Career policy and research output

Bochove, C.A. van

Career policy and research output

Cornelis van Bochove

CWTS Working Paper Series

Paper number	CWTS-WP-2012-013
Publication date	November 1, 2012
Number of pages	20
Email address corresponding author	cbochove@cwts.leidenuniv.nl
Address CWTS	Centre for Science and Technology Studies (CWTS) Leiden University P.O. Box 905 2300 AX Leiden The Netherlands www.cwts.leidenuniv.nl

Universiteit Leiden

CWTS

Centrum voor Wetenschaps- en Technologie Studies

Career policy and research output[1]

Cornelis van Bochove[2]**, CWTS, Leiden University**

October 2012

1. Introduction

University research careers are uniquely different from those of similarly qualified persons in other sectors. In most sectors, young people with a master degree from a university are initially hired for a short probationary period of a few years, and then, if successful, obtain a permanent contract. In research, obtaining a permanent contract ('tenure') often takes fifteen years and the fraction achieving it is far smaller. The initial probationary period until the awarding of a permanent contract typically has three parts: three to five or more years of PhD work, with a relatively low income, a number of years of temporary job (post-doc) contracts without the perspective of a permanent contract, and a number of years of a temporary contract with the perspective of permanence if successful ('tenure track'). Clearly, this extreme selection period reduces the relative attractiveness of research careers and hence of education fields where research is one of the major career opportunities (e.g., many of the sciences).

Unfortunately, there is no comprehensive description in the literature of this peculiar pattern, its international variation or its historical development. Nor is it clear *why* the system is as it is. Most likely, it is due to a combination of institutional and historical factors, and a conviction that it optimizes total research output. Some arguments in favor of the latter focus on the long learning process in research, but the most serious considerations refer to the high variance of research aptitude between individuals, the consequent need for sharp selection, and the long time this requires. This reasoning, however, has never been put on a rigorous base with an explicit formulation of key assumptions and a consistent analysis of their implications. Neither is there a satisfactory review of international salary patterns and personalized funding instruments that mitigate the rigidities of the standard career system.

To provide such review information is part of the research program at CWTS, but beyond the scope of the present paper. Instead, this paper gives a rigorous analysis of the consequences of the pattern mentioned above by means of a simulation model and considers the consequences of variations in crucial parameters and policy instruments on national scientific production. In

[1] Originally, this was a working paper written in early 2009. The findings about the importance of the career system and its technical parameters were the reason to start our research program into the relevant aspects and parameters of the science career system. Once this program has clarified the empirical structure of the international science career system and its historical development, as well as the values of some crucial parameters of identified in this discussion paper, we will refine the analysis given here and rewrite it as a journal article. The purpose of the present discussion paper is to make the analysis available for discussion with specialists in the field of human resources management in science.

[2] I am grateful to Cathelijn Waaijer for many discussions that contributed substantially to my thinking on the issues addressed in this paper and for her critical reading of the draft. Naturally, the responsibility for errors is mine.

particular, with respect to the instruments, we focus on the rate of selection for tenure and the time to tenure. With respect to the parameters, we focus on the distribution of talent and the relation between talent and actual output. In addition, we have to make some assumptions on the relation between age and productivity.

2. The pipeline

We distinguish two basic periods in a research career: before and after tenure. Let the number of years a person works in research be τ, with $\tau = 0$ at the start of research work, $\tau = \tau_f$ at tenure (*f* for faculty) and $\tau = \tau_r$ at retirement. Denote by $m(t, t_0)$ the size at time *t* of the cohort that started at time t_0, let all persons start research at the same age and let retirement age be constant. Total staff at *t* is then given by $\int_{t-\tau_r}^{t} m(t, t_0)dt_0$. We analyze the dynamic equilibrium where the research budget, initial cohort size, selection and wage system are constant over time. Thus the situation of any cohort at time *t* depends on the value of $\tau = t - t_0$ only, not on that of t_0. This implies that we can simply write $m(\tau)$ for the size of any cohort of age τ, and that all properties of the cohort depend on τ alone.

In the initial stage of the research career the size of a cohort is affected by two factors: attrition and selection. Attrition is the autonomous dropping out of researchers; selection is the process where researchers with below average performance are weeded out. If the rates of attrition and selection are constant in the period till tenure, we have:

$$m(\tau) = m_0 e^{(\delta+\sigma)\tau} \qquad -1 < \delta, \sigma \leq 0 \qquad \tau \leq \tau_f \tag{1.1}$$

Where δ is the rate of attrition[1] and σ the rate of selection. At τ_f a fraction φ of the remaining cohort is selected for tenure, implying:

$$m(\tau) = \varphi m_0 e^{(\delta+\sigma)\tau_f} \qquad \tau_f < \tau \leq \tau_r. \tag{1.2}$$

After tenure there is no further selection or attrition until retirement.

3. Wage regimes

Research wage regimes usually are a mixture of two systems: age[2] based or tenure based. In both regimes the initial wage, w_0, can either be on a par with or below the initial wage in non-research jobs at the same level of educational attainment. In age based systems the wage increases fairly rapidly, say at a rate λ_1, for a number of years (τ_1) and more slowly thereafter, say at a rate λ_2. In tenure based systems, there is a sharp break at the moment of obtaining tenure. In its extreme form, used in the past in parts of continental Europe, a full professor got a lifelong employment at a fixed wage (w_f), independent of his age or experience, and no further increases until retirement, except for indexation for rises in the general level of wages. Before tenure, the wage increased with experience to a maximum considerably below the tenure wage. Of course, there are also less extreme forms, where wages of tenured staff do continue to increase, though at a fairly slow rate.

In order to analyze the relation between the choice of the tenure system and the salary system, we will separately consider these two wage regimes. For the age based system, we have:

$$w(\tau) = w_0 e^{\lambda_1 \tau} \qquad \lambda_1 > 0 \qquad \tau \leq \tau_1 \tag{2.1}$$

$$w(\tau) = w_0 e^{\lambda_1 \tau_1 + \lambda_2(\tau - \tau_1)} \qquad 0 < \lambda_2 < \lambda_1 \qquad \tau > \tau_1 \tag{2.2}$$

For the tenure based system:

$$w(\tau) = w_0 e^{\lambda_1 \tau} \qquad \tau \leq \tau_1 \tag{3.1}$$

$$w(\tau) = w_f e^{\lambda_2 \tau} \qquad w_f \geq w_0 e^{\lambda_1 \tau_f} \qquad \tau > \tau_2 \tag{3.2}$$

4. Cohort size

We can now derive the initial cohort size m_0. Let b denote the science budget; for the age based wage system it is identically equal to:

$$b = \int_0^{\min(\tau_f,\tau_1)} m(\tau)w(\tau)d\tau + \int_{\min(\tau_f,\tau_1)}^{\max(\tau_f,\tau_1)} m(\tau)w(\tau)d\tau + \int_{\max(\tau_f,\tau_1)}^{\tau_r} m(\tau)w(\tau)d\tau \tag{4.1}$$

And for the tenure based system:

$$b = \int_0^{\tau_f} m(\tau)w(\tau)d\tau + \int_{\tau_f}^{\tau_r} m(\tau)w(\tau)d\tau \tag{4.2}$$

Substituting (1.1), (1.2) and (3.1), (3.2) for $m(\tau)$ and $w(\tau)$, respectively, choosing the unit of measurement such that $b = 1$ and integrating we obtain for the initial cohort size, or the inflow into the research system:

$$m_0 = \frac{1}{w_0 \psi} \tag{5}$$

Here ψ is given in table 1.

Table 1. Values of ψ for age based and tenure based wage systems

Age based	
$\psi = \frac{e^{(\delta+\sigma+\lambda_1)\tau_f} - 1}{\delta+\sigma+\lambda_1} + \varphi e^{(\delta+\sigma)\tau_f}\left(\frac{e^{\lambda_1\tau_1} - e^{\lambda_1\tau_f}}{\lambda_1} + \frac{e^{(\lambda_1-\lambda_2)\tau_1+\lambda_2\tau_r} - e^{\lambda_1\tau_1}}{\lambda_2}\right)$	$\tau_f \leq \tau_1$
$\psi = \frac{e^{(\delta+\sigma+\lambda_1)\tau_1} - 1}{\delta+\sigma+\lambda_1} + \frac{e^{(\delta+\sigma+\lambda_2)\tau_f - \lambda_2\tau_1} - e^{(\delta+\sigma)\tau_1}}{\delta+\sigma+\lambda_2} e^{\lambda_1\tau_1} + \frac{e^{\lambda_2\tau_r} - e^{\lambda_2\tau_f}}{\lambda_2} \varphi e^{(\lambda_1-\lambda_2)\tau_1+(\delta+\sigma)\tau_f}$	$\tau_f \geq \tau_1$
Tenure based	
$\psi = \frac{w_0(e^{(\delta+\sigma+\lambda_1)\tau_2} - 1)}{\delta+\sigma+\lambda_1} + \frac{\varphi w_f e^{(\delta+\sigma-\lambda_2)\tau_2}(e^{\lambda_2\tau_r} - 1)}{\lambda_2}$	

5. Productivity: learning and selection

Individual research productivity depends on research competence *c* (acquired skills), research aptitude *a* (talent) and random influences. Competence grows with experience, aptitude is constant, and the expectation of random influences is zero. As for competence, there is evidence that it grows rapidly until the age of 40, but continues growing thereafter, though slower. For simplicity, we assume that the periods of higher and lower growth of competence coincide with those of higher and lower growth of wages in the age based wage system. Thus, if *c* is competence, we have:

$$c = c_0 e^{\gamma_1 \tau} \qquad \tau \leq \tau_1 \qquad (6.1)$$

$$c = c_0 e^{\gamma_1 \tau_1 + \gamma_2 (\tau - \tau_1)} \qquad \tau > \tau_1 \qquad (6.2)$$

where $\gamma_1 > \gamma_2$.We will refer, for example in table3 to these two expressions as *c1* and *c2*, respectively.

