Grantmaking*

Jérôme Adda† Marco Ottaviani‡

October 13, 2022

Abstract

The paper develops a model of non-market allocation of resources through grantmaking. On the supply side, the available budget of grants is awarded to applicants who are evaluated most favorably according to the noisy information available to reviewers. On the demand side, stronger candidates are more likely to obtain grants and thus self-select into applying. Leveraging a technique based on the quantile function, we characterize a broad set of allocation rules under which an increase in evaluation noise in a field raises applications in that field—and reduces applications in all other fields. We illustrate the practical relevance of the model by exploiting a change in the budget allocation rule at the European Research Council, showing that a one standard deviation increase in own evaluation noise leads to a 0.3 standard deviation increase in the number of applications and budget share. We also derive some subtle implications for the design of grantmaking institutions in terms of the endogenous choice of noise by fields and the optimal pooling of fields into panels.

Keywords: Non-market resource allocation, grants, applications, grading on a curve, evaluation noise, budget apportionment across fields.

JEL: D83, H81.

*We thank without implication Philippe Aghion, Manuel Arellano, Ricardo Alonso, Christian Bjerke, Estelle Cantillon, Mikhail Drugov, Albin Erlanson, Irwin Feller, Alfonso Gambardella, Nicola Gennaioli, Ian Jewitt, Charlie Johnson, Nenad Kos, Danielle Li, Massimo Marinacci, Andreu Mas-Colell, Meg Meyer, Andrea Prat, Ron Siegel, Timothy Simcoe, William Thomson, Giovanni Ursino, Reinhilde Veugelers, Huseyin Yildirim, John Walsh, Glen Weyl, and Richard Zeckhauser for helpful discussion; James Atkins, Mohamed Badaoui, Francesco Bilotta, Aldo Lucia, Marta Mojoli, and Massimiliano Pozzi for outstanding research assistance; and Federico Pessina, Eugenio Piga, Maik Sälzer, and Biao Yang for proofreading. We gratefully acknowledge research support by the NBER-Sloan Foundation Science of Science Funding program and by the European Patent Office (EPO) Academic Research Programme.

†Bocconi University, BIDSA, CEPR, and IGIER, jerome.adda@unibocconi.it.

‡Bocconi University, BIDSA, CEPR, and IGIER, marco.ottaviani@unibocconi.it.

"The just, then, is a species of the proportionate . . . proportion is equality of ratios

$$\frac{B_i}{B_j} = \frac{a_i}{a_j} \text{ and, therefore, alternando } \frac{B_i}{a_i} = \frac{B_j}{a_j}$$

. . . all men agree that what is just in distribution must be according to merit in some sense, though they do not all specify the same sort of merit, but democrats identify it with the status of freeman, supporters of oligarchy with wealth (or with noble birth), and supporters of aristocracy with excellence."

Aristotle, *Nicomachean Ethics*, Book V, Chapter 3

1 Introduction

Over the sweep of history, artists and scientists have long relied on wealthy patrons and public support to finance their inventions and discoveries. In 1610 Galileo Galilei wrote to his former pupil Cosimo de' Medici, the Grand Duke of Tuscany, subtly asking for financial support to explore the sky with his new powerful telescope. To lure the patron, Galileo named Jupiter's moons he had just discovered the Medician stars and promised "many discoveries and such as perhaps no other prince can match." Cosimo was duly impressed and granted Galileo a full teaching buyout at the University of Pisa.[1]

A more systematic process for funding talented scholars emerged in embryonic form in the first half of the nineteenth century, when science academies in France and England started offering *encouragements* and grants to support worthy projects by their members.[2] To ensure the best use of funds, learned societies began formalizing the application cycle and the review process for the selection of grant recipients. Similar selection procedures had been in place for centuries at university colleges for assigning scholarships to promising students from families with limited means.[3]

With its roots steeped in patronage, grantmaking evolved in the modern era to become an effective method for identifying prospects worthy of funding. As Carnegie, Rockefeller, and Russell Sage and other industrial tycoons turned philanthropists at the beginning of the twentieth century, the private foundations they endowed to "promote the wellbeing of humankind" were inundated by requests for donations. Leveraging their business experience, trustees of these large foundations refined grantmaking as a systematic approach to "wholesale" giving. Modern philanthropic foundations select which applications to fund with the assistance of specialized evaluation panels and delegate to grantees the "retail" implementation of the charitable work.[4]

[1]The quote from Galileo is reported in Westfall (1985, p. 22). For more on Galileo's patronage see also Biagioli (1990) and references therein.

[2]See MacLeod (1971) and Crosland and Gálvez (1989).

[3]Rashdall (1895, p. 200–204) describes the examination procedures for selecting applicants at the first university college, the College of Spain founded at the University of Bologna from a bequest in 1367 and still active today.

[4]See Zunz (2012) and Leat (2016).

As World War II drew to a close, John Maynard Keynes (1945) stewarded the adoption of the grantmaking model with the creation of the Arts Council of Britain by the UK government to "stimulate, comfort and support" independent artistic initiatives in drama, music and painting.[5] At around the same time in the US, Bush (1945), building on his success as director of the wartime Office of Scientific Research and Development, forcefully argued in favor of federal support of the best, curiosity-driven "basic research in the colleges, universities, and research institutes" for a wide range of sciences. In 1946 the National Institutes of Health (NIH) greatly expanded its extramural grants program to cover all areas of biomedical research, while in 1950 the National Science Foundation (NSF) was established to fund basic research across a broad range of scientific disciplines.

As grantmaking grew exponentially in the post-war period, funding organizations developed structured procedures for soliciting and evaluating grant applications.[6] While expert peer reviewers can be entrusted with relative evaluation of projects within a specific field, the apportionment of budget across fields requires balancing opposing requests and is inherently more thorny, given that investment produces returns that are distant in time and difficult to quantify. Universities face similar problems when deciding how to allocate resources and positions across departments.

A number of major funders, like the NIH, the European Research Council (ERC) and the main Canadian funding agencies, adopt a "bottom up" approach based on an apportionment formula that allocates the total available budget to different fields depending on the applications received in each field. These research funding organizations allocate the bulk of their budget B for extramural research in proportion to applications $a_1, a_2, ..., a_N$ received in each field $i = 1, 2, ..., N$, resulting in budget

$$B_i = \frac{a_i}{\sum_{j=1}^{N} a_j} B \tag{PA}$$

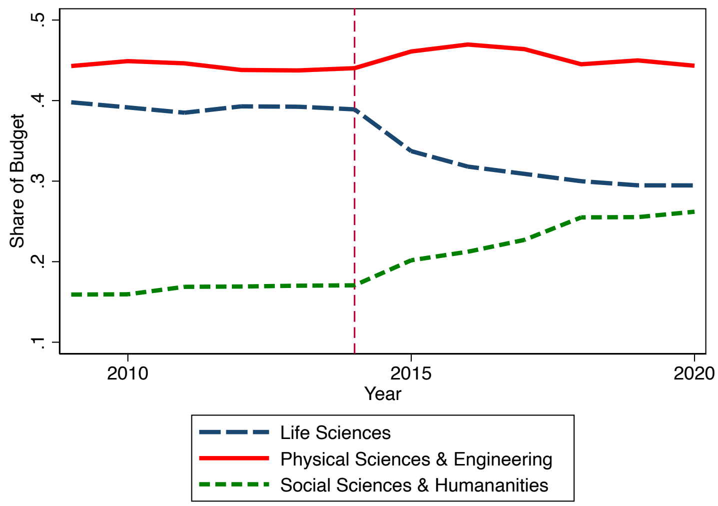
for field i. In such a system, fields compete against each others for funding on the basis of the number of applications that they attract. It is therefore important to understand what drives applications across different fields and how sensitive the funding of a given field is to what happens in other fields. Answering these questions is crucial to improve the design of the funding process.

A recent change in the apportionment rule at the ERC, the largest research funding organization in the EU, suggests that funding rules can have major effects on applications and the resulting allocation. Before 2014, ERC funds were allocated in proportion to applications to panels within each of three disciplinary domains, representing Life Sciences (LS), Physical Sciences and Engineering (PE), and Social Sciences and Humanities (SH). After 2014, the PA formula was applied across the board, so that each panel's budget became proportional to applications received in that panel relative to applications

[5]The US Congress chartered the National Endowment for the Arts in 1965. The grantmaking model for supporting the arts has since been adopted by governments throughout the world, both at the local and national level; see Upchurch (2016).

[6]Nowadays, US federal institutions such as the NIH or the NSF fund research across different fields at the tune of 172 billion dollars per year. The funding budget of the Horizon Europe program amounts to 95.5 billion euros for the period from 2021-2027.

Figure 1: Budget Shares in ERC Funding by Disciplinary Domain.



Notes: This figure shows the evolution of the budget shares for the three large fields covered by ERC funding. The 2014 reform is indicated by the vertical dashed line. *Source:* ERC data.

submitted to all panels belonging to all domains—rather than relative only to applications to the panels within its domain as before. As shown in Figure 1, the reform was followed by a substantial change in relative applications and in budget shares across panels, with a 60 percent increase in the funding for SH panels and a 20 percent decrease for LS panels.

To understand how budget allocation rules and evaluation noise impact the incentives to apply across fields, this paper formulates a foundational model of non-market allocation of resources through grantmaking. Activities are characterized by their ex-ante uncertain merit type, which captures the externality from financing the activity. Proponents of activities can come forth by applying at a cost. The review panel within each field then evaluates and ranks applications based on noisy information, with the aim of selecting the most worthy activities. Evaluation is noisy because it involves an important component of expert judgment with subjective evaluation.

As we argue, a key element determining the allocation of funds across fields is the relative evaluation noise in different fields. While in some fields researchers are likely to agree on the quality and novelty of projects, other fields that lack of a shared paradigm are characterized by more disagreement. The paper derives a general set of conditions under which an increase in evaluation noise in a field increases the applications for grants in that field and its funding, at the expense of other fields. According to our headline result, in fields with little evaluation noise, researchers who are not at the cutting edge refrain from applying because they stand a low chance of being funded, thus reducing available funds in those fields. In turn, noisy fields obtain relatively more funds, which further raise their applications.

Our framework captures key features of a wide range of other non-market resource allocation problems, such as the admission of students across courses or university degree programs, the selection of papers to publish in journals, the funding of business projects in conglomerates, and the selection of individuals to support or businesses to subsidize by government grant-in-aid programs. The next section gives a roadmap of the paper and summarizes the main insights of our analysis mainly in the context of science funding, while referring to other applications in passing. Section 9 casts our contribution within the literature. Section 10 concludes.

2 Roadmap and Main Insights

Grantmaking in a Single Field. To warm up, Section 3 sets the stage by analyzing the baseline specification with a single field populated by a continuum of candidates parametrized by their merit type. Submitting an application is costly, but allows the applicant to obtain a private benefit if the application is successful. The evaluator appraises the merit of each application received on the basis of a noisy signal—allowing for imperfect information is essential to justify the fact that many applicants do not succeed. To showcase the generality of the result, we model the information (or noise) content of the signal following a quantile function approach pioneered by Lehmann (1988). This approach is more applicable to economic problems than the classic Blackwell (1951) approach typically used in economics.

Given the limited budget available for distribution in the field, grants are supplied to the applications that receive sufficiently favorable evaluations. The evaluation on the supply side, in turn, induces candidates to apply only when they perceive a chance of success sufficiently high to compensate for the application cost. Because higher merit applicants receive more favorable evaluations, on the demand side candidates with a merit type above a threshold self-select into applying. We establish the following key comparative statics result: as evaluation becomes noisier, the probability of winning a grant becomes less responsive to the applicant's type, thus increasing the equilibrium amount of applications for given budget.

As we show, the number of grants awarded increases more (or less) than proportionally in applications when the distribution of types has increasing (or decreasing) hazard rate. Intuitively, as applications increase, the distance between the type of the marginal applicants and the average inframarginal applicants increases (or decreases). Thus, an increase in the available budget of grants results in an increase (or decrease) in the average success probability among applicants.[7]

[7] Supplementary Appendix D confronts this prediction with the impact of the spike in funding following the American Recovery and Reinvestment Act (ARRA) of 2009.

Percentiling and Grading on a Curve. When raw scores are used to decide the allocation across fields, specialized panels in each field have an incentive to inflate scores to attract more resources to their field. To counteract the resulting grade inflation across panels, from 1988 the NIH started percentiling scores within each panel (known as study section at the NIH) and introduced the payline system.[8] In each panel grants are assigned to projects that obtain percentiled scores above a level, known as payline, that is equalized across panels. Note that the payline system is equivalent to proportional allocation, given that PA implies that the success rate in field i, defined as the fraction of successful applications in field i

$$p_i = \frac{B_i}{a_i} = \frac{B}{\sum_{j=1}^{N} a_j}, \tag{1}$$

is automatically equalized across all fields, $p_i = p$. Expert evaluators in each panel are then asked to select the most fund-worthy applications so as to exhaust $100 \times p$ percent of the budget requested by the applications in the field.

Similarly, teachers have incentives to give high grades to students to increase enrollment in their classes to the benefit of their department; see Johnson (2013).[9] With grading on a curve, a constant fraction of p students enrolling in a class (or degree program) can be awarded honors, so that the budget of awards pa is proportional to enrollment. The case with constant payline directly captures grading on a curve for a course that awards a given fraction of distinction grades or honors to enrolled students.

As shown in Section 3.2, the constant-payline equilibrium is unique and stable if the type distribution has increasing hazard rate. Multiple equilibria arise only when the type distribution has decreasing hazard rate, so that the marginal type that is added to the applicants' pool as applications increase becomes closer to the average type of the inframarginal applicants.

Paradox of Relative Evaluation. As the evaluation signal in a field becomes less noisy, applications in that field in all stable equilibria unambiguously decrease—and decrease more than under fixed budget. Consider the limit case in which the grantmaker can perfectly evaluate applicants merit types without noise. Under constant payline, only a fraction $p < 1$ of applicants win. With perfect information, candidates know their ranking. Candidates not in the top $100 \times p$ percent of the applicants pool anticipate that they have no chance of succeeding and thus hold off to save the application cost.

[8] See Mandel's (1996, p. 182–188) historical account. "A percentile ranks your application relative to the other applications reviewed by your study section . . . Percentiling counters a phenomenon called "score creep" where study sections give applications increasingly better scores. As a result, scores cluster in the exceptional range, making it impossible to discriminate among applications. Each study section can apply the NIH review criteria differently, scoring either more harshly or more favorably. Percentiling counters these trends by ranking applications relative to others scored by the same study section." https://www.niaid.nih.gov/grants-contracts/understand-paylines-percentiles

[9] Relative grading can also be induced by regulation. For example, according to Texas' Top 10% Rule, students who graduate in the top ten percent of their high school class are guaranteed automatic admission to state-funded universities. See Cullen, Long, and Reback (2013) for an empirical analysis.

Iterating the logic, when the evaluation is perfect the equilibrium always unravels: no candidate applies in the only outcome compatible with equilibrium. Reversing the logic leading to market breakdown in Akerlof (1970), here good types, when they are perceived as such, make competition for scarce grants tougher and thus drive out bad types. But as applications decrease, the pool of grants is proportionally reduced, so that top types dig their own grave. Remarkably, with relative evaluation, symmetric information lead to breakdown. While in classic market settings trading breaks down when information is asymmetric, in our non-market environment asymmetry of information is needed to avoid breakdown. This is the paradox of relative evaluation.

More subtly, we show that unraveling holds when the grantmaker signal is sufficiently informative, provided that the hazard rate of the type distribution is bounded, even when the hazard rate is increasing (e.g., with logistic types). When the type distribution has vanishing hazard rate, as in the Weibull distribution with tail thicker than exponential, there is a stable equilibrium with unraveling for *any* level of noise—and the unraveling equilibrium is *unique* when the evaluation is sufficiently precise.

