C}$ has optimal effort profile (e_1^*, e_2^*) and standards (s_1, s_2). Let $\widetilde{\mathbf{C}}$ have the same standards (s_1, s_2) (so that (Eq2) is satisfied), and set $b_1 := \beta(e_1^*)$, $b_2 := \beta(e_2^*)$, $b_{12} = b_{13} = b_{23} = w_3 = 0$, $w_2 = \omega(b_2)$, and $w_1 = \omega(b_1)$. Lemma A says that $\widetilde{\mathbf{C}}$ is a MNQ contract which also has optimal effort profile (e_1^*, e_2^*). Thus, (Eq1) is satisfied. Lemma A also says that $\widetilde{\mathbf{C}}$ satisfies (12); thus (Eq3) is satisfied.

(b) If $\mathbf{C}$ is the globally $\overline{\Pi}$-maximizing contract, then part (a) yields an MNQ contract $\widetilde{\mathbf{C}}$ which is equivalent to $\mathbf{C}$, hence yields the same value of $\overline{\Pi}$ (by Lemma 2), hence is also $\overline{\Pi}$-maximixing. If $\mathbf{C}$ satisfies (12), then so does $\widetilde{\mathbf{C}}$, by (Eq3). Finally, any MNQ contract automatically satisfies (6) and (9); thus, $\widetilde{\mathbf{C}}$ is admissible. $\square$

For any $e \geq 0$, define $P_e(s) := P(e,s)$. Then $P_e : [0,\infty) \longrightarrow (0,1]$ is a strictly decreasing bijection; hence invertible. Define $\varsigma(e,p) := P_e^{-1}(p)$. It is easy to prove the next result.

Lemma B *For any $e_1, e_2 \geq 0$ and $p_1, p_2 \in [0,1]$, we can achieve the effort profile (e_1, e_2) and retention probabilities (p_1, p_2) with the MNQ contract $(b_1, b_2; s_1, s_2)$ defined by $b_k = \beta(e_k)$ and $s_k = \varsigma(e_k, p_k)$.* □

If $b_{12} = b_{13} = b_{23} = w_3 = 0$, with $w_2 = \omega(b_2)$, and $w_1 = \omega(b_1)$ as specified in Lemma A, and b_1, b_2, s_1 and s_2 are as specified in Lemma B, then equations (18) become:

$$\begin{array}{rrcl} & \overline{\Pi}_1(e_1, e_2; p_1, p_2) & = & -e_1^2; \\ & \overline{\Pi}_2(e_1, e_2; p_1, p_2) & = & k\,\widetilde{R}(e_1, p_1) - e_2^2; \\ \text{and} & \overline{\Pi}_3(e_1, e_2; p_1, p_2) & = & k\,\widetilde{R}(e_1, p_1) + k\,\widetilde{R}(e_2, p_2), \end{array} \tag{B.1}$$

where $\widetilde{R}(e,p) := \overline{R}[e, \varsigma(e,p)]$. Substituting (B.1) and (17) into (19), the expected profit for the University is given by

$$\overline{\Pi}(e_1, e_2; p_1, p_2) \;=\; \frac{-e_1^2 + p_1\left(-e_2^2 + k\,\widetilde{R}(e_1, p_1)\right) + p_1 p_2 k\,\left(\widetilde{R}(e_1, p_1) + \widetilde{R}(e_2, , p_2)\right)}{1 + p_1 + p_1 p_2}. \tag{B.2}$$

Lemma C *Assume hypothesis* (B). *For any $p \in [0,1]$ and $e > 0$, define $\gamma_p(e) := e/\partial_1\,\widetilde{R}(e,p)$. Suppose that, for all $p \in [0,1]$, the function $\gamma_p : \mathbb{R}_{\not{}} \longrightarrow \mathbb{R}_{\not{}}$ is bijective.*

(a) *For any fixed (p_1, p_2), the values of (e_1, e_2) which maximize the value of $\overline{\Pi}(e_1, e_2; p_1, p_2)$ are given by*

$$e_1^*(p_1, p_2) \;:=\; \gamma_{p_1}^{-1}\left(\frac{k\,p_1(1+p_2)}{2}\right) \qquad \textit{and} \qquad e_2^*(p_2) \;:=\; \gamma_{p_2}^{-1}\left(\frac{k\,p_2}{2}\right). \tag{C.1}$$

(b) *In particular, suppose $\widetilde{R}(e,p) = e\,L(p)$ for some function $L : [0,1] \longrightarrow \mathbb{R}_+$. Then $e_1^*(p_1, p_2) \;=\; L(p_1)\,k\,p_1(1+p_2)/2$ and $e_2^*(p_2) \;=\; L(p_2)\,k\,p_2/2$.*

Proof: (a) Differentiating (B.2) we get

$$\begin{array}{rrcl} & \partial_{e_1}\,\overline{\Pi}(e_1, e_2; p_1, p_2) & = & \dfrac{-2e_1 + k\,(p_1 + p_1 p_2)\,\partial_1\,\widetilde{R}(e_1, p_1)}{1 + p_1 + p_1 p_2} \\ \text{and} & \partial_{e_2}\,\overline{\Pi}(e_1, e_2; p_1, p_2) & = & \dfrac{-2p_1 e_2 + k\,p_1 p_2\,\partial_1\,\widetilde{R}(e_2, p_2)}{1 + p_1 + p_1 p_2}. \end{array}$$

To make the numerators of these expressions zero, we need

$$\frac{e_1}{\partial_1\,\widetilde{R}(e_1, p_1)} \;=\; \frac{k\,(p_1 + p_1 p_2)}{2} \qquad \text{and} \qquad \frac{e_2}{\partial_1\,\widetilde{R}(e_2, p_2)} \;=\; \frac{k\,p_2}{2},$$

which is achieved by eqn.(C.1).