Employing a simple linear productivity function, and a proper unit of measurement of output, productivity per researcher, *v,* is:

$$v = c + a \qquad (7)$$

Let average aptitude at the start of a cohort be a_0. As selection proceeds, average aptitude can at most be increased to the value a_s that is the expectation of *a* within the upper fraction *s* of the aptitude distribution. The value of a_s compared to a_0 depends on the shape of the aptitude distribution. In bibliometric measurements of the output of individuals, a Pareto distribution is often found, at least for the tail of the distribution. As shown in appendix 1, this translates into a function:

$$a_s = a_0 s^{\theta} \qquad -1 \leq \theta < 0 \qquad (8)$$

In this specification, a_s remains at a_0 as long as no selection has occurred, and becomes very large if only a small upper fraction remains. The value of the parameter θ is related to the skewness of the underlying Pareto distribution. If θ is close to zero, the distribution is very skewed and concentrated at the minimum aptitude. If θ is close to or equal to minus one, the distribution is flatter, that is, aptitudes vary more among individuals.

The value of *s* follows from (1):

$$s = e^{\sigma \tau} \quad (\tau \leq \tau_f) \qquad s = \varphi e^{\sigma \tau_f} \qquad (9)$$

Thus:

$$a_s = a_0 e^{\sigma \theta \tau} \qquad (\tau \leq \tau_f) \qquad (10.1)$$

$$a_s = a_0 \varphi^{\theta} e^{\sigma \theta \tau_f} \qquad (\tau > \tau_f) \qquad (10.2)$$

The actual average aptitude in any cohort depends on the degree to which the selection process is successful in selecting the upper part of the aptitude distribution. Individual aptitudes are unknown, and only individual *output* is observable. Since the latter depends on random influences as well as on competence and aptitude, output-based selection is prone to errors. However, as long as the random influence at time t is independent of that at time t' and its expectation is zero, the cumulative error in any cohort declines with the age of the cohort at a rate that depends on the distribution function of the errors. In order to obtain an impression of the speed with which the cumulative error may fall with time, we analyzed in a separate paper[3] the selection process for the case of discrete time and uniform distributions of both aptitude and errors. The table below provides the ratio of the expected value of a in the upper half of the output distribution to a_s, for the case that the impact of random influences on output is equal in each period to that of aptitude itself (formally: they have the same absolute standard deviation).

t	$\frac{E(a)}{a_s}$	t	$\frac{E(a)}{a_s}$
1	0.89	6	0.98
2	0.94	7	0.98
3	0.96	8	0.99
4	0.97	9	0.99
5	0.98	10	0.99

Clearly, in this case output-based selection provides a good approximation of aptitude selection, even after only one or two periods (a period could be two years or so). We describe the increase of actual aptitude with selection by a simple transition equation stating that actual aptitude increases from a_0 to a_s in a period of *T* years:

$$a = a_0 \frac{T-\tau}{T} + a_s \frac{\tau}{T} \qquad (\tau \leq T) \qquad (11.1)$$

$$a = a_s \qquad (\tau \geq T) \qquad (11.2)$$

Inserting (10.1) and (10.2) yields four expressions for *a*, depending on the value of τ compared to those of τ_f and *T*, as indicated in table 2.

Table 2. Values of a

$\tau \leq \tau_f, T$	$T \leq \tau < \tau_f$	$\tau, T > \tau_f$	$T \leq \tau_f < \tau$
$a1 = a_0 \frac{T-\tau}{T} + a_0 e^{\sigma\theta\tau} \frac{\tau}{T}$	$a2 = a_0 e^{\sigma\theta\tau}$	$a3 = a_0 \frac{T-\tau_f}{T} + a_0 \varphi^{\theta} e^{\sigma\theta\tau_f} \frac{\tau_f}{T}$	$a4 = a_0 \varphi^{\theta} e^{\sigma\theta\tau_f}$

6. Productivity and total output

Combining the expressions for *c* and *a*, we have to distinguish six regimes for the value of productivity. These regimes occur as combinations of fast (F) and slow (S) granting of tenure with a high (H), medium (M) and low (L) random component in individuals' output.

Table 3 Productivity $v(\tau)$

Case	Tenure	Random comp.	τ	$v(\tau)$	τ	$v(\tau)$	τ	$v(\tau)$	τ	$v(\tau)$
$\tau_1 \le \tau_f \le T$	S	H	$\tau \le \tau_1$	c1+a1	$\tau_1 \le \tau \le \tau_f$	c2+a1	$\tau_f \le \tau$	c2+a3		
$\tau_f \le \tau_1 \le T$	F	H	$\tau \le \tau_f$	c1+a1	$\tau_f \le \tau \le \tau_1$	c1+a3	$\tau_1 \le \tau$	c2+a3		
$\tau_1 \le T \le \tau_f$	S	M	$\tau \le \tau_1$	c1+a1	$\tau_1 \le \tau \le T$	c2+a1	$T \le \tau \le \tau_f$	c2+a2	$\tau_f \le \tau$	c2+a4
$\tau_f \le T \le \tau_1$	F	M	$\tau \le \tau_f$	c1+a1	$\tau_f \le \tau \le \tau_1$	c1+a3	$\tau_1 \le \tau$	c2+a3		
$T \le \tau_1 \le \tau_f$	S	L	$\tau \le T$	c1+a1	$T \le \tau \le \tau_1$	c1+a2	$\tau_1 \le \tau \le \tau_f$	c2+a2	$\tau_f \le \tau$	c2+a4
$T \le \tau_f \le \tau_1$	F	L	$\tau \le T$	c1+a1	$T \le \tau \le \tau_f$	c1+a2	$\tau_f \le \tau \le \tau_1$	c1+a4	$\tau_1 \le \tau$	c2+a4

Total output V is the sum of total output of non-tenured staff and of tenured staff:

$$V = V_{nf} + V_f = \int_0^{\tau_f} m(\tau)\, v(\tau) d\tau + \int_{\tau_f}^{\tau_r} m(\tau) v(\tau) d\tau = m_0 \int_0^{\tau_f} e^{(\delta+\sigma)\tau}\, v(\tau) d\tau +$$
$$\int_{\tau_f}^{\tau_r} \varphi m_0 e^{(\delta+\sigma)\tau_f}\, v(\tau) d\tau.$$

The solution of these integrals is derived in appendix two. Because there are six regimes with respect to productivity and two wage systems, the result, summarized in table 4, is rather taxonomical. This precludes a fruitful algebraic analysis, so we turn to simulation to analyze the results.

Table 4 Solutions of V_{nf} and V_f

SH	$\tau_1 \le \tau_f \le T$	$V_{nf} = m_0(C1 + A1)$	$V_f = m_0(C1 + A1)$
FH FM	$\tau_f \le \tau_1 \le T$ $\tau_f \le T \le \tau_1$	$V_{nf} = m_0(C2 + A1)$	$V_f = m_0(C4 + A3)$
SM SL	$\tau_1 \le T \le \tau_f$ $T \le \tau_1 \le \tau_f$	$V_{nf} = m_0(C1 + A2)$	$V_f = m_0(C3 + A4)$
FL	$T \le \tau_f \le \tau_1$	$V_{nf} = m_0(C2 + A2)$	$V_f = m_0(C4 + A4)$
$C1 = c_0[(\delta+\sigma+\gamma_1)^{-1}(e^{(\delta+\sigma+\gamma_1)\tau_1} - 1) + (\delta+\sigma+\gamma_2)^{-1} e^{\gamma_1\tau_1}(e^{(\delta+\sigma+\gamma_2)\tau_f - \gamma_2\tau_1} - e^{(\delta+\sigma)\tau_1})]$			
$C2 = c_0(\delta+\sigma+\gamma_1)^{-1}[e^{(\delta+\sigma+\gamma_1)\tau_f} - 1]$			
$C3 = c_0\varphi e^{(\delta+\sigma)\tau_f + (\gamma_1-\gamma_2)\tau_1}(e^{\gamma_2\tau_r} - e^{\gamma_2\tau_f})\gamma_2^{-1}$			
$C4 = \varphi c_0 e^{(\delta+\sigma)\tau_f}\{(e^{\gamma_1\tau_1} - e^{\gamma_1\tau_f})\gamma_1^{-1} + e^{\gamma_1\tau_1}(e^{\gamma_2\tau_r - \gamma_2\tau_1} - 1)\gamma_2^{-1}\}$			