Partial Equilibrium. When a field is sizeable, under PA an increase in applications in a field reduces the success rate, still holding constant applications in the other fields. Our general analysis of the partial equilibrium characterizes the allocation resulting in a non-negligible field.[10] We show that when the payline decreases in applications, uniqueness and comparative statics are preserved when the type distribution has increasing hazard rate. Uniqueness is lost with decreasing hazard rate, but all the stable equilibria retain our comparative statics—applications increase in noise.

General Equilibrium across Fields. Building on the partial equilibrium analysis, Section 4 turns to grantmaking across fields where applicants in each field are possibly characterized by different parameters: application cost and grant benefit, type and signal distributions, and noise in the evaluator signal. The general equilibrium takes into account the supply-side interdependence through the budget allocation rule. We derive conditions for sub-proportional budget allocation rules (encompassing fixed budget and PA as special cases) under which equilibrium applications in a field increase when the evaluation in the same field becomes noisier and decrease when the evaluation in other fields becomes noisier.

Empirical Validation. Leveraging the 2014 reform of the ERC budget apportionment rule presented above, Section 5 empirically tests the key comparative statics prediction about the impact of noise on applications. This change in apportionment rule allows us to identify the effect of evaluation noise on the number of applications in each field, relying on a difference-in-difference design. To that effect,

[10] This case is analogous to a partial equilibrium analysis in an international trade model, where the country is large enough to affect the terms of trade. The constant-payline case corresponds to the partial equilibrium for a small country. In our context, a field does not affect the payline when it has a negligible amount of applications relative to the other fields.

we provide novel evidence on evaluation noise across fields using unique data on reviewer grades of grant applications at the Research Council of Norway (RCN). We find stark differences in evaluation noise across fields, with social sciences and applied sciences displaying more noise. We then show that the relative differences in evaluation noise across fields significantly predict changes in the number of applicants following the 2014 reform. Moreover, we find that a one standard deviation change in evaluation noise in a given field leads to an increase in the budget allocation of about 0.3 standard deviation, implying that the effect is sizeable and policy relevant.

Endogenous Evaluation Noise: Game among Fields. Section 6 endogenizes the level of evaluation noise in different fields by analyzing a game played by field representatives defending the professional interests of their field. Through this representative, each field, acting as a collective, has some capacity to introduce noise in the evaluation in their field, for example by affecting the quality of the panel members or introducing some randomization in the signal obtained by the panelists.[11]

If field representatives care about the quality of the research that is financed in their respective field, they face a tradeoff. Increasing noise increases applications, but reduces the average quality of the projects selected. In the resulting Nash equilibrium of the game, fields add noise provided that the initial noise is not too high. When the initial noise is already high in at least one of the fields, in the final Nash equilibrium the noise across fields remains asymmetric as in our baseline analysis, showing robustness. When the initial noise is relatively low in all fields, we show that the addition of noise in all fields resulting in equilibrium levels the playing field, neutralizing the initial asymmetry in noise. When, in addition to being relatively low, the initial noise is sufficiently asymmetric across fields, the final Nash equilibrium allocation results in higher social welfare than the (highly inefficient) initial asymmetric allocation.

Sorting across Fields/Courses. Section 7 extends the analysis to incorporate the demand-side interdependence generated by the ability of candidates to select courses/fields depending on their chance of obtaining a high grade. Grades have a discernible impact on the future of students (Murphy and Weinhardt 2020). Given that students tend to select courses where they expect to obtain better grades, instructors have incentives to grade generously (Achen and Courant 2009). To curb grade inflation universities respond by limiting the fraction of students who can obtain top grades and honors (Johnson 2013).

While in the baseline model candidates choose whether or not to apply/enroll, in this extension they choose one of two courses, for example physics or literature, by comparing their chance of ranking among the top $100\times p$ candidates. In the spirit of Roy's (1953) model of occupational sorting,

[11] In the case of the ERC, the representative could be the chair of the field panel or the member of the scientific council more closely associated with the field.

suppose that candidates have a two dimensional type, corresponding to their mathematical and verbal skill. Holding fixed the acceptance/merit standards, a field attracts more talented candidates when its evaluation becomes less noisy—intuitively, less talented candidates prefer to hide in the noisier field. As we argue, in equilibrium applications increase in the field's noise and decrease in the noise in the other field and the effect is stronger with grading on a curve than under fixed budget.

Design of Funding Rules. Section 8 turns to organizational design questions. Section 8.1 compares the general equilibrium allocation in the baseline model with the optimal allocation for the evaluator. The optimal amount of applications in a field increases in the evaluation noise in another field, contrary to what happens in the equilibrium induced by a sub-proportional allocation rule. Starting from the symmetric allocation resulting when fields have symmetric parameters, general equilibrium applications in a field increase excessively in noise relative to the socially optimal allocation. Evaluator welfare can then be improved by decreasing proportionality in fields characterized by less noisy evaluation.

Pooling Fields and Benchmarking. Section 8.2 considers the impact of pooling a noisier field with a more consensual field into a single panel. Supposing that applicants are still evaluated in the same way, what matters for funding once fields are pooled is the position of a candidate in the mixture distribution of scores in the two fields. Now, candidates evaluated with more (or less) noise are less (or more) likely to be at the top of the distribution. Intuitively, more accurate information increases dispersion in the scores, which matches more closely the underlying type distribution. This way, the more accurate field gains the lion's share of grants within the pooled panel, at the expense of the noisier field. Pooling fields with heterogeneous noise thus dampens the perverse effect of meritocracy on relative evaluation.

3 Grantmaking in a Single Field

This section sets the stage by formulating our baseline model of grantmaking in a single field. The field is populated by a continuum of candidates parametrized by their merit type θ, corresponding to the value created if the project is financed. Candidates know their merit, which follows distribution G in the population, with size normalized to one. For convenience assume that G admits a continuously differentiable and strictly positive density g on a connected support $\left[\underline{\theta}, \overline{\theta}\right]$, possibly unbounded on either side.[12]

To be considered for a grant award, candidates must apply at cost c, the opportunity cost of the time spent preparing the application and describing the work.[13] Applicants who are awarded grants obtain a

[12] We have $G^{-1}(0) = \underline{\theta} = -\infty$ if the support is unbounded below and $G^{-1}(1) = \overline{\theta} = \infty$ if the support is unbounded above.

[13] Application costs can well be sizeable. According to survey evidence by von Hippel and von Hippel (2015) on as-

private benefit v in terms of career advancement and kudos.[14]

An evaluator (review panel) allocates a budget of grants B to applicants on the basis of a noisy signal x about the merit type θ of each applicant. The signal is distributed according to

$$F_\sigma(x \mid \theta), \tag{2}$$

with continuously differentiable and strictly positive density f_σ and connected support, possibly unbounded on either side. We assume that the signal satisfies the monotone likelihood ratio (MLR) property

$$\frac{f_\sigma(x \mid \theta')}{f_\sigma(x \mid \theta)} \text{ increases in } x \text{ for any } \theta' > \theta, \tag{MLR}$$

so that a higher signal indicates higher merit. A key role in our analysis is played by the parameter σ, which measures the noise in the signal in the following non-parametric way

$$F_\sigma\left(F_\sigma^{-1}(q \mid \theta) \mid \theta'\right) \text{ increases in } \sigma \text{ for any } \theta' > \theta \text{ and any percentile } q \in [0,1]. \tag{3}$$

As shown in Lemma 1, an increase in noise according to this criterion (3) corresponds to a reduction in information in the sense of Lehmann (1988): any decision maker with suitably monotonic preferences gains from a reduction in noise.[15]

For algebraic tractability we often illustrate our results for the special case with additive noise where the signal has a location-scale structure, $x = \theta + \sigma\varepsilon$, with noise distribution $F(\varepsilon) = F\left(\frac{x-\theta}{\sigma}\right)$ and support $[\underline{\varepsilon}, \overline{\varepsilon}]$, possibly unbounded on either side. The signal perfectly reveals the merit when $\sigma = 0$ and becomes completely uninformative as $\sigma \to \infty$.[16]

Candidates and the evaluator have common knowledge of the model and its parameters. The evaluator allocates grants to the applicants that generate the most favorable noisy signals. The timing is as follows:

1. Candidates observe their own type θ and decide whether to apply.

tronomers and social and personality psychologists who submitted applications for basic research grants to NASA, the NIH, and the NSF, principal investigators spent on average 116 hours preparing the applications. This represents a major increase to the early day of science funding. For comparison, in 1921 the prominent German biochemist Otto Warburg submitted to the Notgemeinschaft der Deutschen Wissenschaft (Emergency Association of German Science, the forerunner of the Deutsche Forschungsgemeinschaft) a funding application with a single sentence: ‘I require 10,000 marks’; see Koppenol, Bounds, and Dang (2011).

[14] The model can also easily accommodate the addition of an embarrassment or psychological cost d borne by the candidate when the application is turned down. The cost benefit ratio c/v, which determines demand incentives, is then replaced by $(c+d)/(v+d)$.

[15] More precisely, any decision maker with preferences in the general interval dominance ordered class introduced by Quah and Strulovici (2009) obtains a higher expected payoff state by state when σ is reduced. This preference class encompasses as special cases monotone decision problems (Karlin and Rubin 1956) and single-crossing preferences (Milgrom and Shannon 1994).

[16] When the noise is additive, inverting the signal distribution $y = F((x-\theta)/\sigma)$, the quantile function of the signal is $x = \theta + \sigma F^{-1}(y)$. For every percentile y, the quantile difference $\left[\theta + \sigma F^{-1}(y)\right] - \left[\theta + \bar{\sigma} F^{-1}(y)\right]$ decreases in y for $\sigma < \bar{\sigma}$. Equivalently, the quantile transform $\theta + \sigma F^{-1}(F((x-\theta)/\bar{\sigma})) = \sigma x/\bar{\sigma} + (1-\sigma/\bar{\sigma})\theta$ is increasing in θ for $\sigma < \bar{\sigma}$.

2. The evaluator awards the available budget of grants to the applicants on the basis of the signal realizations x.

3.1 Fixed-Budget Equilibrium

To illustrate the logic of the model this section considers the case with fixed budget of grants, B. In general, equilibria have the following monotonic structure, allowing us to solve the model through a simple representation in terms of demand and supply, even though no prices are involved:

- On the supply side, the evaluator awards grants to applications with $x \geq \hat{x}$, because $E[\theta \mid x]$ increases in x by the MLR property.

- On the demand side, candidates with higher merit are more likely to win by MLR, and thus apply for $\theta \geq \hat{\theta}$.

As we show, there always exists a unique fixed-budget equilibrium and this equilibrium is stable. This version of the model allows us to uncover the logic that drives the comparative statics with respect to noise: an increase in noise necessarily raises the amount of applications submitted in the fixed-budget equilibrium.

Application Demand: Self-Selection. Expecting the evaluator to accept whenever the signal is above $\hat{x}$, candidates apply if their benefit from the grant times the expected probability of obtaining a grant outweighs the application cost

$$v\left[1 - F_{\sigma}\left(\hat{x} \mid \theta\right)\right] \geq c. \tag{4}$$

For any given acceptance standard $\hat{x}$, by the MLR property candidates apply if $\theta \geq \hat{\theta}$, where $\hat{\theta}$ is the marginal applicant implicitly defined by

$$1 - F_{\sigma}\left(\hat{x} \mid \hat{\theta}\right) = c/v, \tag{5}$$

the type whose winning probability is equal to the cost benefit ratio.

The top panel of Figure 2 illustrates the signal distribution functions for the marginal type $\hat{\theta}$ and for an inframarginal type $\theta' > \hat{\theta}$. The horizontal axis corresponds to the signal realization x. Note that the distribution for θ' lies to the right of the distribution for $\hat{\theta}$, given that the MLR property implies first-order stochastic dominance. Inverting (5), the acceptance standard $\hat{x}$ that makes type $\hat{\theta}$ indifferent about applying satisfies

$$\hat{x} = F_{\sigma}^{-1}\left(1 - c/v \mid \hat{\theta}\right). \tag{6}$$

This indifference for the marginal type pins down the demand. In the top panel, the winning probability for a given type θ can be visualized as the difference between 1 and the value of the distribution of the

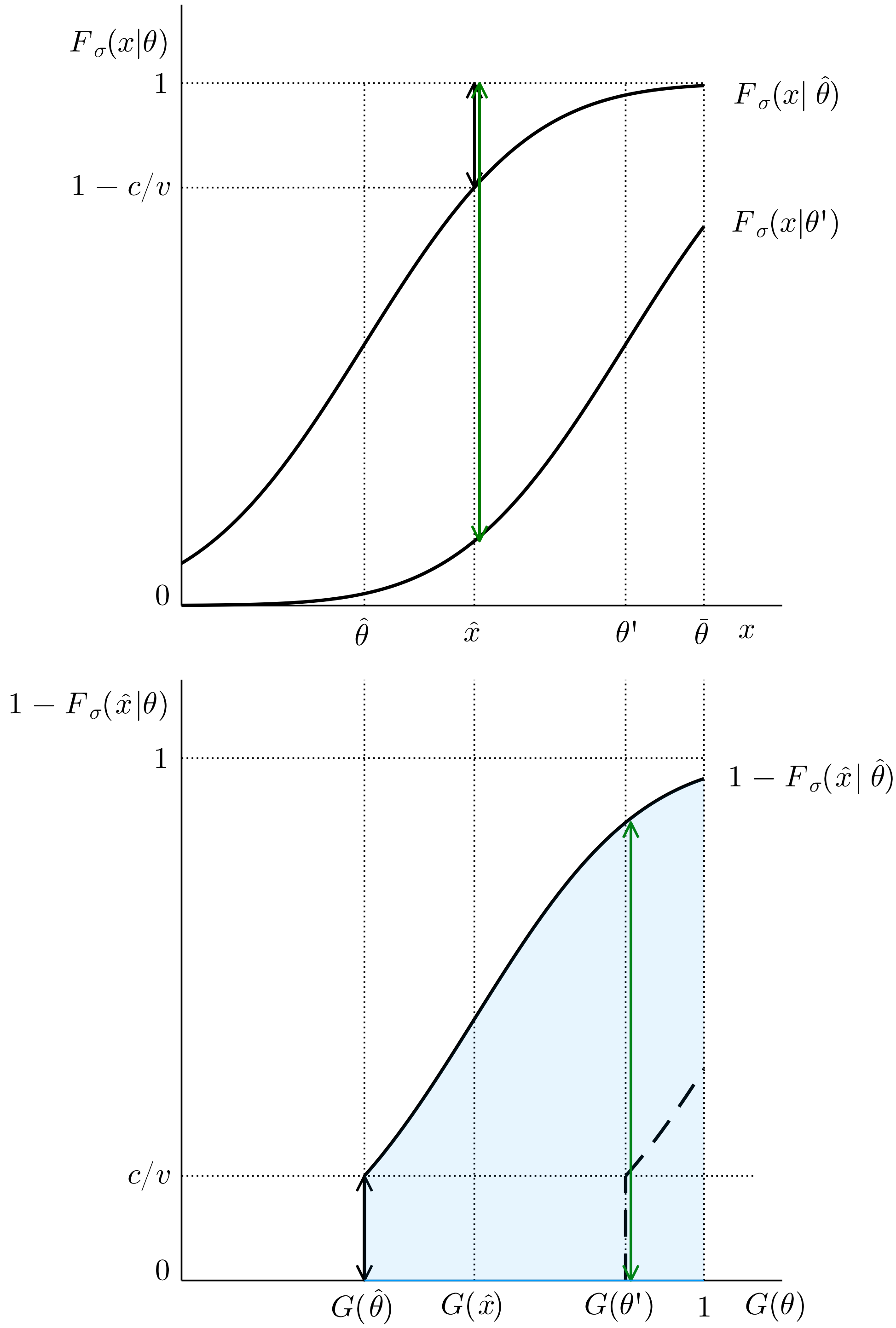


Figure 2: Fixed-budget equilibrium construction. Top panel (a): The black curves depict the signal distributions for $\hat{\theta}$ (curve to the left) and $\theta' > \hat{\theta}$ (curve to the right) with noise σ as a function of the signal realization x. The winning probabilities at acceptance standard $\hat{x}$ for type $\hat{\theta}$ and type θ' are marked in both panels as black and green vertical segments with arrows. Bottom panel (b): The black curve is the winning probability as a function of $G(\theta)$ for a uniform example with $G(\theta) = \theta$. According to the demand condition, the winning probability of the marginal type $\hat{\theta}$ is equal to c/v. The light blue area under the winning probability is the amount of grants awarded to all applicants.

signal computed at $x = \hat{x}$, according to (5): for the marginal type $\hat{\theta}$ the winning probability is equal to c/v.