(b) If $\widetilde{R}(e,p) = e\,L(p)$, then $\partial_1\,\widetilde{R}(e,p) = L(p)$, so $\gamma_p(e) = e/L(p)$, so $\gamma_p^{-1}(x) = L(p)\,x$. Now apply part (a). □

If the hypotheses of Lemma C are satisfied, then the $\overline{\Pi}$-maximization problem is reduced to finding the $(p_1, p_2) \in [0,1]^2$ which maximize the function

$$\hat{\Pi}(p_1, p_2) \quad := \quad \overline{\Pi}[e_1^*(p_1, p_2), e_2^*(p_1, p_2); p_1, p_2]. \tag{C.2}$$

The family of distributions $\{\rho_e\}_{e \in \mathbb{R}_{\neq}}$ is ***tenable*** if it satisfies two conditions:

(T1) $\overline{R}(e, s) = c_1 e + c_2 s$ for some constants $c_1, c_2 \in \mathbb{R}_+$.

(T2) $\varsigma(e, p) = e\, S(p)$ for some function $S : [0,1] \longrightarrow \mathbb{R}_+$, with $S(1) = 0$.

'Tenability' is a technical condition. However, we will later see that all three distribution families in Theorem 1 are tenable.

Lemma D *Suppose $\{\rho_e\}_{e \in \mathbb{R}_{\neq}}$ is tenable. Then hypothesis* (B) *holds. Define $L(p) := c_1 + c_2\, S(p)$. Then $\widetilde{R}(e, p) = e\, L(p)$, so Lemma* C(b) *applies. Furthermore,*

$$\hat{\Pi}(p_1, p_2) \quad = \quad \frac{k^2\, p_1}{4} \left(\frac{p_1 L(p_1)^2 (1 + p_2)^2 + L(p_2)^2 p_2^2}{1 + p_1 + p_1 p_2} \right). \tag{D.1}$$

Thus,

$$\partial_2\, \hat{\Pi}(p_1, p_2) \quad = \quad \frac{k^2\, p_1\, \Xi(p_1, p_2)}{4(1 + p_1 + p_1 p_2)^2}, \qquad \textit{where} \tag{D.2}$$

$$\begin{aligned} \Xi(p_1, p_2) \quad := \quad & 2(1 + p_1 + p_1 p_2) \left(p_1\, L(p_1)^2 (1 + p_2) + L(p_2) L'(p_2) p_2^2 + L(p_2)^2 p_2 \right) \\ & - p_1 \left(p_1 L(p_1)^2 (1 + p_2)^2 + L(p_2)^2 p_2^2 \right). \end{aligned} \tag{D.3}$$

Proof: For any $e \geq 0$, we have $\overline{r}(e) = \overline{R}(e, 0) \underset{\text{(T1)}}{=\!=} c_1 e$; thus, $\overline{r}'(e) = c_1 > 0$ is constant, so $\beta(e) := 2e/\overline{r}'(e) = 2e/c_1$ satisfies condition (B). Now,

$$\begin{aligned} \widetilde{R}(e, p) \quad &= \quad \overline{R}[e, \varsigma(e, p)] \underset{\text{(T1)}}{=\!=} c_1 e + c_2 \varsigma(e, p) \underset{\text{(T2)}}{=\!=} c_1 e + c_2 e\, S(p) \\ &= \quad e(c_1 + c_2 S(p)) = e\, L(p). \end{aligned} \tag{D.4}$$

Equation (D.4) means that Lemma C(b) is applicable, so the functions $e_1^*(p_1, p_2)$ and $e_2^*(p_2)$ are well-defined. We define

$$\widehat{R}_1(p_1, p_2) \quad := \quad \widetilde{R}[e_1^*(p_1, p_2), p_1] \underset{\text{(D.4)}}{=\!=} e_1^*(p_1, p_2)\, L(p_1), \tag{D.5}$$

$$\text{and} \quad \widehat{R}_2(p_2) \quad := \quad \widetilde{R}[e_2^*(p_2), p_2] \underset{\text{(D.4)}}{=\!=} e_2^*(p_2)\, L(p_2). \tag{D.6}$$

Substitute (B.2), (D.5) and (D.6) into (C.2) to obtain

$$\begin{aligned} \hat{\Pi}(p_1, p_2) \quad &= \quad \frac{-e_1^*(p_1, p_2)^2 + p_1 \left(-e_2^*(p_2)^2 + k\, \widehat{R}_1(p_1, p_2) \right) + k\, p_1 p_2 \left(\widehat{R}_1(p_1, p_2) + \widehat{R}_2(p_2) \right)}{1 + p_1 + p_1 p_2} \\ &= \quad \frac{k p_1 (1 + p_2) \widehat{R}_1(p_1, p_2) - e_1^*(p_1, p_2)^2}{1 + p_1 + p_1 p_2} \; + \; p_1 \left(\frac{k p_2 \widehat{R}_2(p_2) - e_2^*(p_2)^2}{1 + p_1 + p_1 p_2} \right). \end{aligned} \tag{D.7}$$