$A1 = a_0 e^{(\delta+\sigma)\tau_f}\left[\frac{1-\tau_f/T}{\delta+\sigma} + \frac{1}{T(\delta+\sigma)^2} + \frac{e^{\sigma\theta\tau_f}}{T(\delta+\sigma+\sigma\theta)}\left(\tau_f - \frac{1}{\delta+\sigma+\sigma\theta}\right)\right] - a_0\left[\frac{1}{\delta+\sigma} + \frac{1}{T(\delta+\sigma)^2} - \frac{1}{T(\delta+\sigma+\sigma\theta)^2}\right]$
$A2 = a_0 e^{(\delta+\sigma)T}\left[\frac{1}{T(\delta+\sigma)^2} - \frac{e^{\sigma\theta T}}{T(\delta+\sigma+\sigma\theta)^2}\right] + \frac{a_0 e^{(\delta+\sigma+\sigma\theta)\tau_f}}{\delta+\sigma+\sigma\theta} - a_0\left[\frac{1}{\delta+\sigma} + \frac{1}{T(\delta+\sigma)^2} - \frac{1}{T(\delta+\sigma+\sigma\theta)^2}\right]$
$A3 = \varphi a_0 e^{(\delta+\sigma)\tau_f}\left(T - \tau_f + \varphi^{\theta}\tau_f e^{\sigma\theta\tau_f}\right)(\tau_r - \tau_f)T^{-1}$
$A4 = a_0\varphi\varphi^{\theta}e^{(\delta+\sigma+\sigma\theta)\tau_f}(\tau_r - \tau_f)$

7. Analysis of the results by simulation

Basic scenario

Table 5 provides an overview of the variations in parameter values we will consider. The first block gives the two parameters that determine just the scale of the model, not its essential mechanisms. We already set the total budget at one. Similarly, the initial wage rate determines the scale of labor; we set it at the rate at which one thousand researchers can be employed. The sum of initial competence and aptitude is the output of a newly starting researcher and determines the unit of measurement of output. We set it at one too. We will not consider variations in entry age, retirement age and 'research maturity age', but set these at 25, 65 and 40, respectively, cf. the second block.

Table 5. Parameter values.

Scale	w_0	**0,001**	**0,002**[3]								
	$c_0 + a_0$	**1**									
Life cycle	τ_r	**40**									
	τ_1	**15**									
Learning and wages	γ_1, λ_1 (%)	0	1	2.5	**5**	10					
	γ_2, λ_2 (%)	0	1	**2.5**	5						
	w_f/w_0	1	2	**3**	5						
Aptitude: distribution and ratio to competence	θ $(\times -1)$	0	0.1	0.33	**0.5**	0.75	1				
	a_0/c_0	**1**									
Attrition and selection	δ $(\% \times -1)$	**0**	3								
	σ $(\% \times -1)$	**0**	3								
	φ (%)	100	80	60	50	40	**30**	20	15	10	5
Randomness	T	40	25	15	**10**	7,5	5	4	3	2	1

As a benchmark, the values in the first column of the next blocks describe the simplest: no learning, a constant wage throughout the career, no attrition or selection before tenure, a very compact aptitude distribution and a high degree of randomness in output. The result is shown in figure 1. Total output is plotted against time to tenure τ_f for different rates of selection. Naturally, without selection output is flat at one thousand. As soon as there is some selection, total output increases to a maximum as time to tenure increases and then starts to falls off again. Thus there is an optimal time to tenure that falls from 20 (halfway to retirement) if all researchers are granted tenure, to 7 if only five percent are tenured.

The relation between selection and output is less straightforward. As selection becomes sharper, output at first increases, but eventually falls off again, if less than about a quarter of

[3] In the age based wage system

the researchers is tenured. Thus there is global maximum output, at a rate of selection of about one quarter and a time to tenure of about twelve.

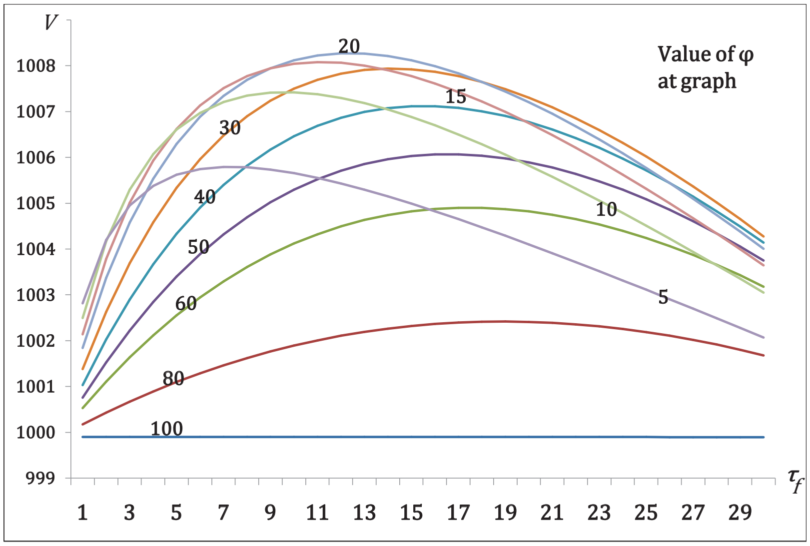

Figure 1 Basic scenario, total research output as function of time to tenure, at different rates of selection

Aptitude distribution

The variations in output are of course very small in this scenario, since the impact of selection on aptitude is very mall. This changes as soon as aptitude distribution is less compact; figure 2 displays the case of a more spread out aptitude distribution, with $\theta = -0{,}5$. This increases the impact of variations in selection and time to tenure on output almost tenfold. Moreover, the global optimum shifts to the left, and to a sharper rate of selection.

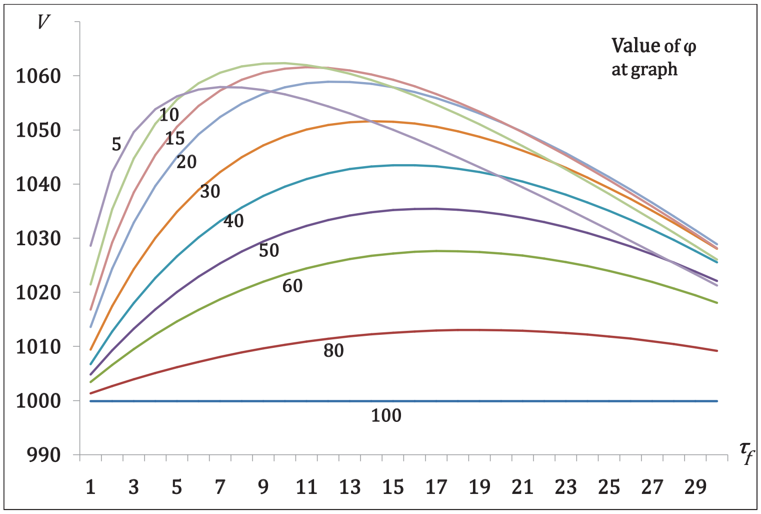

Figure 2 Basic scenario, output as function of time to tenure, at different rates of selection

Both effects increase in strength if the attitude distribution is flatter yet, with θ at its minimum of -1, cf. figure 3. Note that here the global output maximum has shifted to the sharpest rate of selection (5%), and a time to tenure of 7.

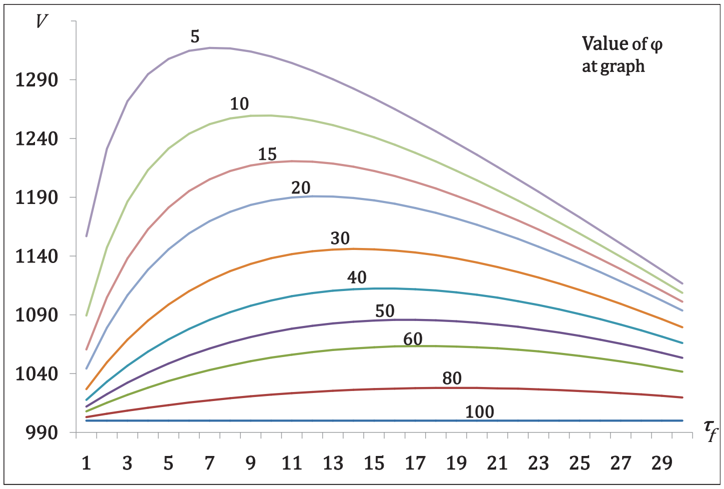

Figure 3 'Flat' aptitude distribution (θ=-1), output as function of time to tenure, at different rates of selection

Randomness of productivity

It should be borne in mind that the basic scenario as shown in figures 1-3 assumes an extremely high level of randomness in productivity (T=40). The aptitude signal in output is largely swamped out by noise. At lower noise levels output is higher and optimal time to tenure falls, cf. figure 4.

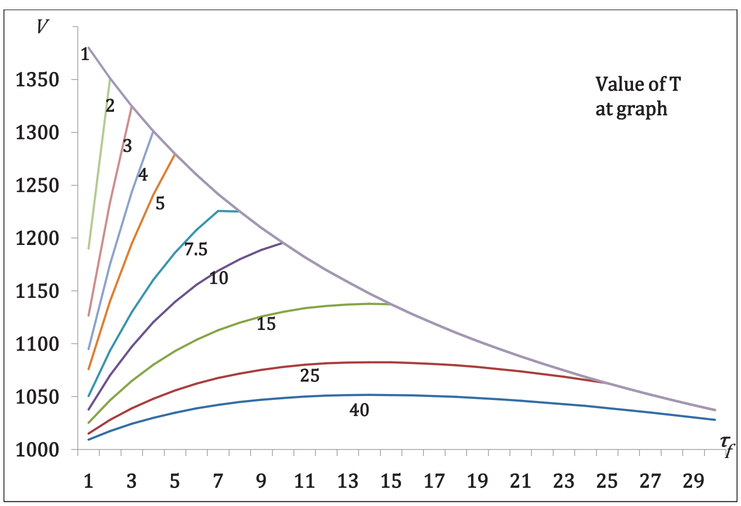

Figure 4 Variation of randomness indicator T, moderate rate of selection (φ=0,3), intermediate aptitude distribution (θ =- 0,5) .