The winning probability for inframarginal type $\theta' > \hat{\theta}$ is higher than c/v, as depicted in the figure. The bottom panel of Figure 2 directly displays the winning probability $1 - F_\sigma(\hat{x}\,|\,\theta)$ as a increasing function of the agent type θ on the horizontal axis, for given acceptance standard $\hat{x}$. Thus, all inframarginal types strictly prefer to apply.

Given that the distribution function increases in the signal but decreases in the type, we can see from (5) that the marginal type, $\hat{\theta}(\hat{x})$, is an increasing function of the acceptance standard, $\hat{x}$. The application demand

$$a^D(\hat{x}) = 1 - G(\hat{\theta}(\hat{x})). \tag{7}$$

is then a downward sloping function of the acceptance standard, $\hat{x}$. As the acceptance standard increases, it becomes more difficult to obtain a grant, inducing fewer candidates to apply. The marginal applicant $\hat{\theta} = G^{-1}(1-a)$ expects to obtain a grant with probability $1 - F\left(\hat{x}\,|\,G^{-1}(1-a)\right)$. Setting the winning probability for the marginal applicant equal to the cost-benefit ratio and solving for the acceptance standard that makes the marginal applicant indifferent, we conclude:[17]

Proposition 1 (Demand) ***(a)*** *The evaluator can induce a applications by setting the acceptance standard at*

$$\hat{x}^D(a) = F_\sigma^{-1}\left(1 - c/v\,|\,G^{-1}(1-a)\right). \tag{8}$$

The inverse demand is downward sloping: to induce more candidates to apply, the evaluator must reduce the acceptance standard.

Grants Awarded: Evaluation. Having derived the demand condition, the second key step of the equilibrium construction turns on the answer to the following question: How many grants must be awarded in order to induce a candidates to apply? According to the demand condition (8), by setting the acceptance standard at $\hat{x}^D(a)$, each type above the marginal, $\theta \geq G^{-1}(1-a)$, self-select into applying and obtains a grant with probability $1 - F\left(\hat{x}^D(a)\,|\,\theta\right)$. The grants awarded are then

$$A(a) = \int_{G^{-1}(1-a)}^{\bar{\theta}} \left[1 - F\left(\hat{x}^D(a)\,|\,\theta\right)\right] g(\theta)\, d\theta, \tag{9}$$

the sum of the winning probability of all applicants, weighted by their density. As applications increase, awards increase through two channels. First, the additional applicants are awarded some grants whenever they clear the acceptance standard. Second, to induce more applications, the acceptance standard $\hat{x}^D(a)$ must be reduced, thus resulting in more awards to inframarginal applicants. Overall:

[17] In the special case with additive noise, the marginal type is $\hat{\theta} = \hat{x} - \sigma F^{-1}(1-c/v)$, demand is $a^D(\hat{x}) = 1 - G\left(\hat{x} - \sigma F^{-1}(1-c/v)\right)$, and inverse demand is $\hat{x}^D(a) = G^{-1}(1-a) + \sigma F^{-1}(1-c/v)$. When information is perfect ($\sigma = 0$), the inverse demand is equal to the counter-quantile function of the type distribution.

Proposition 1 (Grants Awarded: Monotonicity) ***(b)*** *To induce a applicants, the evaluator must award $A(a)$ grants according to (9), an increasing function of a.*

Fixed-Budget Equilibrium. A fixed-budget equilibrium results when the budget of grants available is equal to the budget of grants awarded, according to (9). As in all specifications of the model, equilibrium existence follows by the intermediate value theorem, given that the award function is continuous. An equilibrium is defined to be stable if any local perturbation leads back to the equilibrium. We have:

Proposition 1 (Fixed-Budget Equilibrium) ***(c)*** *There exists a fixed-budget equilibrium. The fixed-budget equilibrium is unique and stable.*

Impact of Noise. What is the impact of an increase in noise to $\bar{\sigma} > \sigma$? At the new level of noise, in general the initial marginal type $\hat{\theta}$ is no longer indifferent. As a first step in the argument, modify the acceptance standard to restore indifference for type $\hat{\theta}$. To ensure that the winning probability for type $\hat{\theta}$ remains constant at the initial level, set the standard at $\hat{y}$ implicitly defined by

$$1 - F_{\bar{\sigma}}\left(\hat{y}|\hat{\theta}\right) = 1 - F_{\sigma}\left(\hat{x}|\hat{\theta}\right). \tag{10}$$

Inverting (10) and substituting (6), we obtain the explicit expression for the adjusted acceptance standard

$$\hat{y} = F_{\bar{\sigma}}^{-1}\left(F_{\sigma}\left(\hat{x}\,|\,\hat{\theta}\right)\,|\,\hat{\theta}\right) = F_{\bar{\sigma}}^{-1}\left(1 - c/v\,|\,\hat{\theta}\right). \tag{11}$$

The top panel of Figure 3 illustrates the construction.

Now, consider an inframarginal type $\theta' > \hat{\theta}$, who strictly prefers to apply at the initial standard $\hat{x}$. At the adjusted standard, $\hat{y}$, the winning probability for type θ' decreases provided that

$$1 - F_{\bar{\sigma}}\left(\hat{y}\,|\,\theta'\right) < 1 - F_{\sigma}\left(\hat{x}\,|\,\theta'\right). \tag{12}$$

We now link this condition to Lehmann informativeness criterion. Substituting (11) and (6) we obtain

$$F_{\sigma}\left(F_{\sigma}^{-1}\left(1 - c/v\,|\,\hat{\theta}\right)\,|\,\theta'\right) < F_{\bar{\sigma}}\left(F_{\bar{\sigma}}^{-1}\left(1 - c/v\,|\,\hat{\theta}\right)\,|\,\theta'\right).$$

This condition holds under (3), which in turn is equivalent to signal $F_{\bar{\sigma}}$ being Lehmann-noisier than signal F_{σ} by Lemma 1 in the Appendix. Intuitively, an increase in noise reduces meritocracy and thus makes the winning probability for any type less responsive to merit. As seen in the bottom panel of Figure 3, when the acceptance standard is adjusted to keep the initial marginal type indifferent, the winning probability of all inframarginal types is reduced, as illustrated by the shift from the black to the dashed red curve.

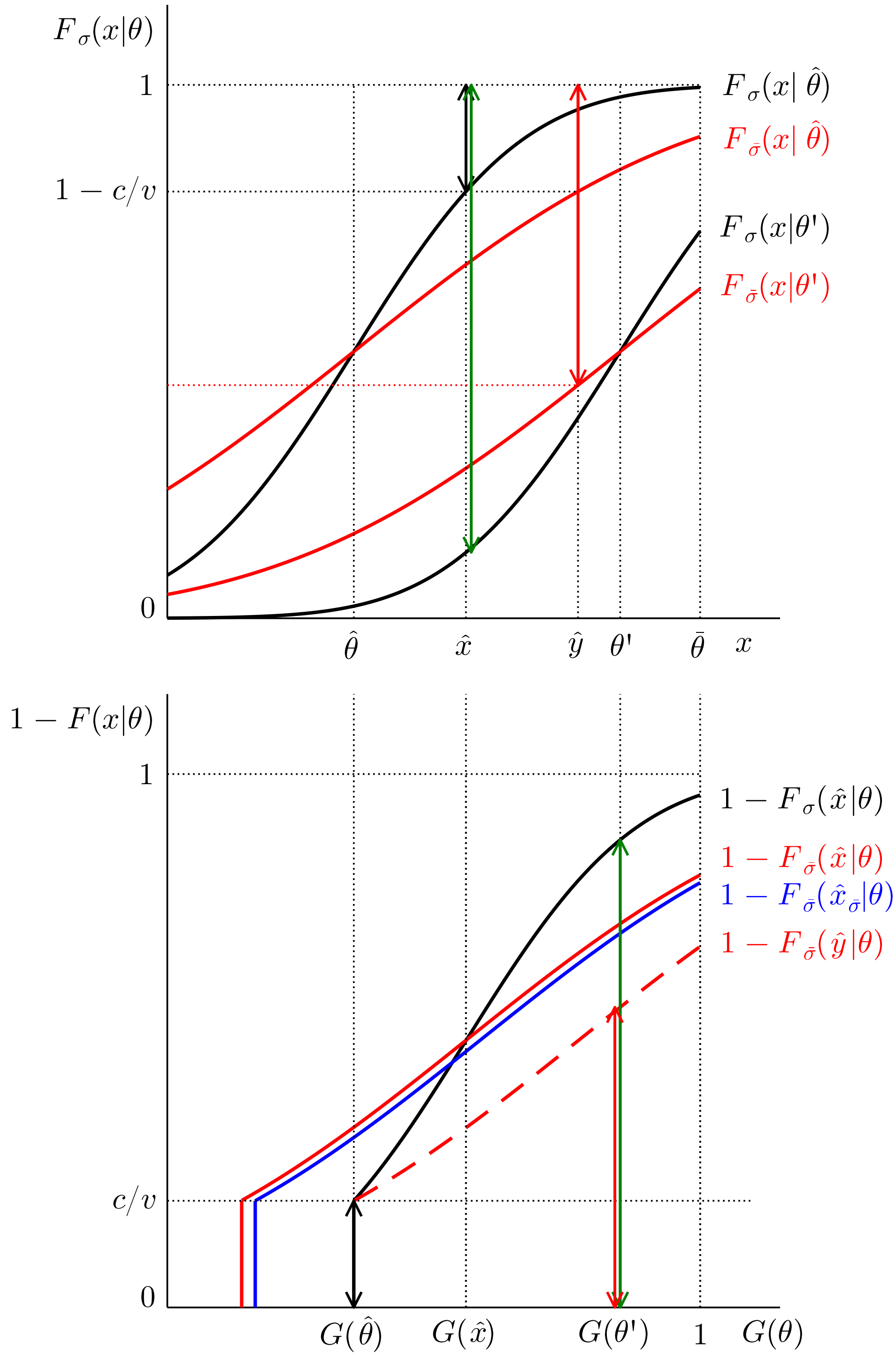


Figure 3: Impact of noise increase to $\bar{\sigma} > \sigma$. Top panel (a): Impact on signal distributions. As noise increases, the signal distributions shift from the black to the red curves. Bottom panel (b): Impact on winning probabilities. If the marginal type θ is held constant (applications do not change), as noise increases the winning probability of inframarginal type θ' is reduced from the green to the red segment, marked with arrows in both panels. Grants awarded under the dashed red curve are below the budget. To spend the initial budget the marginal applicant must be reduced, as shown by the blue segment, equating the area under the blue curve to the area under the black curve.

Weighting (12) by the density of the corresponding inframarginal types and summing (12) over all $\theta' \geq \hat{\theta}$, we conclude that the budget assigned will be under-spent,

$$\int_{\hat{\theta}}^{\bar{\theta}} [1 - F_{\bar{\sigma}}(\hat{y}\,|\,\theta)]\, g(\theta)\, d\theta < \int_{\hat{\theta}}^{\bar{\theta}} [1 - F_{\sigma}(\hat{x}\,|\,\theta)]\, g(\theta)\, d\theta,$$

whenever we retain indifference by the initial marginal type.

Proposition 1 (Impact of Noise on Award Function) ***(d)*** *As noise σ in the evaluator's signal increases, fewer grants are awarded for any given level of applications.*

To re-equilibrate, the acceptance standard, in the new equilibrium for fixed budget B and higher noise $\bar{\sigma}$, must necessarily be reduced below $\hat{y}$, in order to encourage more applications from agents with types below the initial $\hat{\theta}$. Thus, we obtain our keystone comparative statics:

Proposition 1 (Impact of Noise on Fixed-Budget Equilibrium Applications) ***(e)*** *As noise σ in the evaluator's signal increases, fixed-budget equilibrium applications increase.*

The remainder of the paper shows that this comparative statics holds more generally—and is actually strengthened—when the budget allocated to a field increases with applications. Before proceeding, we step back and prod the robustness of this result to the simplifying assumption that candidates have perfect information about their merit.

Noisy Self-Selection. Does evaluation noise increase applications also when candidates have a noisy, rather than perfect, signal t about their type θ? For the case with bilateral noisy information, we can leverage the quantile function approach to easily extend the result once we restrict to parametric signals for which we can prove the Lehmann (1988) property. For example, suppose types are normally distributed, $\theta \sim N(0,1)$, and the evaluator as well as the candidates observe normal and conditionally independent signals, $x|\theta \sim N(\theta, \sigma^2)$ and $t|\theta \sim N(\theta, \tau^2)$, respectively. To decide whether to apply, candidates must now forecast what their type is likely to be. Upon observing signal t, a candidate's updated belief about their type is $\theta|t \sim N\left(\frac{1}{1+\tau^2}t, \frac{\tau^2}{1+\tau^2}\right)$. Candidates with higher signals are more likely to believe that their type is high. Knowing that the evaluator observes a noisy signal, $x|\theta \sim N\left(\theta, \sigma^2\right)$, the candidate's belief about the signal the evaluator observes is $x|t \sim N\left(\frac{1}{1+\tau^2}t, \sigma^2 + \frac{\tau^2}{1+\tau^2}\right)$, so that an increase in σ reduces Lehmann (1988) information. Thus, exploiting the general argument presented above, an increase in the evaluator noise σ makes the winning probability less responsive to the candidate's signal about the type. Applications increase for any given budget of awards, as in the baseline model. In addition, we can establish that an increase in candidate noise τ also reduces Lehmann (1988) information and thus reduces applications.

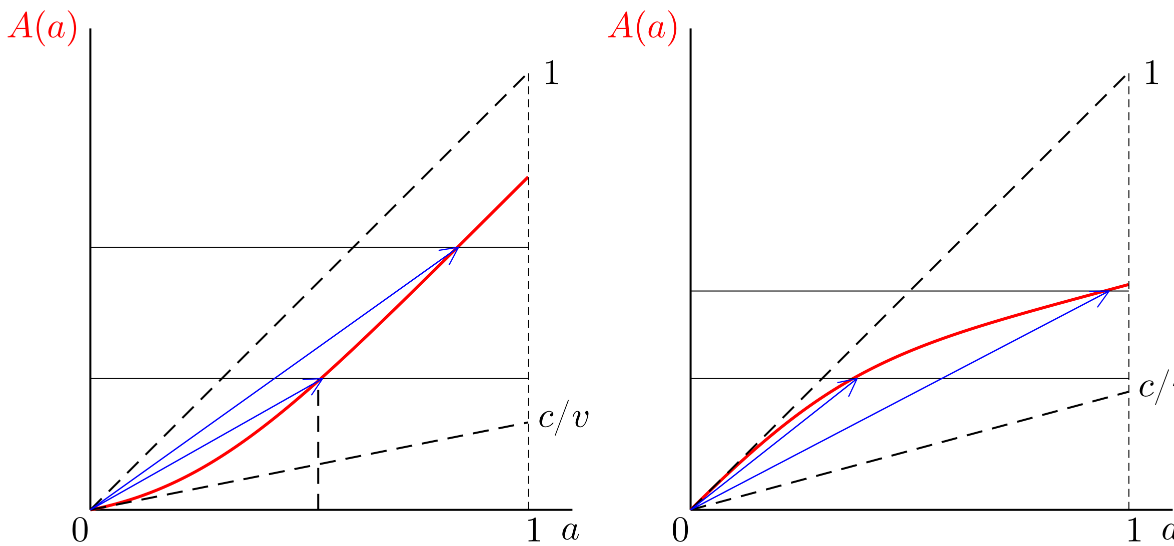


Figure 4: Award function. Left panel (a): Super-proportional, with increasing rays. Right panel (b): Sub-proportional, with decreasing rays.