Now,

$$\begin{aligned}
kp_1(1+p_2)\widehat{R}_1(p_1,p_2) - e_1^*(p_1,p_2)^2 &\underset{\text{(D.5)}}{=\!=} kp_1(1+p_2)\, e_1^*(p_1,p_2)\, L(p_1) - e_1^*(p_1,p_2)^2 \\
&\underset{(*)}{=} k^2p_1^2(1+p_2)^2 L(p_1)^2/2 - k^2p_1^2(1+p_2)^2 L(p_1)^2/4 \\
&= k^2p_1^2(1+p_2)^2 L(p_1)^2/4. \qquad \text{(D.8)} \\
\text{and} \quad kp_2\widehat{R}_2(p_2) - e_2^*(p_2)^2 &\underset{\text{(D.6)}}{=\!=} kp_2\, e_2^*(p_2)\, L(p_2) - e_2^*(p_2)^2 \\
&\underset{(*)}{=} L(p_2)^2\, k^2\, p_2^2/2 - L(p_2)^2\, k^2\, p_2^2/4 \\
&= L(p_2)^2\, k^2\, p_2^2/4. \qquad \text{(D.9)}
\end{aligned}$$

where $(*)$ is Lemma C(b). Substituting (D.8) and (D.9) into (D.7) we get

$$\hat{\Pi}(p_1,p_2) \quad = \quad \frac{k^2p_1^2(1+p_2)^2L(p_1)^2 \;+\; p_1L(p_2)^2\,k^2\,p_2^2}{4(1+p_1+p_1p_2)},$$

which we factor to obtain (D.1). Differentiating (D.1) yields (D.2). □

Lemma E *Suppose* $\{\rho_e\}_{e\in\mathbb{R}_{\neq}}$ *is tenable, and let* Ξ *be as in equation* (D.3)*.*

(a) *Suppose* $\Xi(p_1,p_2) \geq 0$ *for all* $(p_1,p_2) \in [0,1]^2$*. Then the* $\overline{\Pi}$*-maximizing MNQ contract is tenure-track* (i.e. $p_2^* = 1$).

(b) *In this case* $e_1^* = k(c_1 + c_2S(p_1^*))p_1^*$ *and* $e_2^* = kc_1/2$*. Thus, if* $S(p_1) > c_1(1-2p_1)/2c_2p_1$ *then the* $\overline{\Pi}$*-maximizing MNQ contract induces a declining effort profile* (i.e. $e_1^* > e_2^*$)*. In particular, if* $p_1^* > 1/2$*, then* $e_1^* > e_2^*$*.*

Proof: Part (a) follows immediately from eqn.(D.2). Part (b) follows by substituting $p_2^* = 1$ into Lemma C(b); note that $L(1) = c_1 + c_2S(1) = c_1$, because $S(1) = 0$. □

We are now in a position to prove the equivalent of Theorem 1 in the restricted setting onf MNQ contracts. This is the content of the next three lemmas.

Lemma F *Suppose* $\{\rho_e\}_{e\in\mathbb{R}_{\neq}}$ *is the family of uniform distributions from* Theorem 1(a)*. Then the* $\overline{\Pi}$*-maximizing MNQ contract is tenure-track, with a declining effort profile.*

Proof: For all $0 \leq s \leq e$ we have $P(e,s) \;=\; (e-s)/e$; hence $\varsigma(e,p) = e(1-p)$. Also, $\overline{R}(e,s) \;=\; (e+s)/2$. Thus, setting $c_1 = c_2 = \frac{1}{2}$ and $S(p) = 1-p$, we see that $\{\rho_e\}$ is tenable, so we can apply Lemma E. We have $L(p) = (2-p)/2$ in Lemma D. Substitute this expression for $L(p)$ into eqn.(D.3) to get $\Xi(p_1,p_2) \;=\; f(p_1,p_2)/4$, where

$$\begin{aligned}
f(p_1,p_2) \;:=\; & 16\,p_1p_2 + 8\,p_1 + 8\,p_2 - 12\,p_2^2 - 6\,p_1^3p_2 - 4\,p_1^2 - 2\,p_1^3 + 4\,p_2^3 - 8\,p_1p_2^2 \\
& + 4\,p_1^2p_2^2 - 4\,p_1^3p_2^2 + 2\,p_1^4p_2 + p_1^4p_2^2 + p_1^4 - 4\,p_1p_2^3 + 3\,p_1p_2^4.
\end{aligned}$$

Claim 1: $f(p_1,p_2) \geq 0$ *for all* $(p_1,p_2) \in [0,1]^2$.

Proof: Let

$$\begin{aligned} g(p_1,p_2) &:= 16\,p_1p_2 + 8\,p_1 + 8\,p_2 - 12\,p_2^2 - 6\,\underline{p_1^1}p_2 - 4\,\underline{p_1^1} - 2\,\underline{p_1^1} + 4\,p_2^3 - 8\,p_1p_2^2 \\ &\qquad + 4\,p_1^2p_2^2 - 4\,\underline{p_1^2}p_2^2 + 2\,p_1^4p_2 + p_1^4p_2^2 + p_1^4 - 4\,p_1\underline{p_2^2} + 3\,p_1p_2^4 \\ &= 10\,p_1p_2 + 2\,p_1 + 8\,p_2 - 12\,p_2^2 + 4\,p_2^3 - 12\,p_1p_2^2 + 2\,p_1^4p_2 + p_1^4p_2^2 + p_1^4 + 3\,p_1p_2^4. \end{aligned}$$

Claim 1.1: *$g(p_1,p_2) \leq f(p_1,p_2)$, for all $(p_1,p_2) \in [0,1]^2$.*

Proof: Suppose $0 < n < m$. If $0 \leq x \leq 1$ then $x^n \geq x^m$; hence $-x^n \leq -x^m$. We obtained $g(p_1,p_2)$ by taking the expression for $f(p_1,p_2)$ and decreasing the exponents on the underlined negative terms. Each of these terms is made smaller by this change (by previous paragraph); thus, $g(p_1,p_2) \leq f(p_1,p_2)$. $\bigtriangledown$ `Claim 1.1`