Based on intermediate values are used for the rate of selection and the skewness of aptitude distribution, optimal time to tenure no depends almost exclusively on the noise level in the aptitude signal: as soon as the level of aptitude can be fully ascertained from actual output, that is at T, selection for tenure should take place. Only at the highest noise levels ($T \geq 15$) is the optimal time to tenure lower than T. At sharper rates of selection and flatter aptitude distribution (not shown here), the same relation holds, though at somewhat lower T and with substantially higher values of output.

Impact of tenure based wages

That early selection for tenure can boost if randomness is low and wages are constant through life is not surprising. But what if wages increase with age and at tenure? This is considered in figure 5 for the tenure based wage system. We use the 'central' parameter values (bold in table 5) for aptitude distribution, randomness; and also for wages: tenured wage is three times the initial wage; the pre-tenure rate of growth of wages is five percent, post tenure that is down to 2.5 percent. However, we still retain the basic scenario absence of learning, and of pre-tenure selection and attrition.

The pattern in figure 5 is familiar: an optimal time to tenure that is lower as selection is sharper. A difference with the previous cases is of course that the level of output is far smaller; this is not surprising, as fewer researchers can be paid from the same budget now that average lifelong wages are far higher. What is more striking is that output differences between different rates of selection and different times to tenure are far greater (up to two hundred percent) than with flat wages (no more than forty percent in the most extreme case). This does not mean, however, that the optimal time to tenure is shorter; on the contrary, it is ten years if

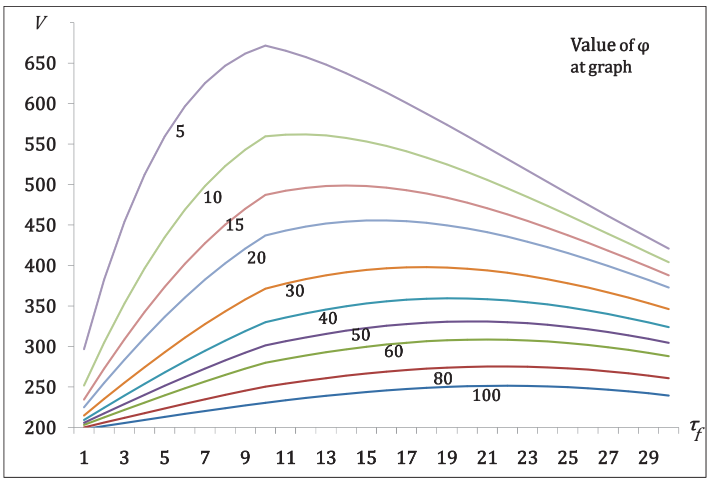

Figure 5 Central scenario, tenure based wages, no learning , no pre-tenure selection or attrition

only five percent is selected for tenure, and a whopping twenty years or more if much more than one quarter is selected. What happens is that if selection is sharp, output is maximized if tenure is granted as soon as aptitudes are completely clear ($\tau_f = T$), but if selection is not sharp, it pays to wait longer.

How sensitive are these patterns to the assumptions made on the distribution of aptitude and on the size of the random component in output? Not very, cf. figures 5a-5c. Even with the flattest distribution of aptitudes ($\theta = -1$) and a very low randomness of individuals' output

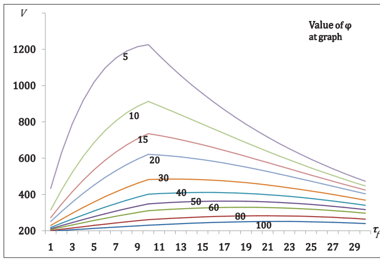

Figure 5a As 5, but θ= - 1

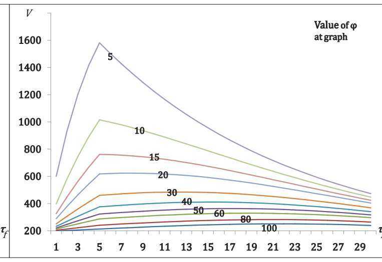

Figure 5b As 5, but θ= - 1 and T = 5

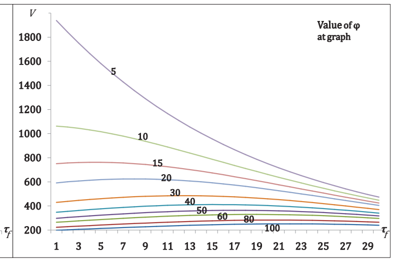

Figure 5c As 5, but θ= - 1 and T = 1

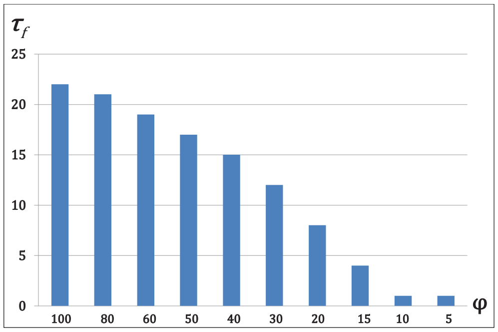

Figure 5d Optimal time to tenure in figure 5c

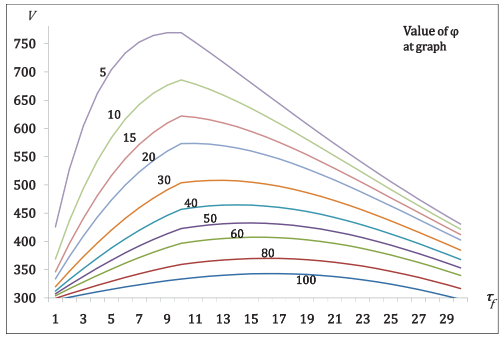

Figure 5e As 5, w_f/w_0=0

(T=1), the same pattern holds, though sharper: at the highest selectivity it is profitable to select and grant tenure after a single year, in spite of the increase in wages this entails. But at lower selectivity, the increase in productivity is rapidly offset by increased cost after tenure, and optimal time to tenure rapidly increases, cf. figure 5d.

The impact of tenure based wages on optimal time to tenure and selection depends heavily on the size of the salary hike at tenure, the ratio w_f/w_0. This is illustrated in figures 5e, where the hike is reduced from 3 to 2 and 5f where optimal times to tenure for these values are compared (the other parameters are from the central scenario). At $w_f/w_0 = 2$ the times to tenure are lower throughout than at 3.

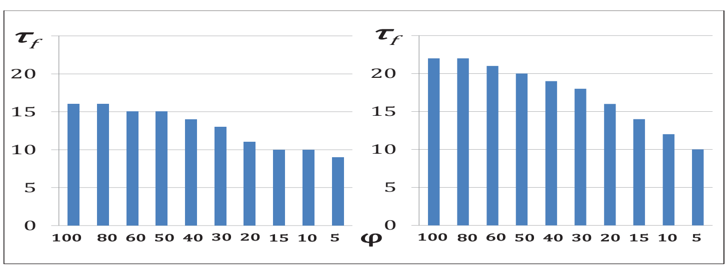

Figure 5f. As 5, optimal time to tenure, w_f/w_0 = 2 and 3

Age based wages

Something similar happens when we shift to age based wages. Since there now is no salary hike at tenure, lifelong labor cost tend to be lower. Naturally, given that the total budget is fixed, this implies higher output, which tends to obfuscate the comparison of the impact of the two systems on optimal time to tenure and rate of selection. To achieve some standardization, figure 6 employs an initial wage rate that is twice as high as that in figure 5, leading to a similar level of output, particularly in the no selection case. For easy comparison, we reproduce figure 5 next to figure 6. Clearly, in the age based system, the optimal time to tenure is much shorter than if wages depend on the tenure decision. In fact, for a very wide range of selection rates, optimal time to tenure now depends on the degree of randomness of output only: as soon as the aptitude distribution is clear, output is optimized by granting tenure, even at moderate rates of selection. This changes, however, at lower rates of randomness, cf. figure 6b, when it is optimal only the sharpest rates of selection to select and grant tenure as soon as aptitude has become clear. If the rate of selection is in the range of 15-60 percent, optimal time to tenure is in the range of 3-9 years, as long as randomness is low enough.