3.2 Partial Equilibrium with Sub-Proportional Budget Allocation

This section considers a single-field model where the budget of grants $B(a)$ depends on applications, where a is the fraction of applicants within the unit-size population of candidates in the field. We restrict attention to the budget rules that are (weakly) increasing and *sub-proportional*

$$\frac{\partial}{\partial a}\frac{B(a)}{a} \leq 0, \tag{13}$$

i.e., the grant budget per application (weakly) decreases in applications.[18] Graphically, the segment that connects any point $(a, B(a))$ in the graph to the origin $(0,0)$ lies entirely below the graph itself. Equivalently, the rays of the function become less steep as a increases, so that none of the area below the graph of the function is hidden from an observer at the origin by the graph itself. Intuitively, (13) relaxes concavity by requiring the average, rather than the derivative, of the function to decrease. Sub-proportional budget encompasses fixed budget, $B(a) = B$, as well as the case with constant payline, $B(a) = pa$, where the fraction of grants is proportional to applications. This formulation allows us to deal with a partial equilibrium version of the full model where the payline $p(a)$ decreases in applications in a field, holding fixed the amount of applications in all other fields.

Shape of Award Function. The characterization of the equilibrium in terms of uniqueness, stability, and comparative statics hinges on the shape of the award function (9), which gives the grant awards necessary to induce a candidates to apply. Notice that the award function in the example depicted in

[18]This is the opposite of a differentiable version of sharshaped, as defined by Marshall and Olkin (2007, p. 690-691).

the bottom panel of Figure 2 is *super-proportional*

$$\frac{\partial}{\partial a}\frac{A(a)}{a} \geq 0,$$

as illustrated in the top panel of Figure 4. This is the opposite of condition (13) we imposed for the budget. Restricting attention to signals with additive noise, this property hinges on monotonicity of the hazard rate of the type distribution:

Proposition 2 (Grant Awarded: Shape) ***(a)*** *The award function $A(a)$ is super-proportional, sub-proportional or linear if the type distribution G has respectively increasing, decreasing or constant hazard rate.*

Remarkably, the curvature of the award function depends only on the type distribution, and it is not affected by the noise distribution. The rest of the section explains the logic behind this central result.

As a preliminary step, through a change of variable $t = G(\theta)$, rewrite the integrand in (9) as a function of the type percentile $t = G(\theta)$

$$A(a) = \int_{1-a}^{1} \underbrace{\left[1 - F\left(\hat{x}^D(a)\,|\,G^{-1}(t)\right)\right]}_{w(\hat{x}(a)|t):=} dt.$$

Thanks to this expression, the budget necessary to induce a applications can be visualized as the area below the winning probability curve $w(\hat{x}(a)|t)$. Equivalently, we can express this area as a rectangle with a basis that spans the integration segment (of length a) and height equal to the average winning probability under a applications, $A(a)/a$, as in the bottom panel of Figure 2. The average winning probability when a fraction a of the population applies is precisely the average of the budget necessary to induce a applications. Graphically, the average winning probability is the slope of the segment connecting $(a, A(a))$ to the origin. By definition of super-proportionality, our claim follows if we demonstrate that $A(a)/a$ has the same monotonicity as the hazard rate of G.

Next, consider an increase in applications from a to $a' > a$. To induce additional applicants, the acceptance standard must be reduced. The winning probability of all inframarginal applicants rises. The additional budget required to finance this operation is precisely the difference between the areas underlying the two winning probability curves in the bottom panel of Figure 2. The crucial point is to understand how the additional budget is divided along type percentiles.

To this end, let

$$t(\hat{x}(a)|w) = G\left(\hat{x}(a) - \sigma F^{-1}(1-w)\right) \tag{14}$$

be the percentile-type winning with probability w under application level a. As w varies, this percentile traces the inverse of the winning probability curve. By monotonicity of $w(\hat{x}(a)|t)$, the top-percentile $1 - t(\hat{x}(a)|w)$ wins with probability greater than w. Clearly, including more applicants dilates the top-percentile of candidates winning with probability w to $1 - t(\hat{x}(a')|w) > 1 - t(\hat{x}(a)|w)$. Indeed, slicing horizontally the winning probability curves in the bottom panel of Figure 2, note that $1-t(\hat{x}(a')|w)$ is equal to $1-t(\hat{x}(a)|w)$, the solid segment, plus the dashed segment. Equivalently, the ratio between the two

$$\rho(w) = \frac{1-t(\hat{x}(a')|w)}{1-t(\hat{x}(a)|w)} \tag{15}$$

is always greater than 1—this is the top-percentile dilation ratio. How does this dilation ratio vary with w? Here is where the shape of the distribution G comes into play. As shown in Lemma 2 in the Appendix, $\rho(w)$ is increasing, decreasing or constant in w if the type distribution respectively has increasing, decreasing or constant hazard rate.

In the boundary case with exponentially distributed types, the top-percentile dilation ratio $\rho(w)$ is a constant. Suppose application level is a, resulting in average winning probability $A(a)/a$. In order to increase applications to $a' > a$, the acceptance standard must be reduced—as a result, the winning probability of all inframarginal applicants goes up. When types are exponential, the fraction of applicants winning with any given probability increases by a constant ratio $\rho > 1$, maintaining their proportion constant between the two application levels. In this case, the relative composition of applicants in terms of winning probability does not change as applicants increase, resulting in a constant average winning probability: $A(a)/a = A(a')/a'$.

When G has increasing hazard rate, as the mass of applicants rises, the top-percentile dilation ratio $\rho(w)$ increases with w. We can visualize this fact in the bottom panel of Figure 2 for an example with uniformly distributed types. As the winning probability rises, the dashed segment, measuring the horizontal displacement between the two winning probability curves, decreases more slowly than the solid segment. In relative terms, the proportion of applicants winning with higher probability increases more than the proportion of applicants winning with lower probability. Thus, relatively more applicants win with higher probability, raising the average winning probability: $A(a')/a' > A(a)/a$. Intuitively, under increasing hazard rate stronger applicants are absorbing relatively more incremental grants than weaker applicants. The budget of awards required to incentivize additional applicants then increases more than proportionally in applications. The logic is reversed with decreasing hazard rate, leading to the opposite result.

Partial Equilibrium. As illustrated in the left panels of Figures 4 and 5, with increasing hazard rate the award function is super-proportional and crosses once and from below the sub-proportional budget function for an interior $a \in (0,1)$ provided that $B'(0) > A'(0)$ and $B(1) < A(1)$.[19] This equilibrium is stable, given that a small increase (or decrease) in a above (or below) the equilibrium level results in an increase (or decrease) in grants awarded above (or below) the budget, thus inducing an adjustment

[19] These two conditions are rather natural. If the hazard rate of the type distribution is unbounded, we have $A'(0) = c/v$, so that $B'(0) > A'(0)$ avoids the trivial case in which the budget is so scarce that nobody applies. Condition $B(1) < A(1)$ imposes that the budget is too scarce to accommodate all applications.

back to the equilibrium:[20]

Proposition 2 (Partial Equilibrium) ***(b)*** *If the type distribution has increasing hazard rate and the budget rule is subproportional, there is a unique partial equilibrium and this equilibrium is stable.*

The right panels of Figures 4 and 5 illustrate sub-proportional award functions resulting when the type distribution has decreasing hazard rate. To understand the role of the hazard rate condition on the equilibrium, consider the special case with proportional budget, $B(a) = pa$, where $p > 0$ represents the grants available per application. In this case with constant payline, when the type distribution has decreasing hazard rate, the unique stable equilibrium is always at the corner. If $A'(0) > p$, unraveling $a = 0$ results in the unique stable equilibrium; if instead $A'(0) < p$, all agents apply $a = 1$ in the unique stable equilibrium.

When types are exponentially distributed, $G(\theta) = 1 - \exp(-\alpha\theta)$, with constant hazard rate equal to α, the award function is proportional, i.e., linear in a with slope increasing in c/v and decreasing in α and in σ.[21] Thus, when there is a boundary level of payline $\check{p} > c/v$ such that any $a \in [0,1]$ is a constant payline equilibrium, for $p < \check{p}$ there is a unique equilibrium with unraveling and for $p > \check{p}$ there is a unique equilibrium with $a = 1$.

Impact of Budget on Success Rate The impact of an anticipated increase in the budget on the success rate, also known as payline—the widely reported fraction of successful applications—depends on the monotonicity of the hazard rate:

Proposition 2 (Impact of Budget on Success Rate) ***(c)*** *If the type distribution has increasing (or decreasing) hazard rate, the equilibrium success rate increases (or decreases) in the budget.*

Intuitively, if the type distribution has increasing hazard rate (or decreasing hazard rate), it also has decreasing (or increasing) residual expectation

$$\frac{\partial E\left[\theta - \hat{\theta} | \theta \geq \hat{\theta}\right]}{\partial \hat{\theta}} < 0 \text{ (or} > 0),$$

see Bagnoli and Bergstrom's (2005) Theorem 6.[22] Under increasing hazard rate, as applications increase, the distance between the average type of the inframarginal applicants and the merit type of the

[20] When $B'(0) < A'(0)$, there is a stable corner equilbrium with unraveling $a = 0$, in which noboby applies. When $B(1) > A(1)$, there is a stable corner equilbrium $a = 1$ in which all agents apply. In all cases, the equilbrium is unique and stable.

[21] For example, when the signal is also exponential, $F(\varepsilon) = 1 - \exp(-\varepsilon)$, the award function is $A(a) = \frac{(c/v)^{\alpha\sigma} - \alpha\sigma c/v}{1-\alpha\sigma} a$, with slope $\lim_{\alpha\to 0} A' = 1$ and $\lim_{\alpha\to\infty} A' = c/v$.

[22] A logconcave density grows at a decreasing rate and declines at an increasing rate. By Prekopa's theorem, logconcavity (logconvexity) of the density $g(\theta)$ implies logconcavity (logconvexity) of the countercumulative distribtion $1 - G(\theta)$, which in turn implies logconcavity (logconvexity) of the right-hand integral $H(\theta) = \int_{\theta}^{\bar{\theta}} \left[1 - G\left(\tilde{\theta}\right)\right] d\tilde{\theta}$, which in turn is equivalent to the fact that the residual expectation $E\left[\theta - \hat{\theta} | \theta \geq \hat{\theta}\right]$ is decreasing (increasing).

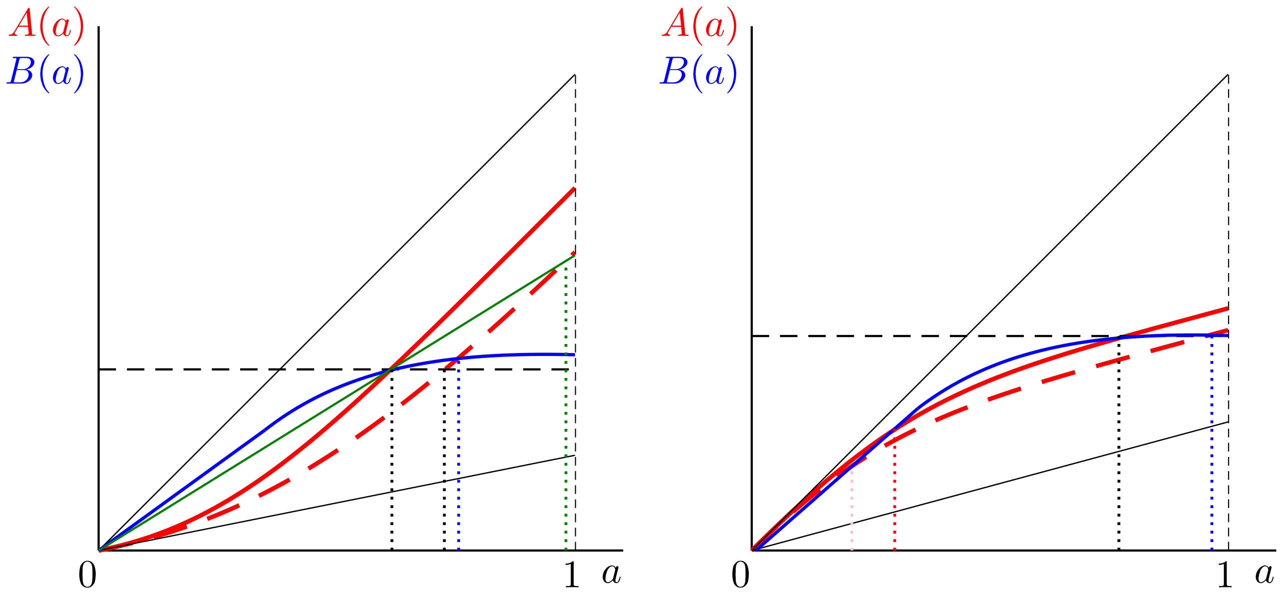


Figure 5: Comparative statics. Left panel (a): A super-proportional award function (red) crosses a sub-proportional budget (blue) only once from below, resulting in a unique stable equilibrium. As noise increases, the award function shifts to the right to the red dashed curve, resulting in an increase in applications larger than under fixed budget (dashed black horizontal segment) but smaller than under constant payline (green straight segment). Right panel (b): With sub-proportional award function multiple equilibria are possible. Here there is an unraveling stable equilibrium, an unstable equilibrium with intermediate applications, and a stable equilibrium with high applications. As noise increases, the unstable equilibrium decreases and the interior stable equilibrium increases.

marginal applicant $\hat{\theta}$ (which is reduced as a goes up) also increases. Thus, inframarginal applicants become stronger relative to the marginal applicant. Given that along the demand curve the success probability of the marginal applicant is fixed at c/v by construction, in equilibrium the average success probability of the inframarginal applicants—the success rate p—must increase in a. The opposite conclusion holds if the type distribution has a decreasing hazard rate. In the boundary case when the type distribution is exponential (with constant hazard rate), applications increase proportionally with the budget, leaving the success rate unchanged. Supplementary Appendix D confronts this prediction with the outcome of the increase in NIH budget following the 2009 Stimulus Bill.

Impact of Noise on Applications. We now return to our headline comparative statics with respect to evaluation noise σ. Recall from Proposition 1.d that an increase in noise σ shifts down the award function. Given that the budget function is increasing and subproportional, if the award function is superproportional, as in the left panel of Figure 5, from Proposition 2 we conclude that applications in the unique and stable equilibrium increase in σ more than under constant budget.[23]

When instead the award function is sub-proportional (at least on a subinterval) we have to take care of the possibility of multiple equilibria, as illustrated in right panel of Figure 5. Given that the award and budget functions are both continuous, equilibria alternate in terms of stability. In accordance with Samuelson's (1947) correspondence principle, the sign of the comparative statics is reversed for unstable equilibria. Our headline result remains valid for all stable equilibria:

Proposition 2 (Partial Equilibrium) ***(d)*** *As noise σ in the evaluator's signal increases, in any stable partial equilibrium the application level a increases.*

In addition, whenever a field's budget increases in applications, application level increases in noise more than under fixed budget, as shown in the figure.