Claim 1.2: *$\partial_1\, g(p_1,p_2) \geq 0$ for all $(p_1,p_2) \in [0,1]^2$.*

Proof: $\partial_1\, g(p_1,p_2) = (8\,p_2 + 4\,p_2^2 + 4)\,p_1^3 + 10\,p_2 + 2 - 12\,p_2^2 + 3\,p_2^4$. Thus, $\partial_1\, g(p_1,p_2) < 0$ if and only if $-p_1^3 > h(p_2)$, where

$$h(p_2) \quad := \quad \frac{2 + 10\,p_2 - 12\,p_2^2 + 3\,p_2^4}{8\,p_2 + 4\,p_2^2 + 4}.$$

The denominator of $h(p_2)$ is clearly positive for $p_2 \in [0,1]$. The numerator of $h(p_2)$ is $H(p_2) := 2 + 10\,p_2 - 12\,p_2^2 + 3\,p_2^4$. It suffices to show that $H(p_2) \geq 0$ for $p_2 \in [0,1]$. But $H'(p_2) = 10 - 24\,p_2 + 12\,p_2^3$ has only one root in $[0,1]$, which corresponds to a (positive) maximum of H. Thus, H has no interior minima in $[0,1]$. Now, $H(0) = 2 > 0$ and $H(1) = 3 > 0$; thus, $H(p_2) > 0$ for all $p_2 \in [0,1]$. Thus, $h(p_2) > 0$ for all $p_2 \in [0,1]$, so it is impossible for $-p_1^3 > h(p_2)$ (because $p_1 > 0$). Thus, $\partial_1\, g(p_1,p_2) \geq 0$. $\bigtriangledown$ `Claim 1.2`

Claim 1.3: *$g(p_1,p_2) \geq 0$ for all $(p_1,p_2) \in [0,1]^2$.*

Proof: Claim 1.2 implies that $g(p_1,p_2)$ is increasing in p_1; thus, it suffices to check that $g(0,p_2) \geq 0$ for all $p_2 \in [0,1]$. But $g(0,p_2) = G(p_2) := 8\,p_2 - 12\,p_2^2 + 4\,p_2^3$. Now, $G'(p_2) = 8 - 24\,p_2 + 12\,p_2^2$ has roots $1 \pm \sqrt{3}/3$. Only one of these roots is in $[0,1]$, and it corresponds to a maximum of G. Also, $G(0) = 0 = G(1)$. Thus, $G(p_2) \geq 0$ for all $p_2 \in [0,1]$. Thus, $g(p_1,p_2) \geq 0$ for all $(p_1,p_2) \in [0,1]^2$. $\bigtriangledown$ `Claim 1.3`

Claims 1.1 and 1.3 together imply that $f(p_1,p_2) \geq 0$ for all $(p_1,p_2) \in [0,1]^2$. $\diamondsuit$ `Claim 1`

Claim 1 and Lemma E(a) imply that the $\overline{\Pi}$-maximizing contract is tenure-track. It remains to demonstrate the declining effort profile. The maximum of $\hat{\Pi}$ occurs along the boundary $p_2 = 1$. Thus, to identify p_1^*, it suffices to maximize

$$\Upsilon(p_1) \quad := \quad \frac{\hat{\Pi}(p_1,1)}{k^2} \quad \underset{\text{(D.1)}}{=\!=} \quad \frac{p\,(16\,p - 16\,p^2 + 4\,p^3 + 1)}{16(1 + 2\,p)}$$

The zeros of

$$\Upsilon'(p_1) \quad = \quad \frac{32\,p_1 + 1 + 24\,p_1^4 - 48\,p_1^3 - 16\,p_1^2}{16\,(1 + 2\,p_1)^2}$$

are the zeros of the numerator $32\,p_1+1+24\,p_1^4-48\,p_1^3-16\,p_1^2$. Only one of these zeros is in the interval $[0,1]$; it is located at $p_1^* \approx 0.8422568359$, and corresponds to a maximum of Υ. Since $p_1^* > 1/2$, Lemma E(b) implies that $e_1^* > e_2^*$. $\Box$

Lemma G *Suppose* $\{\rho_e\}_{e\in\mathbb{R}_{\not<}}$ *is the family of exponential distributions from* Theorem 1(b)*. Then the* $\overline{\Pi}$*-maximizing MNQ contract is tenure-track, with a declining effort profile.*

Proof: We have $P(e,s) = \exp(-s/e)$, so $\varsigma(e,p) = -e\,\ln(p)$. Also, $\overline{R}(e,s) = e+s$. Setting $S(p) = -\ln(p)$ and $c_1 = c_2 = 1$, we see that $\{\rho_e\}_{e\in\mathbb{R}_{\not<}}$ is tenable; thus, we can apply Lemma E. In Lemma D, we have $L(p) = (1-\ln(p))$. Substitute into (D.3) to get

$$\begin{aligned} \Xi(p_1,p_2) &= \lambda(p_1,p_2) + p_1 g(p_1,p_2), \qquad \text{(G.1)}\\ \text{where}\qquad g(p_1,p_2) &:= 2 - p_2^2 + p_1 + 2\,p_2 + 2\,p_1p_2 + p_1p_2^2, \end{aligned}$$

and

$$\begin{aligned} \lambda(p_1,p_2) \;:=\; & \left(2\,p_1 + p_1^2 + p_1^2p_2^2 + 2\,p_1p_2 + 2\,p_1^2p_2\right)\ln(p_1)^2 \;+\; \left(p_1p_2^2 + 2\,p_1p_2 + 2\,p_2\right)\ln(p_2)^2 \\ & -\;\left(4\,p_1 + 4\,p_1p_2 + 4\,p_1^2p_2 + 2\,p_1^2 + 2\,p_1^2p_2^2\right)\ln(p_1) \;-\; \left(2\,p_1p_2 + 2\,p_2\right)\ln(p_2)\,. \end{aligned}$$

Claim 1: $\Xi(p_1,p_2) > 0$ *for all* $(p_1,p_2)\in[0,1]^2$.