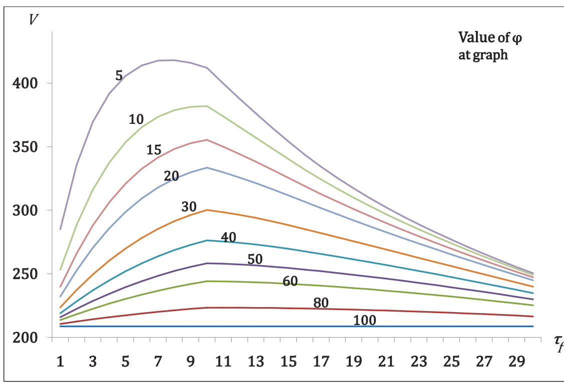

Figure 6 Central scenario, age based wages, initial wage X 2

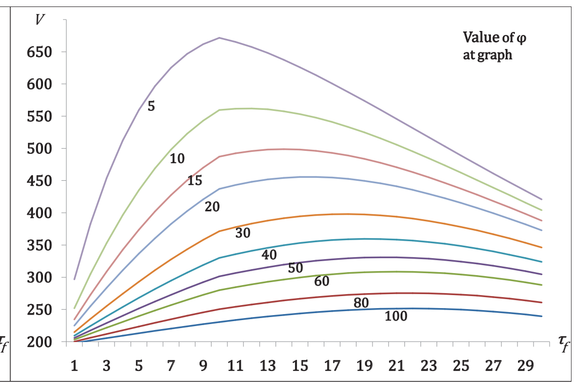

Figure 5 Central scenario, tenure based wages

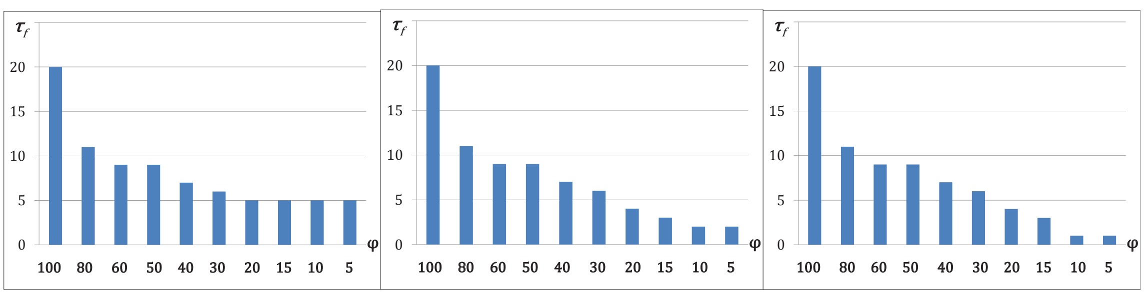

Figure 6b Optimal time to tenure, age based wages as in 6, *T*=5, 2, 1, respectively.

Impact of learning.

In figure 7 learning is introduced; for comparison, figure 6, where there is no learning, is shown. In the first 15 years of the research career competence grows at a rate of five percent, and after that

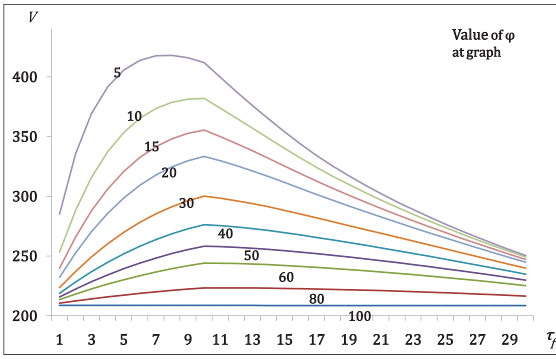

Figure 6 Central scenario, age based wages, no learning, initial wage X 2

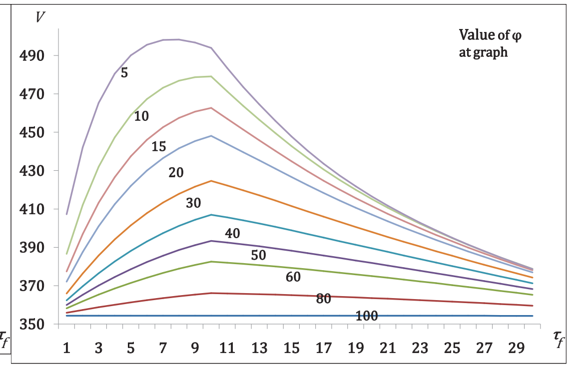

Figure 7 Central scenario, age based wages, learning, initial wage X 2

at 2.5 percent. Both rates are equal to the growth of wages in the same period. Clearly, learning does not alter the pattern of optimal tenure and selection significantly, though of course the level of output is increased substantially. This apparent irrelevance of the presence of learning for the basic policy decisions about selection and tenure vanishes if other patterns of learning prevail. Figure 8 gives three snapshots: one with a higher (ten percent) rate of learning in the initial part of the career, one without learning in the second part of the career, and one with a uniform five percent rate of learning throughout he career.

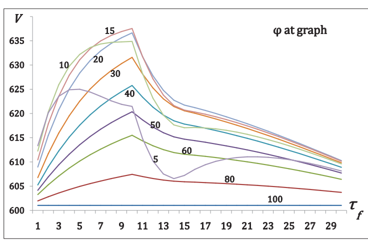

Figure 8a. Central case, age based wages, early learning 10 percent, later learning 2.5 percent

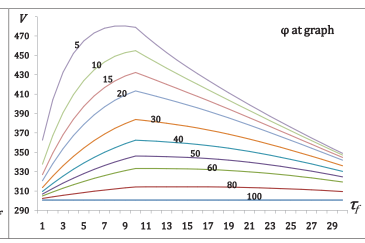

Figure 8b. Central case, age based wages, early learning 5 percent, no later learning

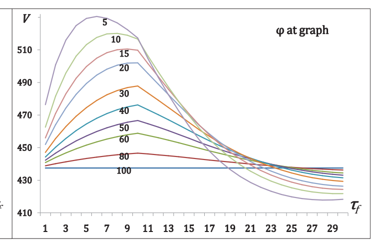

Figure 8c. Central case, age based wages, learning 5 percent throughout

The case of very rapid learning in the first part of the career has some curious features. At this level of learning, the optimal rate of selection is not, as in all cases seen before, the sharpest possible rate (5 percent), but about 15 percent. Optimal time to tenure at that rate is determined by the randomness indicator, T, which stands at ten in this scenario. At still lower selection rates, maximum output drops off; the five percent curve has a remarkable shape, with a very short optimal time to tenure, a sharp drop after ten years, and an actual minimum output at 14 years, which happens to be close to standard times to tenure in the real world. Is this pattern caused by the sharp drop in learning after the first 15 year? No, it is not, as can be seen in figure 8b, where there is no learning after fifteen years. In fact, if learning is constant throughout the career (cf. 8c) one of the curious features of 8a returns, namely the sharp drop of output at high time to tenure. So what is happening?

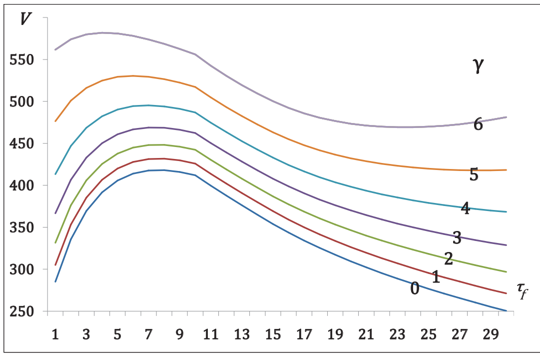

Figure 9a. Uniform rates of learning 0-6, φ=5

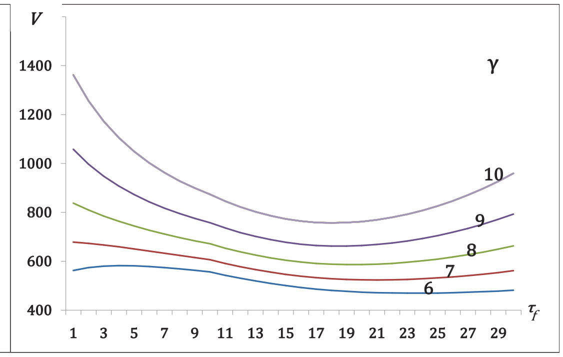

Figure 9b. Uniform rates of learning 6-10, φ=5

In figure 9 the rate of learning is the same throughout the career. Selection is sharp at 5 percent. At low rates of learning, we have the familiar pattern of an optimal time to tenure at 4-7 years, and an output that falls uniformly at higher times to tenure. However, at a rate of learning of six, this changes and output starts to rise again at high times to tenure. If the rate of learning is higher still (9b), we find a bathtub shape. The highest output is obtained if selection and tenure occur after a single year, at higher times to tenure output falls sharply until a minimum is reached at 17 years, and increases again at still higher times to tenure. In fact, if the time to tenure is extended beyond thirty years, output increases to a level beyond the original maximum. Thus, at these rates of learning, it is optimal not to select and grant tenure at all! The reason is clear: if there is a uniformly high rate of learning throughout the career, far in excess of the rate of growth wages, the ratio of output to cost is highest for the oldest researchers, and it is optimal to have as high a number of them as possible.

Attrition and pre-tenure election.

So far, we assumed that all researchers reached time to tenure, and that all selection took place at that time. In figure 10 we consider 3 percent rates of attrition and selection in the pre-tenure period. Here we have a new phenomenon: if there is no or little selection at tenure, the highest output is attained if no tenure is granted at all. This is similar to what we found at high rates of learning: due to the continuous selection, the average productivity of older researchers continues to increase, and outpaces the growth of their wage. The impact of selection at tenure only surpasses the effect of continuation of pre-tenure selection if the rate of selection at tenure is sharp enough.

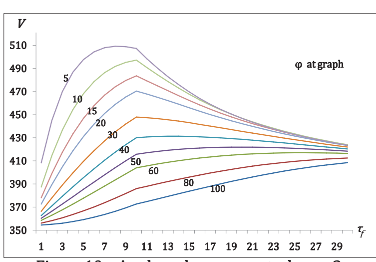

Figure 10a. Age based wages, central case, 3 percent rates of attrition and pre-tenure selection.