Unraveling: The Paradox of Relative Evaluation. Turning to an extreme version of this comparative statics result, consider the outcome resulting when the evaluation is based on a perfect signal without noise, $\sigma = 0$. With fixed budget B, the most efficient allocation results, with the best a agents applying and obtaining the grant with probability 1.

What if instead the budget is proportional to applications $B = pa$, with a constant payline $p < 1$? Given any acceptance standard x, with perfect information all applicants with $\theta \geq x$ anticipate that they will succeed and thus apply in order to secure $v > c$. However, under constant payline only a fraction p of these applicants can succeed. Thus, if $a > 0$, a fraction $1 - p$ of applicants cannot succeed. But then applicants with types below the $1 - p$ quantile of the conditional type distribution among applicants, having perfect information and thus anticipating that they will not succeed, strictly prefer not to apply and save the application cost. The process goes on, until we obtain that the constant-payline equilibrium for $p < 1$ with perfect information always unravels: zero applications are submitted $a = 0$ in the unique equilibrium.[24] This unraveling logic highlights how grading on the curve, if perfect, destroys participation incentives. More generally, we have:

Proposition 2 (Unraveling) ***(e)*** *If the evaluation is based on perfect information, $\sigma = 0$, and $B(a) < a$, in the unique partial equilibrium no candidate applies.*

The result follows immediately from the observation that the award function without noise is $A(a) = a$. When information is perfect, applicants can anticipate how they will be evaluated and thus do not apply when they are sure they will not obtain a grant.

It is worth stressing that while perfect information is sufficient, it is not necessary for unraveling. Actually, well beyond the case with perfect information, unraveling results provided that $B(0) = 0$ and

[23] The comparative statics is strict when the equilibrium is interior, but it holds weakly for corner equilibria.

[24] Or, equivalently, only the highest type $\overline{\theta}$ (measure-zero) applies and is awarded a fraction p of the grant.

$A'(0) < B'(0)$. For type distributions with vanishing hazard rate, $\lim_{\theta\to\infty} g(\theta)/[1-G(\theta)] = 0$, such as Weibull with top tail thicker than exponential, we have $A'(0) = 1$ for all σ; unraveling then results whenever $B'(0) < 1$ for *any level of noise*. More generally, when the limit of the hazard rate of the type distribution is bounded (e.g., with logistic types), we have $A'(0) > c/v$ for any $\sigma < \infty$ with $A'(0)$ decreasing in σ, so that with with constant payline $B'(0) = p > c/v$ unraveling results for $\sigma < \tilde{\sigma}$ for $\tilde{\sigma}$ bounded away from zero.

4 Grantmaking across Fields

We now turn to the problem of grant allocation across fields $i = 1, \ldots, N$, each populated by a continuum of candidates representing the pool of potential applicants. Field i is characterized by specific parameters, such as type distribution G_i, signal noise distribution F_i, noise dispersion σ_i, application cost c_i, and private benefit v_i from obtaining a grant.[25] As in the baseline model, candidates are atomistic and thus they do not take into account the impact of their application decision on the acceptance standard. In each field, the evaluator (think of the review panel) allocates to field i a budget $B_i(a_1, \ldots, a_N)$ dependent on the applications submitted in all fields. In each field grants are awarded to applications with the highest expected merit within the field.

Appendix B characterizes the equilibria in the model with multiple fields, with particular attention to budget rules that satisfy a multidimensional generalization of sub-proportionality (13), condition SPA. As shown in Proposition 4, *quasi-proportional* budget *allocation* rules

$$B_i = \frac{a_i^{\rho_i}}{\sum_{j=1}^{N} a_j^{\rho_j}} B \tag{QPA}$$

with *proportionality coefficients* $\rho_i \in [0,1]$ satisfy SPA. This class encompasses the PA rule used by the ERC, NIH, and Canadian research funding organizations (for $\rho_i = 1$ for all i) and the fixed budget rule adopted by the NSF as well as by UK and Australian agencies (for $\rho_i = 0$ for all i), but more generally allows for field-specific budget responsiveness ρ_i.

If we combine sub-proportionality with increasing hazard rate, we obtain a unique stable equilibrium that preserves the comparative statics we derived for the partial equilibrium—applications increase in own noise—and reverses it for other fields—applications decrease in noise in other fields:

[25] The model can be easily extended to allow for fields to have different size, n_i, and for the individual budget, q_i, that each applicant can request to vary across fields, so that if fraction a_i of candidates apply in field i the total funds requested in the field are $n_i q_i a_i$. In practice, grant calls typically set upper bounds to the size of the award applicants can ask, sometimes depending on the career stage of the applicant. The ERC sets the maximum allowed awards at the same level for all fields. Given that almost all applicants request (and successful applicants are awarded) approximately the maximum allowed, we do not model the individual choice of amount by the applicant. In the more general case in which grant applicants request awards of different size, panel i selects the projects with the highest score so as to distribute the fraction $100 \times p$ of the total funds applied for in field i.

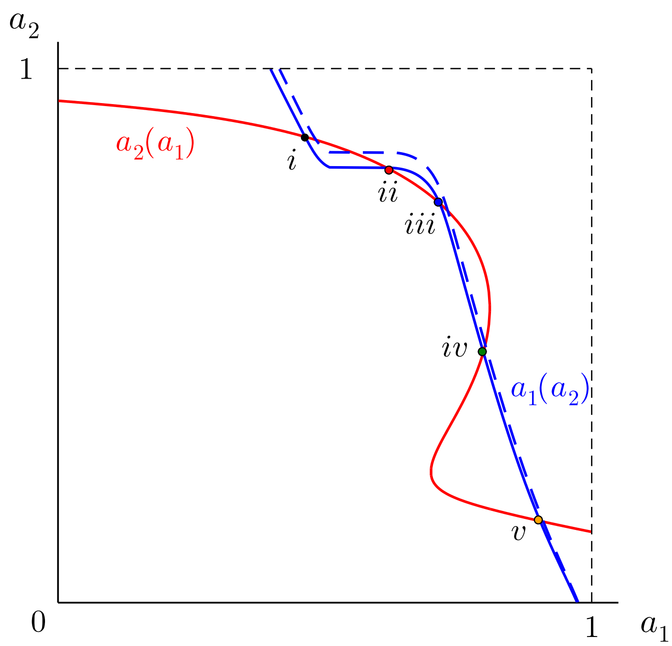


Figure 6: Construction of general equilbria at the crossing of the partial equilibrium correspondences for two fields: $a_j(a_i)$ represents the set of partial equilibrium applications in field j as a function the application level in field i. All stable equilibria, here (i), (iii), and (v), satisfy (16) and (17).

Proposition 3 (Unique General Equilibrium) ***(a)*** *If the type distributions in every field have increasing hazard rate (IHR) and the budget rule is sub-proportional, SPA, the general equilibrium* ***(i)*** *is unique,* ***(ii)*** *stable, and* ***(iii)*** *satisfies the comparative statics that an increase in noise in a field i increases applications in that field*

$$\frac{da_i^E}{d\sigma_i} \geq 0 \tag{16}$$

and decreases applications in any other field j

$$\frac{da_j^E}{d\sigma_i} \leq 0. \tag{17}$$

To understand this result, note that by Proposition 1.e, for any given budget size, noisier fields tend to attract more applications. As the budget in a field increases in applications, the increased number of applications results in an increase in the budget, which in turn induces a further increase in applications. If applications increase less than proportionally with the budget, as is the case when the type distribution has increasing hazard rate, and the budget is sub-proportional in applications, the process converges to a unique interior equilibrium that features more applications in the noisier field and fewer applications in the other fields.

Multiple Equilibria. For general type distributions violating IHR, we lose equilibrium uniqueness, as already noticed for the partial equilibrium. However, Appendix B shows that the comparative statics for all stable equilibria remains well behaved for symmetric budget rules such as PA:

Proposition 3 (Multiple General Equilibria) ***(b)*** *For general type distributions and proportional budget rule, PA, in any stable general equilibrium, applications in a field increase in the noise of that field, (16).*

Figure 6 illustrates the construction of the general equilibrium and the logic of the comparative statics result with $N = 2$ fields in an example featuring multiple equilibria. In each of two fields types follow a mixture of two normal distributions with a non-monotonic hazard rate (increasing for low types, decreasing for intermediate types, and increasing again for high types). To understand the shape of field 2's partial equilibrium correspondence (red curve in the figure) $a_2(a_1)$, given applications a_1 in field 1, note that the level of applications in field 1 impacts the equilibrium in field 2 only through the budget function $B_2(a_1, a_2)$, which decreases in a_1. As can be seen from Figure 5.b, the budget reduction created by the increase in a_1 results in a decrease in applications at any stable partial equilibrium—and in an increase in applications at any unstable partial equilibrium. The two decreasing arms of the partial equilibrium correspondence $a_2(a_1)$ in Figure 6 depict stable partial equilibria, while the increasing arm depicts an unstable interior partial equilibrium. A similar construction applies to field 1's partial equilibrium correspondence $a_1(a_2)$ (blue curve in the figure).

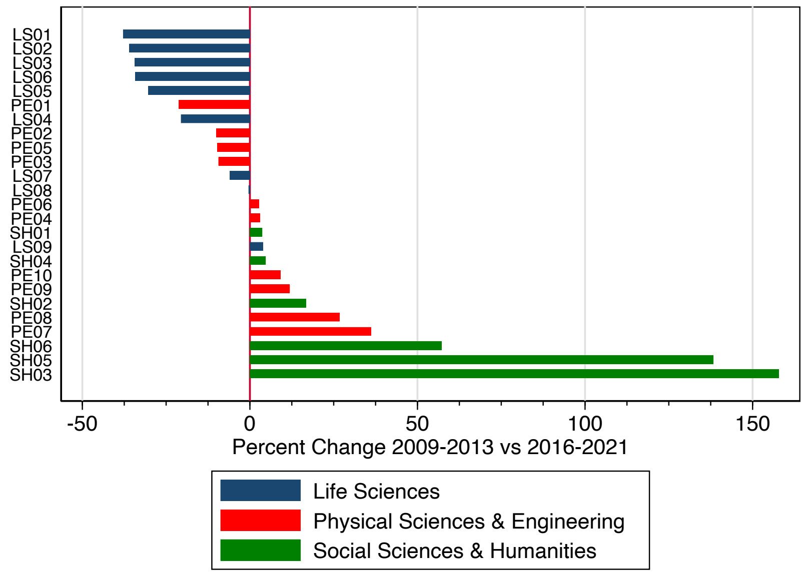
The points of intersection of the partial equilibrium correspondences $a_2(a_1)$ and $a_1(a_2)$ are general equilibria. In this example there are five interior general equilibria, marked by colored dots in the figure. A general equilibrium is stable when $a_2(a_1)$ crosses $a_1(a_2)$ from below at points at which both $a_2(a_1)$ and $a_1(a_2)$ are downward sloping. Here, general equilibria (i), (iii), and (v) are stable, while (ii) is general-equilibrium unstable ($a_2(a_1)$ crosses $a_1(a_2)$ from above) and (iv) is partial-equilibrium unstable ($a_2(a_1)$ is upward sloping at the crossing).

Consider an increase in noise in field 1. By Proposition 2.d, $a_1(a_2)$ shifts to the right (dashed blue curve). All stable equilibria satisfy both own and cross comparative statics, (16) and (17), given that at a stable equilibrium $a_2(a_1)$ slopes down and is flatter than the downward-sloping $a_1(a_2)$.

5 Empirical Validation: The 2014 ERC Funding Reform

This section exploits the natural experiment of the 2014 reform of the European Research Council (ERC) funding rules to test and quantify the central prediction of our theory about the impact of noise on applications and budget shares.

Figure 7: Changes in Budget Shares in ERC Funding by Panel.



Notes: This figure shows the relative change in funding for each ERC panels for the years 2009-2013 compared to 2016-2021, leaving out the years around the 2014 reform. LS: Life Sciences, PE: Physical Sciences and Engineering, SH: Social Sciences and Humanities. LS01 covers molecular biology, biochemestry, structural biology and molecular biophysics. SH3 covers demography, sociology, anthropology, education and communication. *Source:* ERC data.

5.1 Institutional Background

The ERC funding scheme was set up in 2007 by the European Union and has since then funded over 10,000 researchers across all fields of research with a budget of about 1.7 billion euros per year. Before 2014, ERC used to set the budget for each of the three disciplinary domains, respectively at about 39 percent for Life Sciences (LS), 44 percent for Physical Sciences and Engineering (PE), and 17 percent for Social Sciences and Humanities (SH). Within each of these domains, the budget was then allocated to panels in proportion to the budgetary demand by proposals submitted to the panels within the same domain according to PA. From 2015 the ERC started allocating funds proportionally across all panels, making a panel's budget dependent on the applications to panels from all three domains.[26] As shown in Figure 1, the relative budgets of the three domains were stable until the date of the reform, but started to diverge after 2015 with a sharp decline in the budget share devoted to LS and an increase for SH.

The reform had also consequences within domains as shown in Figure 7. Within SH, panels such as environmental studies, geography, and demography (SH3), cultural studies (SH5), and history (SH6)

[26]For example, up to 2014 SH1's (economics and management) budget depended only on applications submitted to SH panels, but from 2015 it was linked to applications to all panels in LS, PE, and SH.

saw an increase in the share of their budget, whereas the relative budget of economics and management (SH1) stayed constant. Within LS, a number of basic-research panels ranging from molecular biology (LS1) to neurosciences (LS5), saw a sustained decline, whereas more applied panels like non-medical biotechnology (LS9) had a small increase in their budget share. Whether these large changes in budget allocation can be explained by evaluation noise requires a more precise evaluation that we conduct below.

5.2 Econometric Specification

Proposition 3 relates the number of applications a_{ist} to evaluation noise, σ_i, where s indicates the seniority of the grant call.[27] To each panel we associate a panel group G_{it} within which budget allocations are made in year t, where N_{it} indicates the number of panels in G_{it}. Group membership is changing over time, due to the reform of the ERC funding described above. Before the reform, panels belonging to the same domain were competing for funds only with other panels belonging to the same domain; thus, panels are assigned to three groups depending on their domain LS, PE, and SH. After the reform, panels started competing for the overall budget regardless of the domain—thus all panels are assigned to a single group regardless of their domain.

The model developed above implies that the level of applications in a given panel i depends both on the reviewer noise in that panel as well as the noise in the panels in the same group. For the empirical analysis, we hypothesize that applications in a panel depend on the difference $\sigma_i - \bar{\sigma}_t(i)$, where $\bar{\sigma}_t(i)$ is the average of the reviewer noise in the relevant group to which the panel belongs, G_{it}. We estimate the following econometric model with a classic difference-in-difference structure

$$a_{ist} = \alpha_{is} + \alpha_t + \beta_a \left[\sigma_i - \bar{\sigma}_t(i)\right] + \varepsilon_{ist}. \tag{18}$$

We allow for panel times seniority fixed effects (α_{is}) as well as year fixed effects (α_t). The identifying variation derives from the reform that changed the funding allocation and its specific effect across panels. We supplement the analysis by relating the resulting share of the budget allocation to each panel and seniority call, B_{ist}, to evaluation noise in a similar way as in equation (18) and we denote the marginal effect of evaluation noise on the budget share by β_B. The regression assumes that the disturbances ε_{ist} are potentially heteroskedastic, contemporaneously correlated across panels, and autocorrelated.

5.3 Measuring Evaluation Noise

In order to estimate equation (18), we construct a measure of evaluation noise σ_i for each panel. We relate evaluation noise, σ_i, to agreement among grades of reviewers evaluating grant applications in

[27]The ERC has annual calls for three separate levels of seniority: Starting, Consolidator, and Advanced grants.

the field represented in a given panel. Panels in which a larger share of reviewers agree on a grade are panels with lower noise.