Proof: $\lambda(p_1,p_2) \geq 0$ for all $(p_1,p_2)\in[0,1]^2$, because $\ln(x)^2 \geq 0$ for all $x>0$, and $-\ln(x)\geq 0$ for all $x\in(0,1]$. Thus, it suffices to show $g(p_1,p_2)>0$. Let $h(p_2) := -p_2^2 + 2p_2 + 2$.

Claim 1.1: $g(p_1,p_2) > h(p_2)$ *for all* $p_1,p_2>0$.

Proof: Write $g(p_1,p_2)$ as polynomial in p_2 to get: $g(p_1,p_2) = (-1+p_1)\,p_2^2 + (2+2\,p_1)\,p_2 + 2 + p_1$. If $p_1>0$, then $-1+p_1 > -1$, $2+2p_1>2$ and $2+p_1>2$. Thus, each p_2-coefficient of $g(p_1,p_2)$ is strictly larger than the corresponding coefficient of $h(p_2)$, for any $p_1>0$. Thus, $g(p_1,p_2)>h(p_2)$ for all $p_1,p_2>0$. $\bigtriangledown$ `Claim 1.1`

Now, $h(0) = 2 > 0$, $h(1) = 3 > 0$, and h has no extremal points in $[0,1]$; thus $h(p_2)>0$ for all $p_2\in[0,1]$. Thus, Claim 2.1 implies that $g(p_1,p_2)>0$ for all $(p_1,p_2)\in[0,1]^2$. Thus, eqn.(G.1) implies that $\Xi(p_1,p_2)\geq 0$ for all $(p_1,p_2)\in[0,1]^2$, as desired. $\diamond$ `Claim 1`

Claim 2 and Lemma E(a) imply that the $\overline{\Pi}$-maximizing contract is tenure-track. It remains to demonstrate the declining effort profile. The maximum of $\hat{\Pi}$ occurs along the boundary $p_2 = 1$. Thus, to identify p_1^*, it suffices to maximize

$$\Upsilon(p_1) \;:=\; \frac{\hat{\Pi}(p_1,1)}{k^2} \;\underset{\text{(D.1)}}{=\!=}\; \frac{p_1\left(4\,p_1 - 8\,p_1\ln(p_1) + 4\,p_1\ln(p_1)^2 + 1\right)}{4(1+2\,p_1)}.$$

The zeros of

$$\Upsilon'(p_1) \;=\; \frac{-8\,p_1\ln(p_1) + 8\,p_1\ln(p_1)^2 + 8\,p_1^2\ln(p_1)^2 + 1 - 8\,p_1^2}{4\,(1+2\,p_1)^2}$$

are the zeros of the numerator $-8\,p_1 \ln(p_1) + 8\,p_1 \ln(p_1)^2 + 8\,p_1^2 \ln(p_1)^2 + 1 - 8\,p_1^2$. This is a transcendental function, and it is not possible to find closed-form expressions for its zeros. However, numerically, the numerator has only one zero, located at $p_1^* \approx 0.7121849555$; this corresponds to the unique maximum of $\Upsilon(p_1)$. Since $p_1^* > 1/2$, Lemma E(b) implies that $e_1^* > e_2^*$. □

Lemma H *For any $\alpha > 1$, let $\{\rho_e^\alpha\}_{e\in\mathbb{R}_{\not-}}$ be the family of power law distributions from* Theorem 1(c). *There exist $\underline{\alpha}, \overline{\alpha} \in (1,\infty)$ such that, if $\alpha \in (1,\underline{\alpha})$ or $\alpha \in (\overline{\alpha},\infty)$, then the $\overline{\Pi}$-maximizing MNQ contract is tenure-track, with a declining effort profile. In particular, this holds if $\alpha = 2$.*

Proof: For any $\alpha > 1$, we have $P_\alpha(e,s) = \left(\frac{e}{e+s}\right)^\alpha$; thus $\varsigma_\alpha(e,p) = e(p^{-1/\alpha} - 1)$. Also, $\overline{R}_\alpha(e,s) = \frac{e+\alpha s}{\alpha-1}$. Thus, setting $S_\alpha(p) := (p^{-1/\alpha} - 1)$, $c_1 = 1/(\alpha-1)$ and $c_2 = \alpha/(\alpha-1)$, we see that $\{\rho_e^\alpha\}_{e\in\mathbb{R}_{\not-}}$ is tenable. In Lemma D, we have

$$L_\alpha(p) \quad = \quad \frac{\alpha p^{-1/\alpha}}{(\alpha-1)} - 1 \quad \text{thus} \quad L'_\alpha(p) \quad = \quad \frac{-1}{(\alpha-1)\,p^{\frac{\alpha+1}{\alpha}}}.$$