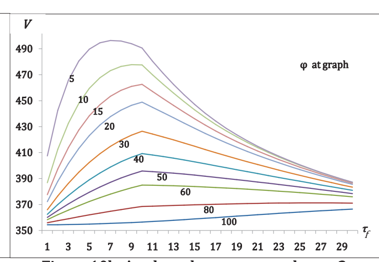

Figure 10b. Age based wages, central case, 3 percent rate of attrition, no pre-tenure selection.

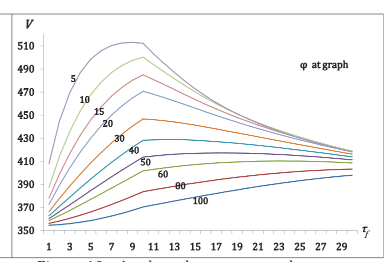

Figure 10c. Age based wages, central case, no attrition, 3 percent pre-tenure selection.

This, however is only part of the story, as is seen in 10b, where there is only attrition, not pre-tenure selection. If selection at tenure is of minor importance, it is still optimal not to grant tenure at all, but to let the attrition continue all the way to retirement.

8. Conclusion

We developed a model to analyze the relation between the design of the career system in research and the total output of research systems. We focused in particular on the appointment to permanent positions ('tenure') and the sharpness of selection at that appointment. Our results show that there may be a strong impact of the time to tenure and the degree of selection on research output. The two most essential factors determining this impact are the skewdness of the distribution of research talent ('aptitude') and the size of random influences on individuals' research output. The larger the latter, the more time is needed to determine the aptitude of individuals and the longer the optimal time to tenure is.

Output and optimal time to tenure are also influenced by the degree to which research competence is acquired by experience and the extent to which this learning process continues at a later age. Finally, there is a considerable impact on optimal time to tenure and rates of selection of the wage system. A wage system that depends not on tenure decisions but simply on seniority (age) leads to shorter optimal times to tenure and might very well be more attractive to prospective young researchers than a system where wages are dependent on tenure, in spite of the fact that the former leads to a lower average wage.

These conclusions imply that it is important to obtain empirical evidence on a number of characteristics of lifelong research output:

1. The distribution of research output of individuals as an indication of the distribution of aptitude
2. The fluctuations of an individual's research output from period to period, as an indication of the random influences on output
3. Systematic trends in individuals' research output during their career, as an indication of learning.

Reasonable assumptions on these characteristics indicate that a redesign of current career systems, with much faster, but strongly selective, tenure appointments, will not just increase the attractiveness of research as a career, but also increase research output. Therefore, research into the actual empirical value of the characteristics involved is highly relevant for the science of science policy.

Appendix 1 The mean of the upper part of Pareto distributions

Let a random variable a have a pdf $f(x)$ $(0 \leq x_m \leq x \leq x_{max})$, cdf $F(x)$ and a mean a_0 . Consider an upper fraction s of the total distribution. The mean a_s of this fraction is the mean of the conditional distribution of x, given $x > x_s$, with x_s defined such that $F(x_s) = 1 - s$. Since the conditional pdf is $sf(x)$, the conditional expectation is $a_s = \frac{1}{s}\int_{x_s}^{x_{max}} xf(x)dx$, where x_s is defined by $\int_{x_m}^{x_s} f(x)dx = 1 - s$. Let f be the Pareto distribution, with pdf and cdf, respectively: $f(x) = \frac{kx_m^k}{x^{k+1}}$ $(x_m > 0\,, k > 0, x_m \leq x)$, $\qquad F(x) = 1 - (\frac{x_m}{x})^k$. The mean is $a_0 = \frac{kx_m}{k-1}$, provided $k > 1$. For s we have $1 - s = 1 - (\frac{x_m}{x_s})^k$, implying $x_s = x_m s^{-1/k}$. Thus we obtain for the conditional mean: $a_s = \frac{1}{s}\int_{x_s}^{\infty} k(\frac{x_m}{x})^k dx = \frac{1}{s}\{\frac{kx_m^k}{(1-k)x^{k-1}}\}|_{x_s}^{\infty} = \frac{kx_m^k}{s(k-1){x_s}^{k-1}} = \frac{kx}{(k-1)s^{\frac{1}{k}}} = a_0\, s^{-1/k}$.

This result holds for $k > 1$ only. If $0 < k < 1$, no finite mean of the distribution exists. Nevertheless values of k below one are interesting. They reflect extreme inequality in the population with respect to the variable concerned. For these values a mean can once more be calculated if we assume that there is a maximum value x_{max}. In this case the integral of the original pdf over the x-domain does not have the value one: $\int_{x_m}^{x_{max}} \frac{kx_m^k}{x^{k+1}} dx = 1 - (\frac{x_m}{x_{max}})^k$. Thus the pdf and cdf of the modified distribution are $f(x) = \frac{1}{1-\rho^k}\frac{kx_m^k}{x^{k+1}}$ and $F(x) = \frac{1}{1-\rho^k}\{1 - \left(\frac{x_m}{x}\right)^k\}$ Define $\rho = \frac{x_m}{x_{max}}$. The mean now is $a_0 = \frac{1}{1-\rho^k}\int_{x_m}^{x_{max}} k(\frac{x_m}{x})^k dx = \frac{k}{1-k}\frac{\rho^k x_{max} - x_m}{1-\rho^k}$. To derive the conditional mean, first note that $1 - s = \frac{1}{1-\rho^k}\{1 - (\frac{x_m}{x_s})^k\} \longrightarrow x_s = \frac{x_m}{\rho[s\rho^{-k}+1-s]^{1/k}}$. Thus we now find: $a_s = \frac{1}{s}\frac{1}{1-\rho^k}\int_{x_s}^{x_{max}} k(\frac{x_m}{x})^k dx = \frac{1}{s}\frac{1}{1-\rho^k}\left\{\frac{kx_m^k}{(1-k)x_{max}^{k-1}} - \frac{kx_m^k}{(1-k)x_s^{k-1}}\right\}$ This can be written as

$$\frac{a_s}{a_0} = \frac{1}{s} - \frac{1}{s}\frac{1}{\rho^{k-1}-1} - \frac{1}{s}[s\rho^{-1}(\rho^{k-1}\ - 1)^{\frac{k}{k-1}} + (1-s)\rho^{\frac{1}{k}}(\rho^{k-1}\ - 1)^{\frac{k}{k-1}}]^{\frac{k-1}{k}}$$

For $x_{max} \to \infty$ we have $\rho \to 0, \rho^{k-1} \to \infty\,, (\rho^{k-1}\ - 1) \to \rho^{k-1}$,

And $\lim_{x_{max}\to\infty} \frac{a_s}{a_0} = \frac{1}{s} - \frac{1}{s}\lim_{x_{max}\to\infty}[s\rho^{-1}\left(\rho^{k-1}\right)^{\frac{k}{k-1}} + (1-s)\rho^{\frac{1}{k}}(\rho^{k-1})^{\frac{k}{k-1}}]^{\frac{k-1}{k}} =$

$$= \frac{1}{s} - \frac{1}{s}\lim_{x_{max}\to\infty}[s\rho^{k-1} + (1-s)\rho^{\frac{1}{k}}\rho^k]^{\frac{k-1}{k}} = \frac{1}{s} - \frac{1}{s}[\lim_{x_{max}\to\infty} s\rho^{k-1}]^{\frac{k-1}{k}} = \frac{1}{s}$$

In the main paper we use the equation $a_s = a_0 s^{\theta}$, where $\theta = -\frac{1}{k} < 0$. Here k can be both larger than one $(\theta > -1)$ and between zero and one, provided there exists in the latter case a maximum value for the variable concerned. For large values of this maximum, the upper mean can be approached by the above equation with $\theta = -1$. Since there is no ready interpretation of θ in terms of a parameter of a Pareto distribution for $\theta < -1$, we exclude this possibility and impose $-1 \leq \theta < 0$.

As for the interpretation of equation $a_s = a_0 s^\theta$, in Pareto distributions, the lower the value of k, the flatter the distribution is, cf figure 1 ($x_m = 1$).

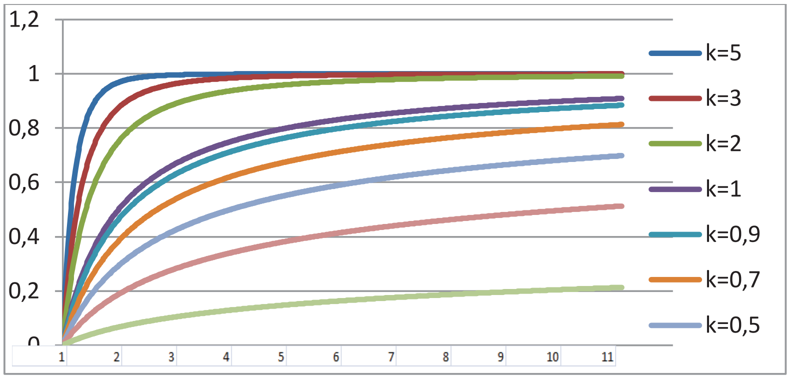

Figure 2. Cdf of Pareto distributions

Since we may write s= $(\frac{a_s}{a_0})^{-k}$, it is seen immediately that a given ratio a_s/a_0 corresponds to a higher value of s (that is, a larger fraction of the original distribution remaining) the higher k and the lower $-k$ and θ are. That is, with a flatter distribution selection need not be sharp to achieve the same gain in average aptitude.