Given that reviewer evaluations at the ERC are not available, we quantify evaluation noise by research field using data on grant evaluations at the Norwegian Research Council (RCN). We obtained complete data including grades for the universe of RCN applications, whether successful or not over an extended period from 2002 to 2021. We focus on all the proposals that were submitted within the FRIPRO program, which funds curiosity-driven academic research proposed by researchers across all disciplines in a similar way as the ERC. The median amount of the grant is about 1.3 million dollars in 2020 and is awarded to individual researchers, rather than research centres. The referees are mostly senior scholars of international standing with a median age of 52 and based in 42 different countries, with the UK, Germany, and Sweden being the most frequent origin in 2020-2021. Given that the setting is comparable to the ERC, we assume that the evaluation dispersion at the RCN and the ERC are similar.

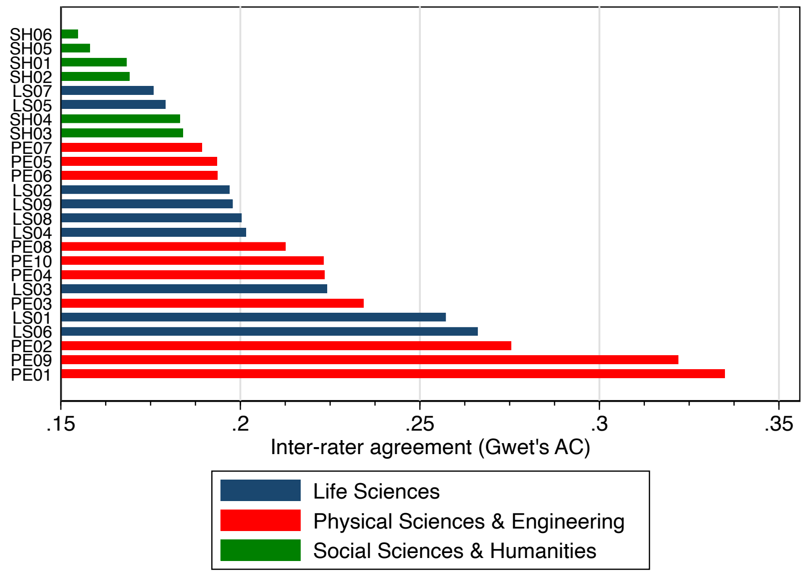
To evaluate reviewer agreements in the panels defined by the ERC, we need to assign each RCN application to one of the 25 ERC panels. We then measure reviewer agreements based on the grades of each application in a given panel. We explain each step below.

The FRIPRO program is divided into broad domains based on the fields of the applicant. Given that FRIPRO domains are similar to those used at the ERC, it is straightforward to assign each application to one of the ERC domains (LS, PE, and SH).

The next step is to assign each RCN application to one of the ERC panels within each domain. There are 6 panels in the SH domain, 10 in PE, and 9 in LS. To this end, we exploit text information to construct a prediction algorithm to assign applications into panels. We use text from both titles and abstracts of ERC and RCN applications. We use a total of 10,962 ERC applications (corresponding to the universe of successful applications between 2007 and 2020) and 9,964 RCN applications (including both successful and unsuccessful applications). In both sets of applications, the text (title and abstract) describing the research project has a total of about 2,100 characters corresponding to about 300 words. The classification is done using machine learning techniques combining BERT (Biredictional Encoder Representations from Transformers) with a neural network algorithm to classify each application. Appendix C.1 provides more details on the method. The validation accuracy ranges from 74 to 83 percent. In the main analysis, we allocate each application to an ERC panel, based on the highest predicted probability. As a robustness check, we also perform a bootstrap analysis where we randomly assign a given application to a panel according to the vector of probabilities of belonging to a particular panel.

Having assigned RCN applications to ERC panels, we develop a measure of evaluation noise for each ERC panel based on the agreement among reviewer grades in the RCN applications assigned to that panel. Overall, we have 39,077 observations of RCN reviewer grades, or about 4.05 grades per application. We calculate a measure of reviewer agreement based on all RCN applications that we assigned to a particular ERC panel, using the grades that each reviewer gave to the application.

Figure 8: Interrater Agreement by ERC Panel.



Notes: This figure shows the interrater agreement, computed as Gwet's AC, based on reviewer grades for RCN funding applications. Each application has been assigned to an ERC panel, based on text analysis. LS: Life Sciences, PE: Physical Sciences and Engineering, SH: Social Sciences and Humanities. SH06 covers history, whereas PE01 covers mathematics. *Source:* RCN and ERC data.

We borrow methods and statistics developed in education and psychology for measuring reviewer agreement based on a comparison of grading patterns of multiple evaluators. The simplest measure is the percentage agreement between reviewers, i.e., the number of times any pair of reviewers agree on the same grade divided by the number of possible pairs. This measure tends to over-estimate the amount of agreement as it does not take into account chance agreements that would occur (Cohen 1960, 1968). To adjust for chance agreements, measures such as Cohen's kappa, Fleiss' kappa, Gwet's AC and Brennan AC have been developed; see Gwet (2014) for a review and Appendix C for further details. We compute these different measures and we find them to be highly correlated with each other, with pair-wise correlations ranging between 0.84 to 0.99.

For illustration, Figure 8 plots the interrater agreement computed as Gwet's AC. The largest agreement among reviewers are in panels belonging to the PE domain, and in particular in mathematics (PE01) and universe sciences (PE09), while the lowest agreement occurs in the SH domain, especially in history (SH06) and in cultural studies (SH05). Table 3 in Appendix C provides a complete list of different agreement measures across all panels, with standard errors. As a measure of evaluation noise we use minus the interrater agreement measure.

Table 1: Effect of Evaluation Noise on Funding Outcomes

Evaluation noise based on:	Requested Funding (β_a)	Budget Shares (β_B)
Percent agreement	0.305***	0.372***
	(0.061)	(0.119)
Cohen kappa	0.303***	0.291*
	(0.078)	(0.15)
Fleiss kappa	0.31***	0.301**
	(0.079)	(0.152)
Gwet AC	0.305***	0.375***
	(0.06)	(0.118)
Brennan AC	0.305***	0.372***
	(0.061)	(0.119)
(0.129)		
Observations	845	850

Note: This table shows the effect of evaluation noise computed from different inter-reliability agreement on the requested funding, the probability of funding and allocated budget shares, see equation (18). The coefficients are expressed in standard deviation effects. Each cell is a separate regression, controlling for time and panel*seniority fixed effects. Panel-corrected standard errors are calculated using a Prais-Winsten regression, where a panel*seniority specific AR(1) process is assumed. This also allows the error terms to be panel*seniority specific, heteroskedastic, and contemporaneously correlated across panels*seniority groups. Significance level: *** p<0.01, ** p<0.05, * p<0.1.

5.4 Effect of Evaluation Noise on Applications

The estimated coefficients β_a and β_B from equation (18) are displayed in Table 1. The table displays five regressions, based on different definitions of the evaluation noise. All effects are expressed in standard deviation changes. Column 1 displays the results for grant applications. A one standard increase in the evaluation noise in a field increases applications in that field by 0.3 of a standard deviation. The effect is rather similar when using different definitions of reviewer agreement (and therefore evaluation noise). Given the econometric model, the comparative statics predictions of Proposition 3 translate to

$$\frac{\partial a_{ist}}{\partial \sigma_i} = \beta_a > 0 \qquad \frac{\partial a_{ist}}{\partial \sigma_{i'}} = -\frac{\beta_a}{N_{it}} < 0,\ i' \in G_{it}. \tag{19}$$

We thus find empirical confirmation for the predictions of our theory. Turning to the effect of evaluation noise on grant allocation shares, we also find consistent and statistically significant effects that

$$\frac{\partial B_{ist}}{\partial \sigma_i} > 0. \tag{20}$$

A standard deviation increase in own evaluation noise increases the budget of that field by 0.3 to 0.4 of a standard deviation, or by 0.5 percent from a baseline of 4 percent. In all four specifications, the reform led to a significant change in the level of funding (and therefore applications), with the ERC fields with the lowest evaluation noise lagging behind. Appendix Table 4 shows that the results are robust to using a probabilistic classification of proposals into ERC panels.

Overall, we conclude that a simple model that accounts for differences in evaluation noise across fields is able to explain the changes in ERC budget allocations that occurred after the reform, even at the finer 25-panel subdivision.

6 Endogenous Evaluation Noise: Game among Fields

Our baseline analysis takes the evaluation noise in each field as exogenously given. However—given that under proportional apportionment the application level and thus the budget allocated to a field increases in noise—each field acting as a collective might be tempted to raise its noise level, for instance by reducing the quality of panelists. Coordination at the level of each field could be achieved through a representative appointed by the scholarly association in the field. Similarly, in the application to grading, teachers in each course could easily add noise to their grades. To analyze these situations, this section sketches a game-theoretic extension of the model in which fields independently choose the noise level in their own evaluation process.

As a proof of concept, consider two fields, $i = 1,2$, with additive noise and an identical distribution of types, $G_i(\theta) = G(\theta)$. In a first stage, suppose that each field i acts as a player and simultaneously sets its noise level σ_i aiming to maximize the merit of the funded projects in the field

$$U_i(\sigma_i, \sigma_{-i}) := \int_{G^{-1}(1-a_i)}^{\overline{\theta}} \theta \left[1 - F\left(\frac{G^{-1}(1-a_i) + \sigma_i F^{-1}(1-c/v) - \theta}{\sigma_i}\right)\right] g(\theta) d\theta, \tag{21}$$

where $a_i = a_i(\sigma_i, \sigma_{-i})$ is the level of applications that results in the general equilibrium in the second stage and σ_{-i} is the noise in the other field.[28] Suppose that the action set for field i is $[\sigma_0^i, \infty)$: field i can voluntarily increase its level of noise, but cannot decrease it below a set level σ_0^i corresponding to the field's initial "intrinsic" noise level. While decreasing the level of noise is prohibitively costly, the field can freely increase the level of noise above σ_0^i.

The noise levels (σ_i, σ_{-i}) chosen in the first-stage game are publicly observed. In the second stage, candidates in each field apply and the total budget B is allocated to the two fields in proportion to applications. For any given (σ_i, σ_{-i}), the second-stage equilibrium is then determined by the solution of the system

$$\int_{G^{-1}(1-a_i)}^{\overline{\theta}} \left[1 - F\left(\frac{G^{-1}(1-a_i) + \sigma_i F^{-1}(1-c/v) - \theta}{\sigma_i}\right)\right] g(\theta) d\theta = \frac{a_i}{a_1 + a_2} B$$

for $i = 1,2$.

[28] Clearly, if instead fields only cared about maximizing the number of grants assigned to their field, each field would aim to make its signal as noisy as possible.

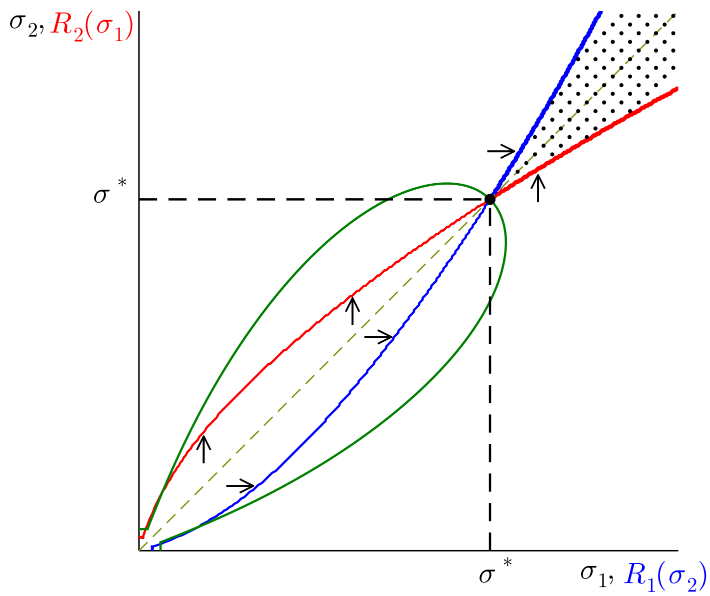


Figure 9: Equilibrium regimes in the field game with normal noise and normal types. The blue and red curves are the best replies and the green curve is level curve for the total payoff at (σ^*, σ^*).

For an example with ε_i and θ_i normally distributed, Figure 9 displays field 2's best reply $\sigma_2 = R_2(\sigma_1)$ in red as a function of field 1's noise and similarly $\sigma_1 = R_1(\sigma_2)$ in blue. To understand the shape of the best replies, note that when evaluation in the other field is perfect, $\sigma_{-i} = 0$, it is enough for field i to set an infinitesimal σ_i to obtain the entire budget. As σ_{-i} increases, field i obtains less budget, resulting in reduced applications. Then, provided that the expected merit of the marginally funded applicant is positive, it becomes optimal to increase noise to obtain a larger budget. On the one hand, holding fixed the level of applications, an increase in noise reduces the effectiveness of evaluation and thus has a negative direct effect on the field's payoff—this effect becomes stronger as noise increases. On the other hand, as noise increases, equilibrium applications rise, in turn increasing the budget allocated to the field. At the best reply level of noise, the negative effect associated with the reduced quality in winning candidates is exactly offset by the positive marginal effect of obtaining more budget. Raising the level of noise past this level reduces the field's payoff. Best replies are upward sloping for low levels of noise and concave—increasing noise has diminishing marginal returns.[29]

Depending on the initial level of noise $\sigma^0 = \left(\sigma_1^0, \sigma_1^0\right)$, there is always a unique stable equilibrium, with a basin of attraction equal to the entire action profile. As illustrated in Figure 9, there exists a benchmark level of noise $\sigma^* > 0$, such that there are three equilibrium regimes depending on the parameters:

1. **Low initial noise in all fields**: When $\sigma_1^0 \leq \sigma^*$ and $\sigma_2^0 \leq \sigma^*$, both fields sets their noise to the

[29]Best replies can eventually decrease if the expected type in the population of candidates is negative and the fraction of applicants is sufficiently high, for example because the budget is high relatively to c/v.

same level σ^*, resulting in the symmetric equilibrium (σ^*, σ^*). For these initial conditions, noise is equalized across fields—resulting in a fully-levelled playing field in the equilibrium of the field game.

2. **Highly asymmetric initial noise across fields**: When $\sigma_1^0 \geq \sigma^*$ and $\sigma_2^0 \leq R_2(\sigma_1^0)$, the equilibrium is on the red curve $\left(\sigma_1^0, R_1(\sigma_1^0)\right)$, as illustrated by the vertical arrows. In this case, field 2 increases its noise up to $R_2(\sigma_1^0)$, while field 1, which would prefer to decrease its noise, keeps it at the initial σ_1^0. Symmetrically, when $\sigma^* \leq \sigma_2^0$ and $\sigma_1^0 \leq R_1(\sigma_2^0)$, the equilibrium is on the blue curve $\left(R_1(\sigma_2^0), \sigma_2^0\right)$, as illustrated by the horizontal arrows. For parameters in this second region, the increase in noise by the less noisy field only partly levels the playing field—part of the initial asymmetry in noise persists in equilibrium.

3. **High initial noise in all fields**: When $\sigma_1^0 \geq R_1(\sigma_2^0)$ and $\sigma_2^0 \geq R_1(\sigma_1^0)$ (and thus $\sigma_1^0 \geq \sigma^*$ and $\sigma_2^0 \geq \sigma^*$), both fields do not modify their noise levels. For these parameters, the equilibrium is at $\left(\sigma_1^0, \sigma_2^0\right)$, equal to the initial level in both fields—all the initial asymmetry in noise persists.