Substituting into eqn.(D.3) we get $\Xi_\alpha(p_1,p_2) = \xi(p_1,p_2)/(\alpha-1)^2$, where $\xi(p_1,p_2) :=$ $2\,p_1 + 2\,p_2 - 4\,pa - 4\,p_2 a - 2\,p_1 p_2^{\frac{\alpha-2}{\alpha}}\alpha + p_1^2 + 4\,p_1^{\frac{2\,\alpha-1}{\alpha}} a p_2 + 2\,p_1^{\frac{2\,\alpha-1}{\alpha}} p_2^2\alpha + 4\,p_1 p_2 + p_1 p_2^{2\frac{\alpha-1}{\alpha}}\alpha^2 - 2\,p_2^{\frac{\alpha-2}{\alpha}}\alpha - 4\,p_2^{\frac{\alpha-1}{\alpha}}\alpha^2 + p_1 p_2^2 + 2\,p_1 p_2^{\frac{\alpha-2}{\alpha}}\alpha^2 + 6\,p_1 p_2^{\frac{\alpha-1}{\alpha}}\alpha + 2\,p_1^{\frac{\alpha-2}{\alpha}}\alpha^2 + 2\,p_1^{\frac{\alpha-2}{\alpha}}\alpha^2 p_2 + p_1^2 p_2^2 + 4\,p_1^{\frac{\alpha-1}{\alpha}}\alpha - 2\,p_1^{\frac{2\,\alpha-1}{\alpha}}\alpha^2 + 2\,\alpha^2 p_2 + p_1^{2\frac{\alpha-1}{\alpha}} p_2^2\alpha^2 + 4\,p_1\alpha^2 p_2 - 8\,p_1 p_2 a + 2\,p_1\alpha^2 - 2\,p_1 p_2^{\frac{\alpha-1}{\alpha}} - 4\,p_1^{\frac{\alpha-1}{\alpha}}\alpha^2 p_2 - 4\,p_1^{\frac{2\,\alpha-1}{\alpha}}\alpha^2 p_2 + p_1^{2\frac{\alpha-1}{\alpha}}\alpha^2 + 2\,p_1^{\frac{2\,\alpha-1}{\alpha}}\alpha - 2\,p_1 p_2^{\frac{2\,\alpha-1}{\alpha}} - 2\,p_2^{\frac{\alpha-1}{\alpha}} + 2\,p_2^{\frac{\alpha-2}{\alpha}}\alpha^2 + 6\,p_2^{\frac{\alpha-1}{\alpha}}\alpha + 4\,p_1 p_2^{\frac{2\,\alpha-1}{\alpha}}\alpha + 2\,p_1^2 p_2 + p_1\alpha^2 p_2^2 - 2\,pap_2^2 - 2\,p_1^{\frac{2\,\alpha-1}{\alpha}} p_2^2\alpha^2 + 2\,p_1^{2\frac{\alpha-1}{\alpha}}\alpha^2 p_2 - 2\,p_1 p_2^{\frac{2\,\alpha-1}{\alpha}}\alpha^2 - 2\,p_1 p_2^{2\frac{\alpha-1}{\alpha}}\alpha - 2\,p_1^2\alpha + 2\,p_1^2\alpha^2 p_2 - 4\,p_1^2 p_2 a + p_1^2 p_2^2\alpha^2 - 2\,p_1^2 p_2^2\alpha + p_1^2\alpha^2 - 4\,p_1 p_2^{\frac{\alpha-1}{\alpha}}\alpha^2 + 4\,p_1^{\frac{\alpha-1}{\alpha}} a p_2 - 4\,p_1^{\frac{\alpha-1}{\alpha}}\alpha^2$. Thus, it suffices to show that $\xi_\alpha(p_1,p_2) > 0$ for all $(p_1,p_2) \in [0,1]^2$. As promised in Theorem 1(b), we consider only the asymptotics as $\alpha \searrow 1$ or $\alpha \to \infty$.

Asymptotics as $\alpha \searrow 1$. We have

$$\lim_{\alpha\searrow 1} \xi_\alpha(p_1,p_2) \quad = \quad \xi_1(p_1,p_2) \quad := \quad (1-p_1) + 2\,p_2 + p_2^2 + 2\,\frac{p_2}{p_1} + \frac{2}{p_1}. \tag{H.1}$$

Now, $\xi_1(p_1,p_2)$ is positive for all $(p_1,p_2) \in [0,1]^2$, because $(1-p_1) \geq 0$ if $p_1 \leq 1$, and all the other terms in expression (H.1) are nonegative. Thus, if α is small enough, then $\xi_\alpha(p_1,p_2) > 0$ for all $(p_1,p_2) \in [0,1]^2$; hence Lemma E(a) implies that the $\overline{\Pi}$-maximizing contract is tenure-track.

Indeed, if $\alpha = 2$, we have $\xi_2(p_1,p_2) = 12 - 6\,\sqrt{p_2} + 10\,p_2 - 8\,p_1^{3/2} p_2 - 2\,p_1\,p_2^{3/2} - 8\,\sqrt{p_1}p_2 - 4\,p_1^{3/2} p_2^2 - 4\,p_1^{3/2} - 8\,\sqrt{p_1} - 6\,p_1\,\sqrt{p_2} + 12\,p_1\,p_2 + 5\,p_1\,p_2^2 + p_1^2 p_2^2 + p_1^2 + 2\,p_1^2 p_2 + 10\,p_1$. A numerical plot reveals that $6 < \xi_2(p_1,p_2) < 16$ for all $(p_1,p_2) \in [0,1]^2$. Thus, the $\overline{\Pi}$-maximizing contract is tenure-track when $\alpha = 2$.

It remains to demonstrate the declining effort profile. For all $p \in [0,1]$, we have $\lim_{\alpha \searrow 1} S_\alpha(p) - \frac{c_1(1-2p)}{2c_2 p} = \frac{1}{2p} > 0$. Thus, if α is small enough, then Lemma E(b) implies that $e_1^* > e_2^*$, as desired.