Appendix 2. Derivation of the solutions for research output.

Using table 3, the output of non-tenured and tenured staff is given in tables A.1 and A.2. The basic formulas for the solutions of the integrals in these tables are given in table A.3.

Table A.1 Non-tenured staff output.

SH	$\tau_1 \leq \tau_f \leq T$	$\int_0^{\tau_1} \text{m1}(\text{c1} + \text{a3})d\tau + \int_{\tau_1}^{\tau_f} \text{m1}(\text{c2} + \text{a3})d\tau$
FH	$\tau_f \leq \tau_1 \leq T$	$\int_0^{\tau_f} \text{m1}(\text{c1} + \text{a1})d\tau$
SM	$\tau_1 \leq T \leq \tau_f$	$\int_0^{\tau_1} \text{m1}(\text{c1} + \text{a1})d\tau + \int_{\tau_1}^{T} \text{m1}(\text{c2} + \text{a1})d\tau + \int_T^{\tau_f} \text{m1}(\text{c2} + \text{a2})d\tau$
FM	$\tau_f \leq T \leq \tau_1$	$\int_0^{\tau_f} \text{m1}(\text{c1} + \text{a1})d\tau$
SL	$T \leq \tau_1 \leq \tau_f$	$\int_0^{T} \text{m1}(\text{c1} + \text{a1})d\tau + \int_T^{\tau_1} \text{m1}(\text{c1} + \text{a2})d\tau + \int_{\tau_1}^{\tau_f} \text{m1}(\text{c2} + \text{a2})d\tau$
FL	$T \leq \tau_f \leq \tau_1$	$\int_0^{T} \text{m1}(\text{c1} + \text{a1})d\tau + \int_T^{\tau_f} \text{m1}(\text{c1} + \text{a2})d\tau$

$\text{m1} = m_0 e^{(\delta+\sigma)\tau}$

Table A.2 Tenured staff output.

SH	$\tau_1 \leq \tau_f \leq T$	$\int_{\tau_f}^{\tau_r} \text{m2}(\text{c2} + \text{a3})d\tau$
FH	$\tau_f \leq \tau_1 \leq T$	$\int_{\tau_f}^{\tau_1} \text{m2}(\text{c1} + \text{a3})d\tau + \int_{\tau_1}^{\tau_r} \text{m2}(\text{c2} + \text{a3})d\tau$
SM	$\tau_1 \leq T \leq \tau_f$	$\int_{\tau_f}^{\tau_r} \text{m2}(\text{c2} + \text{a4})d\tau$
FM	$\tau_f \leq T \leq \tau_1$	$\int_{\tau_f}^{\tau_1} \text{m2}(\text{c1} + \text{a3})d\tau + \int_{\tau_1}^{\tau_r} \text{m2}(\text{c2} + \text{a3})d\tau$
SL	$T \leq \tau_1 \leq \tau_f$	$\int_{\tau_f}^{\tau_r} \text{m2}(\text{c2} + \text{a4})d\tau$
FL	$T \leq \tau_f \leq \tau_1$	$\int_{\tau_f}^{\tau_1} \text{m2}(\text{c1} + \text{a4})d\tau + \int_{\tau_1}^{\tau_r} \text{m2}(\text{c2} + \text{a4})d\tau$

$\text{m2} = m_0 \varphi^\theta e^{(\delta+\sigma)\tau_f}$

Table A.3 Primitive functions of the components of V_{nf} and V_f

Component	Primitive function $\int f(\tau)d\tau$
m1 c1=$e^{(\delta+\sigma)\tau} c_0 e^{\gamma_1 \tau}$	$c_0 e^{(\delta+\sigma+\gamma_1)\tau}(\delta + \sigma + \gamma_1)^{-1}$
m1 c2=$e^{(\delta+\sigma)\tau} c_0 e^{\gamma_1 \tau_1 + \gamma_2(\tau-\tau_1)}$	$c_0 e^{(\gamma_1-\gamma_2)\tau_1 + (\delta+\sigma+\gamma_2)\tau}(\delta + \sigma + \gamma_2)^{-1}$
m1 a1=$e^{(\delta+\sigma)\tau}\{a_0 \frac{T-\tau}{T} + a_0 e^{\sigma\theta\tau} \frac{\tau}{T}\}$	$a_0 e^{(\delta+\sigma)\tau}[\frac{1-\tau/T}{\delta+\sigma} + \frac{1}{T(\delta+\sigma)^2} + \frac{e^{\sigma\theta\tau}}{T(\delta+\sigma+\sigma\theta)}\left(\tau - \frac{1}{\delta+\sigma+\sigma\theta}\right)]$
m1 a2=$e^{(\delta+\sigma)\tau} a_0 e^{\sigma\theta\tau}$	$a_0 e^{(\delta+\sigma+\sigma\theta)\tau}(\delta + \sigma + \sigma\theta)^{-1}$
V_f	
m2 c1=$e^{(\delta+\sigma)\tau_f} c_0 e^{\gamma_1 \tau} \varphi$	$\gamma_1^{-1} \varphi c_0 e^{(\delta+\sigma)\tau_f + \gamma_1 \tau}$
m2 c2=$\varphi e^{(\delta+\sigma)\tau_f} c_0 e^{\gamma_1 \tau_1 + \gamma_2(\tau-\tau_1)}$	$\gamma_2^{-1} c_0 \varphi e^{(\gamma_1-\gamma_2)\tau_1 + \gamma_2 \tau + (\delta+\sigma)\tau_f}$
m2 a3=$\varphi e^{(\delta+\sigma)\tau_f}[a_0 \frac{T-\tau_f}{T} + a_0 \varphi^\theta e^{\sigma\theta\tau_f} \frac{\tau_f}{T}]$	$\varphi a_0 e^{(\delta+\sigma)\tau_f}(T - \tau_f + \varphi^\theta \tau_f e^{\sigma\theta\tau_f})\tau T^{-1}$
m2 a4 $= a_0 \varphi^\theta e^{\sigma\theta\tau_f} \varphi e^{(\delta+\sigma)\tau_f}$	$a_0 \varphi \varphi^\theta e^{(\delta+\sigma+\sigma\theta)\tau_f} \tau$
Note: $\int x\alpha e^{bx} dx = \alpha\beta^{-1} e^{\beta x}(x - \beta^{-1})$	

The results of applying these solutions to the integrals in tables A.1 and A.2 are provided in tables A.4 and A.5. Reshuffling yields the expressions given in table 4 in the main text.

Table A.4 Solution of the integrals of V_{nf} (omitting m_0)