What is the effect of the increase in noise on the total payoff in the two fields, $U_1(\sigma_1, \sigma_2) + U_2(\sigma_1, \sigma_2)$? The solid curve corresponds to the level curves of the total payoff achieved at (σ^*, σ^*). Strikingly, we conclude that when fields are only allowed to increase (but not to decrease) their noise, starting off from a relatively low but sufficiently asymmetric level of initial noise, the addition of noise in the field game can generate a gain in total payoff. For example, suppose that initially the noise levels are $(\sigma_0^1, \sigma_0^2) = (1, 1/3)$ outside the green isopayoff, but within the parameters that lead to (σ^*, σ^*). The social planner gains by allowing field 2 to raise optimally its level of noise to $R_2(1)$, to which field 1 replies with $R_1\left(R_2\left(1\right)\right)$, eventually reaching (σ^*, σ^*) with total payoff

$$U_1(\sigma^*, \sigma^*) + U_2(\sigma^*, \sigma^*) > U_1(\sigma_0^1, \sigma_0^2) + U_2(\sigma_0^1, \sigma_0^2).$$

In this second-best world, the improvement in efficiency associated with a more balanced allocation of the budget across fields is larger than the reduction in efficiency due to the less meritocratic allocation within fields.

7 Sorting across Fields/Courses

To isolate the effect of the supply-side interdependence induced by the budget allocation rule, our baseline model restricts candidates to apply in a single field. However, in the context of grantmaking, researchers who work at the crossroad between fields often have some leeway in choosing the field where they stand a better chance of funding. Similarly, university students, when selecting their major field and elective courses, might take into account their chance of obtaining an honors degree, which is typically awarded to the top 10 or 15 per cent of students in the class. This section extends the model to incorporate the demand-side interdependence generated by the ability of candidates to select which field to apply in—or which course to enroll in.

In the spirit of Roy (1951), suppose that candidates are characterized by two dimensions of talent, θ_1 and θ_2, with identical and independent distributions, $G_i(\theta_i) = G(\theta_i)$. Candidates choose to apply in either of two fields, where field i evaluates dimension i of the applicant's θ_i through noisy signal $x_i = \theta_i + \sigma_i \varepsilon$ satisfying the MLR property. For example, candidates who apply to physics are evaluated in terms of their mathematical talent, while verbal talent matters for literature candidates.

On the supply side, awards are allocated either (i) through a fixed budget, $B_i < 1/2$, or (ii) in proportion to applications, $a_i B$, where $B < 1$ represents the total budget and a_i the number of applications in field i.[30] On the demand side, we set for simplicity the application cost to zero in both fields, $c_i = 0$. Nevertheless, given that candidates can submit a single application, they face an opportunity cost equal to the foregone probability of winning a grant in the other field. In equilibrium candidates choose the field that maximizes their winning probability.

With either budget allocation rule, by the MLR property the evaluator implements a cutoff acceptance policy to assign grants according to $x_i \geq \hat{x}_i$. The equilibrium is characterized by (i) the demand-side indifference condition

$$1 - F\left(\frac{\hat{x}_1 - \theta_1}{\sigma_1}\right) = 1 - F\left(\frac{\hat{x}_2 - \theta_2}{\sigma_2}\right), \tag{22}$$

which defines an upward sloping indifference boundary $\hat{\theta}_2(\theta_1) = \hat{x}_2 + \frac{\sigma_2}{\sigma_1}(\theta_1 - \hat{x}_1)$ in the space (θ_1, θ_2) such that for any given θ_1 types $\theta_2 \leq \hat{\theta}_2(\theta_1)$ apply to field 1 and otherwise apply to field 2 and (ii) supply-side budget equations for each field

$$\begin{aligned}
\int_{\underline{\theta}}^{\bar{\theta}} \int_{\underline{\theta}}^{\hat{\theta}_2(\theta_1)} \left[1 - F\left(\frac{\hat{x}_1 - \theta_1}{\sigma_1}\right)\right] g(\theta_1) g(\theta_2)\, d\theta_2 d\theta_1 &= B_1, \\
\int_{\underline{\theta}}^{\bar{\theta}} \int_{\hat{\theta}_2(\theta_1)}^{\bar{\theta}} \left[1 - F\left(\frac{\hat{x}_2 - \theta_2}{\sigma_2}\right)\right] g(\theta_1) g(\theta_2)\, d\theta_2 d\theta_1 &= B_2.
\end{aligned}$$

To illustrate the construction, start off from initial noise levels σ_1 and σ_2, resulting in equilibrium acceptance standards $\hat{x}_1$ and $\hat{x}_2$. The solid black line in Figure 10 illustrates the indifference boundary resulting with symmetric noise $\sigma_1 = \sigma_2$ and acceptance standards $\hat{x}_1 = \hat{x}_2$, with axes expressed in terms of type percentiles $(G(\theta_1), G(\theta_2))$.[31] What is the impact of an increase in σ_2/σ_1, the noise in field 2 relative to field 1, on equilibrium applications in the two fields for the case with fixed budget?

First, the change in relative noise has an impact on selection. Holding fixed the acceptance standards $(\hat{x}_1, \hat{x}_2)$, the increase in σ_2/σ_1 induces an anti-clockwise rotation of the indifference boundary (22)

[30] If budget were abundant $B > 1$, grants would always be awarded to the entire population.

[31] Beyond the symmetric case, the indifference boundary $\hat{\theta}_2(\theta_1)$ in the type percentile space is non-linear.

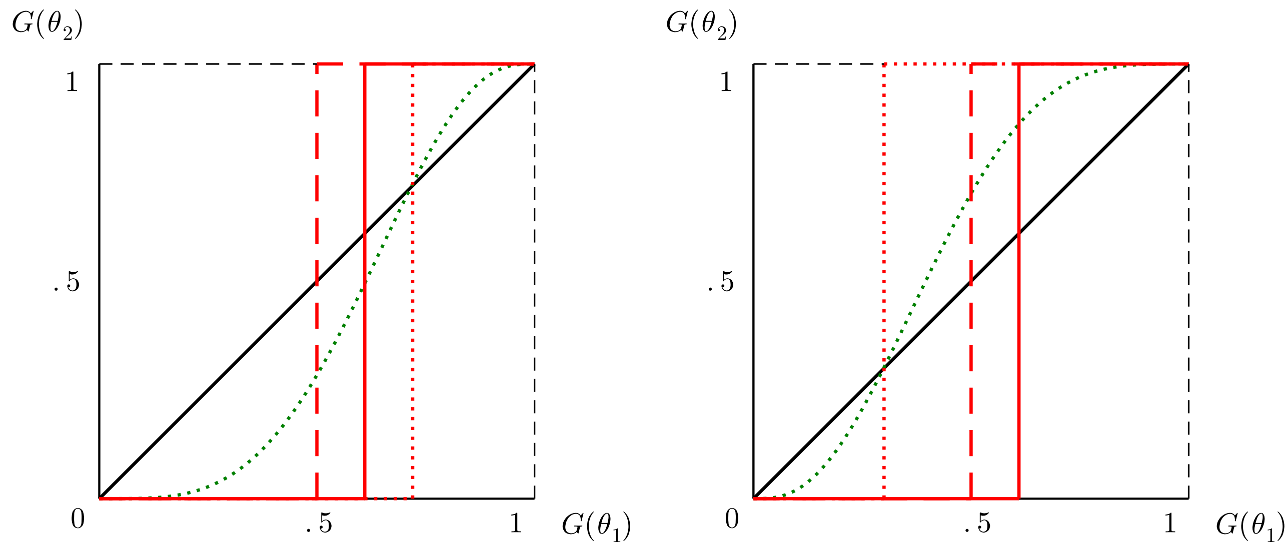


Figure 10: Comparative statics with respect to an increase in noise σ_2. Panel (a) on the left represents case (a) in which the increase in noise in step 1 results in a reduction in a_2 holding fixed $(\hat{x}_1, \hat{x}_2)$ at the initial level. Panel (b) on the right represents the opposite case.

around $(G(\hat{x}_1), G(\hat{x}_2))$, corresponding to the dotted green curve. Types in the upper-right region to the right of the dashed red curve and above the black curve—who are highly talented in both dimensions—have an incentive to flee the relatively noisier field 2 and join field 1, where they are now relatively more likely to clear the acceptance bar. Intuitively, the winning probability of these candidates is now higher in the relatively more meritocratic field 1, even though they are even more talented in dimension θ_2 than θ_1. At the same time, candidates in the lower-left region (to the left of the dashed red curve and below the black curve) with lower talent in dimension θ_1 now find the noisier field 2 more attractive, even though they are relatively worse in dimension θ_2 than θ_1. Overall, the more meritocratic field 1 attracts more talented candidates, while less talented candidates prefer to hide in the noisier field 2.

To understand how noise impacts the level of applications in the two fields, note that as a result of the first step, application levels either (a) decrease or (b) increase, depending on the relative size of the regions of types switching field, as represented respectively by the two panels in Figure 10. As a proof of concept, consider the extreme case in which evaluation becomes perfect in field 1, $\sigma_1' = 0$, resulting in a vertical indifference boundary (red dotted curves). The second step consists in adjusting the acceptance standard in field 1 until applications in field 1 are reset to the initial level. This is achieved at $\hat{x}_1 = G^{-1}(1/2)$, given that we started from a symmetric situation. In case (a), $\hat{x}_1$ should be reduced to increase applications by translating the indifference boundary to the dashed red line—by construction the area to the right of the dashed curve and to the left of the black curve (high-merit applications gained) is equal to the area to the left of the dashed curve and to the right of the black curve (low-merit applications lost). A similar construction applies to case (b), when $\hat{x}_1$ should instead be increased to move the indifference boundary to the left and thus reduce applications in field 1.

Third, with perfect information all applicants in field 1 are awarded a grant for sure.[32] Having restored applications to the initial level, grant awards would be $1/2 > B_1$, thus overspending the initial budget. When the budget is fixed, to re-equilibrate the imbalance in the budget, $\hat{x}_1$ must necessarily increase relative to the level in the second step, shifting the indifference boundary to the right, as represented by the red curves in the two panels of figures. Hence, in the new equilibrium a_1 must decrease to $a_1' = B_1 < 1/2$ and thus a_2 must increase to $a_1' = 1 - B_1 > 1/2$.

Finally, turn to the outcome under proportional budget allocation. Unraveling results in field i when evaluation is perfect in that field ($\sigma_i = 0$) or completely noisy in the other field ($\sigma_{-i} \to \infty$). Under proportional apportionment all types can guarantee the average winning probability by applying to field 2—thus, candidates who would win with probability below this level leave field 1, and the process continues until field 1 unravels, $a_1 = 0$. More generally, when candidates choose field with normal types and normal noise we verified numerically that (i) equilibrium applications increase in the field's noise and decrease in the noise in the other field and (ii) the effect is stronger under proportional than fixed budget.[33]

8 Organization of Funding with Noisy Evaluation

8.1 Design of Funding Rules

The optimal allocation for the grantmaker maximizes the total merit across fields

$$\sum_{i=1}^{N} \int_{G_i^{-1}(1-a_i)}^{\bar{\theta}_i} \theta \left[1 - F_i\left(\frac{\hat{x}^D(a_i) - \theta}{\sigma_i}\right)\right] g_i(\theta)\, d\theta$$

subject to the demand system $\hat{x}_i^D(a_i) = G_i^{-1}(1-a_i) + \sigma_i F_i^{-1}(1 - c_i/v_i)$ given the total budget available for distribution

$$\sum_{i=1}^{N} A_i(a_i) = \sum_{i=1}^{N} \int_{G_i^{-1}(1-a)}^{\bar{\theta}_i} \left[1 - F_i\left(\frac{\hat{x}_i^D(a_i) - \theta}{\sigma}\right)\right] g_i(\theta)\, d\theta = B.$$

To illustrate how the equilibrium compares to the optimal allocation, consider initially two symmetric fields with normally distributed types and signals and a PA budget rule. The identical equilibria in the first and second field are represented in the left and right panel of Figure 11 by the black dot marked

[32] In general, the composition of applicants in field 1 has now improved in the first-order stochastic order. Actually, the density for types below (above) a critical level $\tilde{\theta}$ is reduced (increased), implying stochastic dominance.

[33] We also verified that the main results of the paper extend to the field choice model with normal noise when types in each field follow the generalized normal distribution (also known as exponential power) with density $g(\theta) = \beta e^{-(|\theta-\mu|/\alpha)^\beta} / [2\alpha\Gamma(1/\beta)]$, encompassing the Laplace ($\beta = 1$), normal ($\beta = 2$), and uniform ($\beta \to \infty$) distributions. When the type distribution has increasing hazard rate ($\beta > 1$), the equilibrium is unique; unraveling results when the upper tail is thicker than exponential, $\beta < 1$.

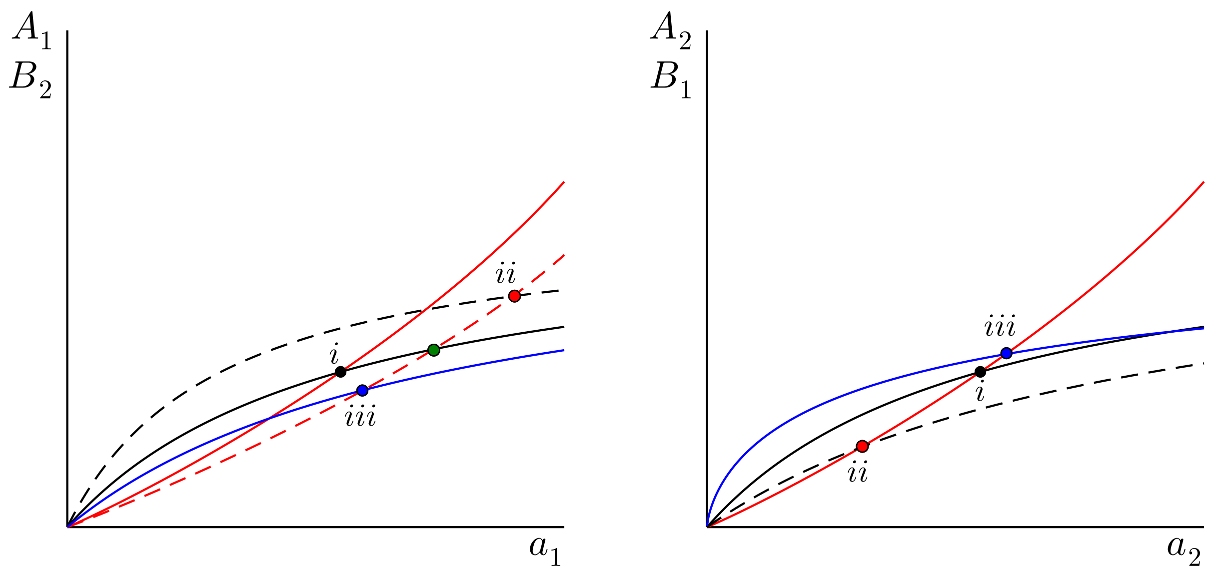


Figure 11: Optimal design of responsiveness of allocation rule. The left (a) and right (b) panels represent equilibria in field 1 and 2 respectively, where award and budget functions cross, The black dots (i) are the initial symmetric equilibria. The red dots (ii) are the equilibria with PA allocation following an increase in noise in field 1. The blue dots are the optimal allocations, which can be implemented with a subproportional budget rule in which the budget is less responsive in field 1 than 2.

as (i), at the crossing of the red curve $A_i(a_i)$ and the black curve $B_i(a_i)$. In this symmetric setting, the PA equilibrium allocation is also optimal.