Asymptotics as $\alpha \to \infty$. A computation reveals that $\lim_{\alpha\to\infty} \xi_\alpha(p_1,p_2) = \Xi(p_1,p_2)$, where $\Xi_\alpha(p_1,p_2)$ is exactly as in eqn.(G.1) from the exponential case. Thus, Claim 2 implies that $\lim_{\alpha\to\infty} \xi_\alpha(p_1,p_2) > 0$ for all $(p_1,p_2) \in [0,1]^2$. Thus, if α is sufficiently large, then Lemma E(a) implies that the $\overline{\Pi}$-maximizing contract is tenure-track.

It remains to demonstrate the declining effort profile. Substituting $L_\alpha(p) = \frac{\alpha p^{-1/\alpha}}{(\alpha-1)} - 1$ into eqn.(D.1) and differentiating yields

$$\partial_1 \hat{\Pi}_\alpha(p_1,1) \quad = \quad k^2 \, \frac{8\alpha^2 f_\alpha(p_1) \; + \; \alpha\, g_\alpha(p_1) \; + \; h_\alpha(p_1)}{4(\alpha-1)^2(1+2\,p_1)^2}, \tag{H.2}$$

where $f_\alpha(p) := \left(p^2 + p^{2-\frac{2}{a}} - 2p^{2-\frac{1}{a}}\right) + \left(p + p^{1-\frac{2}{a}} - 2p^{1-\frac{1}{a}}\right)$, $g_\alpha(p) := -16\,p^{2\,\frac{a-1}{a}} - 8\,p^{\frac{a-2}{a}} + 32\,p^{\frac{2\,a-1}{a}} + 24\,p^{\frac{a-1}{a}} - 16\,p^2 - 16\,p$, and $h_\alpha(p) := 1 + 8\,p - 16\,p^{\frac{2\,a-1}{a}} - 8\,p^{\frac{a-1}{a}} + 8\,p^2$.

Claim 1: *If $p \in (0,1)$, then $f_\alpha(p) > 0$.*

Proof: If $p \in (0,1)$ then the function $x \mapsto p^x$ is convex. Thus, $p^x + p^y > 2p^{(x+y)/2}$. Setting $x = 2$ and $y = 2 - \frac{2}{\alpha}$, we get $p^2 + p^{2-\frac{2}{a}} > 2p^{2-\frac{1}{a}}$. Setting $x = 1$ and $y = 1 - \frac{2}{\alpha}$, we get $p^1 + p^{1-\frac{2}{a}} > 2p^{1-\frac{1}{a}}$. Thus, each of the two bracketed terms in $f_\alpha(p)$ is strictly positive; thus $f_\alpha(p) > 0$, as desired. ◇ Claim 1

In the limit as $\alpha \to \infty$, the term $(\alpha-1)^2$ in the denominator of expression (H.2) annihilates all terms in the numerator except $f_\alpha(p)$. Thus, if α is extremely large, then the sign of $\partial_1 \hat{\Pi}_\alpha(p_1,1)$ is the same as the sign of $f_\alpha(p_1)$, and $f_\alpha(p_1) > 0$ by Claim 3. Thus, $\partial_1 \hat{\Pi}_\alpha(p_1,1) > 0$ for all $p \in (0,1)$; thus, the optimal value of p_1 is $p_1^* = 1$.

This means that, if α is large enough, then $p_1^* > 1/2$; thus, Lemma E(b) implies that $e_1^* > e_2^*$, as desired. □

Proof of Theorem 1. Lemmas F, G, and H state that, under any of the hypotheses (a), (b) or (c), the $\overline{\Pi}$-maximizing element in the space of MNQ contract is tenure-track, and induces a declining profile of effort. Thus, by Proposition 4, the same statement is true for the $\overline{\Pi}$-maximizing element in the space of *all* contracts. Thus, by Proposition 3, the same statement is true for the $\overline{\Pi}$-maximizing element in the space of *raid-proof* contracts. This proves Theorem 1. □

References

Aghion, Phillipe, Dewatripont, Mathias and Stein, Jeremy C. "Academic freedom, private-sector focus, and the process of innovation." *Rand Journal of Economics* 39 (2008): 617-35.

Arrow, Kenneth J. "Economics of Welfare and the Allocation of Resources for Invention." In *The Rate and Direction of Inventive Activity*, edited by R. Nelson (1962). Princeton University Press, Princeton, NJ.

Barbezat, Debra A., and Hughes, James W. "The Effect of Job Mobility on Academic Salaries." *Contemporary Economic Policy* 19 (2001): 409-23.

Beckman, Klaus and Schneider, Andrea. "The interaction of publications and appointments: New evidence on academic economists in Germany." *Education Economics* (2012) forthcoming.

Bernhardt, Dan, and Scoones, David. "Promotion, Turnover and Preemptive Wage Offers." *American Economic Review* 83 (1993): 771-91.

Bok, D. "Toward Education of Quality." *Harvard Magazine* (May/June 1986): 49-64.

Carmichael, H. Lorne. "Incentives in Academics: Why Is There Tenure?" *Journal of Political Economy* 96(3) (June 1988): 453-72.

Cater, Bruce.; Lew, Byron, and Pivato, Marcus. "The Robust Efficiency of Tenure Contracts in Academic Employment." *unpublished manuscript, Trent University.*

Cater, Bruce.; Lew, Byron, and Smith, Barry. "A Theory of Tenure-Track Contracts." *Education Economics* 16(2) (June 2008): 203-18.

Currie, Jan., and Newson, Janice. (Eds.) *Universities and Globalization: Critical Perspectives*, Thousand Oaks, CA: Sage.

Diamond, Arthur M. "The Life Cycle Research Productivity of Mathematicians and Scientists." *Journal of Gerentology* 41 (1986): 520-25.