	First integral	Second integral	Third integral
SH	$\frac{c_0}{\delta+\sigma+\gamma_1}(e^{(\delta+\sigma+\gamma_1)\tau_1}-1)+a_0e^{(\delta+\sigma)\tau_1}$ $\left[\frac{1-\tau_1/T}{\delta+\sigma}+\frac{1}{T(\delta+\sigma)^2}+\frac{e^{\sigma\theta\tau_1}}{T(\delta+\sigma+\sigma\theta)}\left(\tau_1-\frac{1}{\delta+\sigma+\sigma\theta}\right)\right]$ $-a_0\left[\frac{1}{\delta+\sigma}+\frac{1}{T(\delta+\sigma)^2}-\frac{1}{T(\delta+\sigma+\sigma\theta)^2}\right]$	$\frac{c_0}{\delta+\sigma+\gamma_2}(e^{(\gamma_1-\gamma_2)\tau_1+(\delta+\sigma+\gamma_2)\tau_f}-e^{(\delta+\sigma+\gamma_1)\tau_1})$ $+a_0e^{(\delta+\sigma)\tau_f}\left[\frac{1-\frac{\tau_f}{T}}{\delta+\sigma}+\frac{1}{T(\delta+\sigma)^2}+\frac{e^{\sigma\theta\tau_f}\left(\tau_f-\frac{1}{\delta+\sigma+\sigma\theta}\right)}{T(\delta+\sigma+\sigma\theta)}\right]$ $-a_0e^{(\delta+\sigma)\tau_1}\left[\frac{1-\frac{\tau_1}{T}}{\delta+\sigma}+\frac{1}{T(\delta+\sigma)^2}+\frac{e^{\sigma\theta\tau_1}\left(\tau_1-\frac{1}{\delta+\sigma+\sigma\theta}\right)}{T(\delta+\sigma+\sigma\theta)}\right]$	
FH	$\frac{c_0}{\delta+\sigma+\gamma_1}\left(e^{(\delta+\sigma+\gamma_1)\tau_f}-1\right)+a_0e^{(\delta+\sigma)\tau_f}$ $\left[\frac{1-\tau_f/T}{\delta+\sigma}+\frac{1}{T(\delta+\sigma)^2}+\frac{e^{\sigma\theta\tau_f}}{T(\delta+\sigma+\sigma\theta)}\left(\tau_f-\frac{1}{\delta+\sigma+\sigma\theta}\right)\right]$ $-a_0\left[\frac{1}{\delta+\sigma}+\frac{1}{T(\delta+\sigma)^2}-\frac{1}{T(\delta+\sigma+\sigma\theta)^2}\right]$		
SM	$\frac{c_0}{\delta+\sigma+\gamma_1}(e^{(\delta+\sigma+\gamma_1)\tau_1}-1)+a_0e^{(\delta+\sigma)\tau_1}$ $\left[\frac{1-\tau_1/T}{\delta+\sigma}+\frac{1}{T(\delta+\sigma)^2}+\frac{e^{\sigma\theta\tau_1}}{T(\delta+\sigma+\sigma\theta)}\left(\tau_1-\frac{1}{\delta+\sigma+\sigma\theta}\right)\right]$ $-a_0\left[\frac{1}{\delta+\sigma}+\frac{1}{T(\delta+\sigma)^2}-\frac{1}{T(\delta+\sigma+\sigma\theta)^2}\right]$	$\frac{c_0}{\delta+\sigma+\gamma_2}\left(e^{(\gamma_1-\gamma_2)\tau_1+(\delta+\sigma+\gamma_2)T}-e^{(\delta+\sigma+\gamma_1)\tau_1}\right)$ $+a_0e^{(\delta+\sigma)T}\left[\frac{1}{T(\delta+\sigma)^2}+\frac{e^{\sigma\theta T}\left(T-\frac{1}{\delta+\sigma+\sigma\theta}\right)}{T(\delta+\sigma+\sigma\theta)}\right]-$ $a_0e^{(\delta+\sigma)\tau_1}\left[\frac{1-\frac{\tau_1}{T}}{\delta+\sigma}+\frac{1}{T(\delta+\sigma)^2}+\frac{e^{\sigma\theta\tau_1}\left(\tau_1-\frac{1}{\delta+\sigma+\sigma\theta}\right)}{T(\delta+\sigma+\sigma\theta)}\right]$	$\frac{c_0e^{(\gamma_1-\gamma_2)\tau_1}}{\delta+\sigma+\gamma_2}\left(e^{(\delta+\sigma+\gamma_2)\tau_f}-e^{(\delta+\sigma+\gamma_2)T}\right)$ $+\frac{a_0(e^{(\delta+\sigma+\sigma\theta)\tau_f}-e^{(\delta+\sigma+\sigma\theta)T})}{\delta+\sigma+\sigma\theta}$
FM	$\frac{c_0}{\delta+\sigma+\gamma_1}(e^{(\delta+\sigma+\gamma_1)\tau_f}-1)+a_0e^{(\delta+\sigma)\tau_f}$ $\left[\frac{1-\tau_f/T}{\delta+\sigma}+\frac{1}{T(\delta+\sigma)^2}+\frac{e^{\sigma\theta\tau_f}}{T(\delta+\sigma+\sigma\theta)}\left(\tau_f-\frac{1}{\delta+\sigma+\sigma\theta}\right)\right]$ $-a_0\left[\frac{1}{\delta+\sigma}+\frac{1}{T(\delta+\sigma)^2}-\frac{1}{T(\delta+\sigma+\sigma\theta)^2}\right]$		
SL	$\frac{c_0}{\delta+\sigma+\gamma_1}(e^{(\delta+\sigma+\gamma_1)T}-1)+a_0e^{(\delta+\sigma)T}$ $\left[\frac{1}{T(\delta+\sigma)^2}+\frac{e^{\sigma\theta T}\left(T-\frac{1}{\delta+\sigma+\sigma\theta}\right)}{T(\delta+\sigma+\sigma\theta)}\right]-$ $a_0\left[\frac{1}{\delta+\sigma}+\frac{1}{T(\delta+\sigma)^2}-\frac{1}{T(\delta+\sigma+\sigma\theta)^2}\right]$	$\frac{c_0}{\delta+\sigma+\gamma_1}(e^{(\delta+\sigma+\gamma_1)\tau_1}-e^{(\delta+\sigma+\gamma_1)T})+$ $\frac{a_0}{\delta+\sigma+\sigma\theta}(e^{(\delta+\sigma+\sigma\theta)\tau_1}-e^{(\delta+\sigma+\sigma\theta)T})$	$\frac{c_0e^{(\gamma_1-\gamma_2)\tau_1+(\delta+\sigma+\gamma_2)\tau_f}-e^{(\delta+\sigma+\gamma_1)\tau_1}}{\delta+\sigma+\gamma_2}$ $+a_0\frac{e^{(\delta+\sigma+\sigma\theta)\tau_f}-e^{(\delta+\sigma+\sigma\theta)\tau_1}}{\delta+\sigma+\sigma\theta}$
FL	$\frac{c_0}{\delta+\sigma+\gamma_1}(e^{\gamma_1T}-1)+a_0e^{(\delta+\sigma)T}$ $\left[\frac{1}{T(\delta+\sigma)^2}+\frac{e^{\sigma\theta T}\left(T-\frac{1}{\delta+\sigma+\sigma\theta}\right)}{T(\delta+\sigma+\sigma\theta)}\right]-$ $a_0\left[\frac{1}{\delta+\sigma}+\frac{1}{T(\delta+\sigma)^2}-\frac{1}{T(\delta+\sigma+\sigma\theta)^2}\right]$	$\frac{c_0}{\delta+\sigma+\gamma_1}(e^{(\delta+\sigma+\gamma_1)\tau_f}-e^{(\delta+\sigma+\gamma_1)T})+$ $\frac{a_0}{\delta+\sigma+\sigma\theta}(e^{(\delta+\sigma+\sigma\theta)\tau_f}-e^{(\delta+\sigma+\sigma\theta)T})$	

Table A.5. Solution of the integrals of V_f (omitting m_0)

	First integral	Second integral
SH	$c_0\varphi e^{(\gamma_1-\gamma_2)\tau_1+(\delta+\sigma)\tau_f}(e^{\gamma_2\tau_r}-e^{\gamma_2\tau_f})\,\gamma_2^{-1}+$ $\varphi a_0e^{(\delta+\sigma)\tau_f}(T-\tau_f+\varphi^\theta\tau_fe^{\sigma\theta\tau_f})(\tau_r-\tau_f)T^{-1}$	
FH	$\varphi c_0e^{(\delta+\sigma)\tau_f}(e^{\gamma_1\tau_1}-e^{\gamma_1\tau_f})\gamma_1^{-1}+$ $\varphi a_0e^{(\delta+\sigma)\tau_f}(T-\tau_f+\varphi^\theta\tau_fe^{\sigma\theta\tau_f})(\tau_1-\tau_f)T^{-1}$	$c_0\varphi e^{(\gamma_1-\gamma_2)\tau_1+(\delta+\sigma)\tau_f}(e^{\gamma_2\tau_r}-e^{\gamma_2\tau_1})\gamma_2^{-1}+$ $\varphi a_0e^{(\delta+\sigma)\tau_f}(T-\tau_f+\varphi^\theta\tau_fe^{\sigma\theta\tau_f})(\tau_r-\tau_1)T^{-1}$
SM	$c_0\varphi e^{(\gamma_1-\gamma_2)\tau_1+(\delta+\sigma)\tau_f}(e^{\gamma_2\tau_r}-e^{\gamma_2\tau_f})\,\gamma_2^{-1}+$ $+a_0\varphi\varphi^\theta e^{(\delta+\sigma+\sigma\theta)\tau_f}(\tau_r-\tau_f)$	
FM	$\varphi c_0e^{(\delta+\sigma)\tau_f}(e^{\gamma_1\tau_1}-e^{\gamma_1\tau_f})\gamma_1^{-1}+$ $\varphi a_0e^{(\delta+\sigma)\tau_f}(T-\tau_f+\varphi^\theta\tau_fe^{\sigma\theta\tau_f})(\tau_r-\tau_f)T^{-1}$	$c_0\varphi e^{(\gamma_1-\gamma_2)\tau_1+(\delta+\sigma)\tau_f}(e^{\gamma_2\tau_r}-e^{\gamma_2\tau_1})\gamma_2^{-1}++$ $\varphi a_0e^{(\delta+\sigma)\tau_f}(T-\tau_f+\varphi^\theta\tau_fe^{\sigma\theta\tau_f})(\tau_r-\tau_f)T^{-1}$
SL	$c_0\varphi e^{(\gamma_1-\gamma_2)\tau_1+(\delta+\sigma)\tau_f}(e^{\gamma_2\tau_r}-e^{\gamma_2\tau_f})\,\gamma_2^{-1}+$ $+a_0\varphi\varphi^\theta e^{(\delta+\sigma+\sigma\theta)\tau_f}(\tau_r-\tau_f)$	
FL	$\varphi c_0e^{(\delta+\sigma)\tau_f}(e^{\gamma_1\tau_1}-e^{\gamma_1\tau_f})\gamma_1^{-1}+$ $+a_0\varphi\varphi^\theta e^{(\delta+\sigma+\sigma\theta)\tau_f}(\tau_1-\tau_f)$	$c_0\varphi e^{(\gamma_1-\gamma_2)\tau_1+(\delta+\sigma)\tau_f}(e^{\gamma_2\tau_r}-e^{\gamma_2\tau_1})\,\gamma_2^{-1}$ $+a_0\varphi\varphi^\theta e^{(\delta+\sigma+\sigma\theta)\tau_f}(\tau_r-\tau_1)$

The results of applying these solutions to the integrals in tables A.1 and A.2 and reshuffling yields the expressions given in table 4 in the main text.

[1] A positive value for delta is not logically impossible: it would mean that a cohort acquires additional researchers after its start; with the same characteristics (wage, average aptitude, level of competence) the other researchers in the cohort have at τ. This can be (and in fact is) achieved by international migration. Analyzing this is , however, beyond our scope

[2] Strictly: experience

[3] Cornelis van Bochove (2010) Expectation of a uniform random variable with uniform observation errors after selection of the highest observations, Statistical Papers, DOI 10.1007/s00362-009-0304-y