Freeman, Smith. "Wage Trends as Performance Displays Productive Potential: A Model and Application to Academic Early Retirement." *Bell Journal of Economics* 8(2) (Autumn 1977): 419-43.

Goodwin, Thomas, and Sauer, Raymond. "Life cycle productivity in academic research: Evidence from cumulative publication histories of academic economists." *Southern Economic Journal* 61 (1995): 728-43.

Gumport, Patricia J. "The contested terrain of academic program reduction." *Journal of Higher Education* 64 (1993): 283311.

Gumport, Patricia J. "Academic restructuring: Organizational change and institu- tional imperatives." *Higher Education* 39 (2000): 6791.

Hearn, James C. "The Teaching Role of Contemporary American Higher Education: Popular Imagery and Organizational Reality." In *The Economics of American Higher Education*, edited by W.E. Becker and D.R. Lewis (1992). Kluwer Academic, Boston.

Hendrickson, Robert M. "Removing Tenured Faculty for Cause." *Education Law Reporter* 44 (1988): 483-94.

Hogan, Chad. "Enforcement of Implicit Employment Contracts Through Unionization." *Journal of Labor Economics* 18(1) (2002): 171-95.

Holmstrom, Bengt, and Milgrom, Paul. "Aggregation and Linearity in the Provision of Intertemporal Incentives." *Econometrica* 55(2) (1987): 308-28.

Hutchinson, E. Bruce, and Zivney, Terry. "The publication profile of economists." *Journal of Economic Education* 26 (1995): 59-79.

James, Estelle. "Decision Processes and Priorities in Higher Education." In *The Economics of American Universities*, edited by S.A. Hoenack and E.L. Collins (1990). State University of New York Press, Albany.

Kahn, Charles., and Huberman, Gur. "Two-Sided Uncertainty and 'Up-or-Out' Contracts." *Journal of Labor Economics* 6 (October 1988): 423-44.

Kenny, Lawrence W. and Studley, Roger E. "Economists Salaries and Lifetime Productivity." *Southern Economic Journal* 62 (1995): 382-93.

Lazear, Edward P. "Why is there Mandatory Retirement?" *Journal of Political Economy* 87 (1979): 1261-84.

Lazear, Edward P. "Raids and Offer Matching." In *Research in Labor Economics* 8 Part A, edited by Ronald G. Ehrenberg (1986). JAI Press, Greenwich, CT.

Levin, Sharon G., and Stephan, Paula E. "Research Productivity over the Life Cycle: Evidence for Academic Scientists." *American Economic Review* 81(1) (1991): 114-32.

Lovain, Timothy B. "Grounds for Dismissing Tenured Post-Secondary Faculty for Cause." *Journal of College and University Law* 10 (1983/84): 419-33.

MacLeod, W. Bentley., and Malcomson, James M. "Implicit Contracts, Incentive Compatibility, and Involuntary Unemployment." *Econometrica* 57 (1989): 447-80.

McDowell, John M. "Obsolescence of Knowledge and Career Publication Profiles: Some Evidence of Differences among Fields in Costs of Interrupted Careers." *American Economic Review* 72 (1982): 752-68.

McKenzie, Richard B. "In Defense of Academic Tenure." *Journal of Institutional and Theoretical Economics* 152(2) (June 1996): 325-41.

McPherson, Michael., and Schapiro, Morten Owen. "Tenure Issues in Higher Education." *Journal of Economic Perspectives* 13(1) (Winter 1999): 85-98.

Morris, Arval A. *Dismissal of Tenured Higher Education Faculty: Legal Implications of the Elimination of Mandatory Retirement.* Topeka KS.: National Organization on Legal Problems in Education (1992).

Nelson, Richard R. "The Simple Economics of Basic Scientific Research." *Journal of Political Economy* 67 (1959): 297-306.

Nussbaum, Martha C. *Not for profit : why democracy needs the humanities.* Princeton, N.J. : Princeton University Press (2010).

Oster, Sharon M. and Hamermesh, Daniel S. "Aging and Productivity among Economists." *Review of Economics and Statistics* 80 (1998): 154-6.

Pearce, David G. and Stacchetti, Ennio. "The Interaction of Implicit and Explicit Contracts in Repeated Agency." *Games and Economic Behavior* 23 (1998): 75-96.

Polachek, Solomon W. and Siebert, W. Stanley. *The Economics of Earnings* Cambridge: Cambridge University Press (1993).

Rhoades, Gary. *Managed professionals: Unionized faculty and restructuring academic labor* Albany: State University of New York (1998).

Rothschild, Michael., and White, Lawrence J. "The Analytics of the Pricing of Higher Education and Other Services in which the Customers Are Inputs." *Journal of Political Economy* 103 (1995): 573-86.

Siow. Aloysius. "Are first impressions important in academia?" *Journal of Human Resources* 26 (1991): 236-55.

Siow. Aloysius. "Some Evidence on the Signalling Role of Research in Academia." *Economic Letters* 54 (1997): 271-6.

Siow, Aloysius. "Tenure and other Unusual Personnel Practices in Academia." *Journal of Law, Economics and Organization* 14 (1998): 152-73.

Slaughter, Sheila. "Retrenchment in the 1980s: The politics of prestige and gender." *Journal of Higher Education* 64 (1993): 250282.

Slaughter, Sheila, and Rhoades, Gary. "The emergence of a competitiveness research and development policy coalition and the commercialization of academic science and technology." *Science, Technology, and Human Values* 21 (1996): 303339.

Waldman, Michael. "Up-or-Out Contracts: A Signaling Perspective." *Journal of Labor Economics* 8 (April 1990): 230-50.