As noise dispersion σ_1 in the first field increases, the award function in field 1 shifts down to the dashed red curve in the left panel, so that the partial equilibrium applications increase in field 1 from the level corresponding to the black dot (i) to the green dot. The right panel shows the reduction in the budget in field 2 to the dashed curve due to the increase in applications in field 1, as we adjust to the general equilibrium represented by the red dot (ii). In turn, this reduction in applications in field 2 increases the budget available in field 1 to the dashed curve in the left panel, leading to a further increase in applications, eventually resulting in a new general equilibrium at the red dot (ii).

As noise increases in field 1, it becomes optimal for the grantmaker to transfer some of the overall budget from the noisier field 1 to the relatively more accurate field 2, resulting in the grantmaker optimal allocation (iii) marked by the blue dots. Departing from proportional allocation PA, the grantmaker can implement this optimal allocation within the QPA class of budget rules by reducing proportionality ρ_2 in the second field.

8.2 Pooling Fields

In the baseline model each panel evaluates a single field and is characterized by a field-specific level of evaluation noise for all applicants in the panel. In reality, panels at research funding organizations are

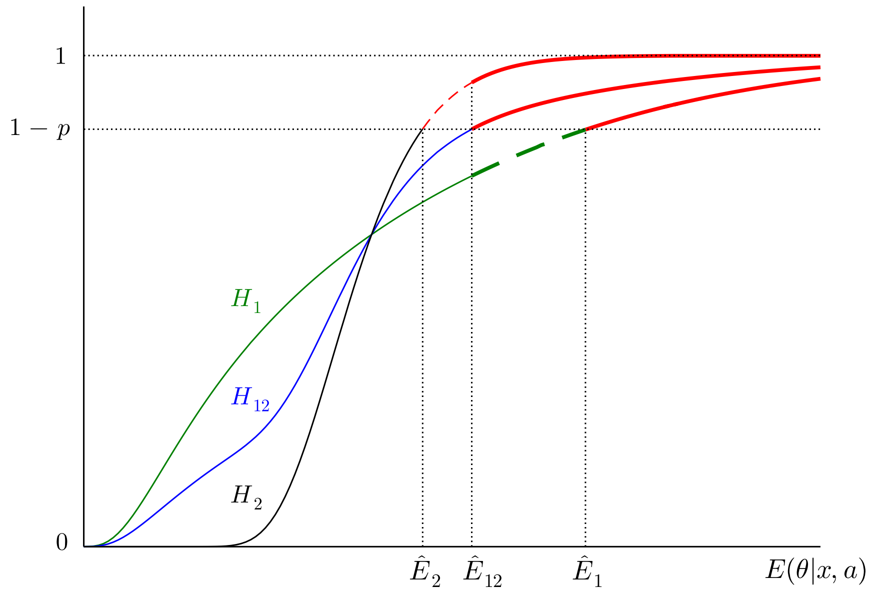


Figure 12: Impact of pooling fields. The distributions of scores for field 1 and 2 are represented in green and black respectively, When fields are evaluated in isolation, grantees correspond to the top segments of the separate distributions of scores for field 1 (green) and 2 (black). When fields are merged, scores follow the mixture distribution (blue) resulting in a loss of awards in field 1 (dashed red) and a gain in awards in field 2 (dashed green).

typically assigned applications that belong to different fields. What is the impact of pooling heterogeneous fields into a single panel, relative to assigning each field to a separate panel?

As a proof of concept, suppose there are two fields within the same discipline. Think of the basic and clinical research within a medical specialty such as pancreatic cancer research. Suppose that evaluation is noisier for clinical than basic research, $\sigma_2 > \sigma_1$.

It is useful to reinterpret the selection of grantees in a constant payline equilibrium for a single field as follows. Express the acceptance standard, rather than in terms of the signal x, in terms of the corresponding posterior expectation $E_{\sigma_i}[\theta|a_i,x]$ about the application merit θ computed via Bayes' rule. Given a_i, the constant payline acceptance standard expressed in terms of the posterior expectation (or score) $E_{\sigma_i}[\theta|a_i,x]$ is then

$$1-H_i\left(\hat{E}_i\right)=p.$$

Given that the score $E_{\sigma_i}[\theta|a_i,x]$ is an increasing function of x by the MLR property, the two representations are clearly equivalent when all applicants are evaluated with a common signal structure F_{σ_i}, as in the baseline model. The expected merit score of the marginally accepted candidate satisfies $E_{\sigma_i}[\theta|a_i,\hat{x}_i]=\hat{E}_i$, linking $\hat{x}_i$ and $\hat{E}_i$.

For a field evaluated in isolation with noise σ_1 and given application level a_1, the score $E_{\sigma_1}[\theta|a_1,x]$ is distributed according to H_1. Under constant payline the marginal score is $\hat{E}_1$, as illustrated by the green curve in Figure 12 when both types and signals are normally distributed. Similarly, for a noisier field evaluated in isolation, the black curve corresponds to distribution H_2 of scores $E_{\sigma_2}[\theta|a_2,x]$, resulting in marginal score $\hat{E}_2$. As illustrated for this example, a reduction in noise (or increase in accuracy) induces a mean-preserving clockwise rotation in the score distribution.[34]

Turn now to the case in which applicants in the two fields are pooled in the same panel. Suppose that applications are still evaluated by the same experts in each field and that σ_1 and σ_2 remain unaffected. The pooled scores in the joint panel are distributed according to a mixture of H_1 and H_2, with weights determined by the relative level of applications in the two fields

$$H_{12} = \frac{a_1}{a_1+a_2}H_2 + \frac{a_2}{a_1+a_1}H_2,$$

corresponding to the dashed blue curve in Figure 12 for the case $a_1 = a_2$. The resulting marginal score for given payline p is now $\hat{E}_{12}$, solving $1 - H_{12}\left(\hat{E}_{12}\right) = p$.

For a realistically low payline—when the winning scores with pooled fields is above the rotation point—we have

$$\hat{E}_2 < \hat{E}_{12} < \hat{E}_1,$$

as illustrated in the figure. Intuitively, the winning proposals above the payline are disproportionately originating from the more accurate field a, where scores are more extreme. Applicants in field 1 with scores between $\hat{E}_{12}$ and $\hat{E}_1$ (the green dashed segment in the figure) are now awarded grants at the expense of applicants in field 2 with scores between $\hat{E}_2$ and $\hat{E}_{12}$ (red dashed segment). The more accurate field is able to increase the fraction of successful applications above the payline p and thus enjoys a higher effective payline, $1 - H_1\left(\hat{E}_{12}\right)$. Conversely, the noisier field experiences a lower effective payline, $1 - H_2\left(\hat{E}_{12}\right)$.

Through this mechanism, pooling fields with heterogeneous noise dampens the perverse effect of meritocracy on the level of applications. The more consensual field obtains the lion's share of grants within the panel. This pattern is in line with Martin, Lindquist, and Kotchen's (2008) empirical finding that basic research has a higher success rate than clinical research at the NIH, where paylines across panels (also known as study sections) are nevertheless equalized. Clinical research suffers from being less consensual because it is pooled with basic research within the same panels, consistent with our prediction. If clinical studies and basic science were regrouped in separate panels, their success rate would be automatically equalized. However, according to our analysis, more applications would be submitted for clinical studies and fewer for basic science.

Noisier fields thus have a strong incentive to split from more consensual fields and lobby to have their own separate panel. Not only will the fraction of accepted applications increase for noisier fields

[34] In the limit as signal noise $\sigma \to \infty$, the distribution of the posterior expectation becomes a step function at the prior $E[\theta|a]$. As $\sigma \to 0$, the distribution of the posterior expectation converges to the prior distribution, $G(\theta|a)$.

that set up their own panel, but also incentives to apply will be stepped up, in turn resulting in an increase in awards. Conversely, more consensual fields prefer to be merged with noisier fields.

While under proportional allocation fields that are assigned to separate panels have a perverse incentive to increase noise relative to the other panels, pooling with other fields induces a virtuous incentive to decrease their noise relative to other fields in the same panel, thus gaining awards at the expense of other fields within the same panel.

8.3 Benchmarking

This logic can also shed light on a benchmarking practice adopted by the NIH, according to which percentiles are computed by pooling scores across recent evaluation cycles at the same panel, also known as study sections. As explained by National Institutes of Health (1988), percentiles for applications in each evaluation cycle are calculated by pooling current scores with scores given by the same study sections to the applications evaluated in the preceding two cycles, a system that is still in operation today.[35]

What might look like an inessential tweak to the payline system has important consequences. If some applications were submitted in the previous two circles, $a_{t-1} + a_{t-2} > 0$, even if in the current cycle evaluation were perfect, $\sigma_t = 0$, some budget would always be available for distribution. Hence, unraveling would not result. More generally, benchmarking dampens the impact of noise on applications by reducing the responsiveness of the budget to applications.

Similar to pooling, benchmarking can actually reverse the perverse comparative statics of proportional allocation with respect to noise. By improving its accuracy in this cycle compared to the previous cycle, a panel is able to increase the fraction of successful applications above the payline. Under the reasonable assumption that reviewers aim at assigning as many grants as possible to applicants in their panel (possibly at the expense of other panels), they now have an incentive to be more accurate than in the previous cycle, so as to increase dispersion in the posterior expectation and thus increase the number of funded applications in the panel. Through this channel, the NIH method of computing percentiles relative to the applications previously evaluated by the same panel incentivizes accurate evaluation, triggering virtuous incentives to increase accuracy, in contrast to the vicious incentives highlighted in our baseline analysis.

9 Contribution to Literature

Economists have given short shrift to grantmaking. While we perform a largely positive analysis of commonly used non-market resource allocation schemes, most previous work focuses on normative

[35] See https://www.niaid.nih.gov/grants-contracts/understand-paylines-percentiles for a detailed account.

aspects. In a pioneering application of marginal analysis, Peirce (1867) sketches the normative theory of resource allocation across research fields for a planner. As stressed at least since Arrow (1962), market forces tend to underprovide research, mostly because invention is non-rival. Governments, however, have limited information about the benefits of research in different fields. Weisbrod (1963) offers an early attempt to quantify the social benefits of medical research across diseases.[36] Weinstein and Zeckhauser (1973) link the problem of the optimal allocation of budget to fields to the decision theoretic approach underlying hypothesis testing.

Turning to positive analyses, Wildavsky (1964) describes the incremental nature of the budget apportionment process for determining government funding of the NIH in the early days; our static model abstracts from dynamic considerations.[37] Zuckerman and Merton (1971) notice that acceptance rates at leading scholarly journals vary across academic disciplines, with higher rejection rates in social sciences and humanities compared to physical sciences; our analysis shows that the performance of allocation rules with proportional elements is particularly problematic when fields are heterogeneous.[38] Rejection rates also vary along similar lines across directorates at the National Science Foundation.[39]

In terms of theory, Lazear (1997) outlines a lottery model of research funding (researchers can increase their chance of obtaining a grant by buying more tickets) but abstracts away from self-selection and noisy evaluation on which we focus. Scotchmer (2004, Chapter 8) formulates a simple dynamic model of demand for funding where high-type researchers self-select into applying and are disciplined to deliver because they expect to be funded in the future. Building on a setting with continuous types and scale-location signals similar to ours, Leslie (2005) sketches the demand side for submissions to academic journals—in addition to a complete analysis of the demand side, we add (noisy) evaluation on the supply side and characterize the equilibrium depending on the budget allocation rule.[40] See also Stephan's (2012, Chapter 6) discussion of science funding and Azoulay and Li's (2020) overview of the fledgling empirical literature on grant funding for science.[41]

In our model the application cost, akin to what Nichols and Zeckhauser (1982) call an ordeal, induces more worthy applicants to self-select. In our model, the evaluator uses an additional noisy signal about the applicant's type so that the application cost acts as an endogenous screening device.

[36] In a review of the NIH, Zeckhauser (1967) also argues that disease burden should guide funding choices.

[37] See also the formalization by Davis, Dempster, and Wildavsky (1964). Savage (1999) gives a historical account of the influence process behind university earmarks in comparison to merit-based public funding of research.

[38] Zukerman and Merton (1971, page 77) write: ". . . the more humanistically oriented the journal, the higher the rate of rejecting manuscripts for publication; the more experimentally and observationally oriented, with an emphasis on rigour of observation and analysis, the lower the rate of rejection." Referee please take notice.

[39] Cole and Cole's (1981) landmark study documents differences in agreement among reviewers (as measured by inter-rater reliability) across fields at the NSF.

[40] See also Cotton (2013) and Taylor and Yildirim (2011), focusing on discrimination issues, which we skirt.

[41] Gans and Murray (2012) overview the main funding sources available for scientists (government, private firms' internal R&D, and foundations), with a focus on comparing their different disclosure and openness requirements. Boudreau, Guinan, Lakhani, and Riedl (2016) investigate the role of the intellectual distance between evaluators' expertise and the research proposals in systematically shaping funding outcomes.

The noise in the evaluation process thus plays a key role in our model as in the literature on statistical discrimination, pioneered by Phelps (1972) and surveyed by Moro and Fang (2001). In that strand, Cornell and Welch (1996) argue that competition for ranking in a tournament discriminates against candidates the evaluator is *less* informed about. Our base model moots this channel by focusing on an evaluator who is equally informed about applicants belonging to the same field. The new effect we uncover, instead, operates across fields. Competition within a field with more noisy evaluation becomes closer to a lottery and thus encourages more applications. In turn, when the budget of grants available to a field increases in applications, the evaluator ends up inefficiently discriminating against candidates evaluated with less noise—the opposite of Cornell and Welch's outcome.

Within the agency literature, Che, Dessein, and Kartik (2013), Alonso (2018), and Frankel (2021) largely focus on how to optimally constrain biased evaluators—in our model, instead, evaluators within each field are unbiased. While our model zooms in on the noisy evaluation process of applicants, the literature on tournaments and contests—from Lazear and Rosen (1981) to O'Keeffe, Viscusi, and Zeckhauser (1984), Moldovanu and Sela (2001), Che and Gale (2003), Siegel (2009), Gross and Bergstrom (2019), and Fang, Noe, and Strack (2020)—mostly focuses on the incentives of contestants to exert effort, from which we abstract. Closer to our setting, Morgan, Sisak, and Várdy (2018) analyze the incentives of applicants to select different fields in a setting with exogenous supply, while we focus on endogenously determining the supply through the budget allocation.[42]

At a technical level, we leverage Lehmann's (1988) quantile-function approach to derive sharp predictions on the impact of evaluation noise.[43] Exploiting the structure of the problem, where evaluation noise in a field affects the other fields only through the budget allocation rule, we are able to obtain unambiguous comparative statics. Our results linking comparative statics to stability are in line with Samuelson's (1947) correspondence principle; see Hale et al. (2014) for an overview of the tools. Relative to the literature on fair division of resources among claimants, recently summarized by Thomson (2019), our model endogenizes the claims (applications are costly) and introduces imperfect verification (evaluation is noisy).

10 Conclusion

Our analysis emphasizes the central role of evaluation noise across fields in the allocation of resources. By developing a non-parametric approach to information, we derive the testable comparative statics prediction that applications increase in noise in all stable equilibria. In addition to empirically validating

[42]We also abstract away from dynamic considerations. See Board, Meyer-ter-Vehn, and Sadzik (2020) for a model of recruitment where the accuracy of evaluation endogenously depends on past recruits; Moisson and Tirole (2020) for a foray into the dynamics of cooptation; and Bardhi, Guo, and Strulovici (2020) for a characterization of when costly experimentation amplifies or dampens small differences in ability.

[43]This approach is little known in economics, with the notable exception of Persico (2000).

this result, we extend the analysis to allow candidates to choose field or course, as is most relevant in applications to course selection. Noisier fields are more attractive for weaker candidates who win with lower probability, thus reinforcing our baseline comparative statics.

We also show that incentives of fields to add noise in their evaluation tend to rebalance initial asymmetries, to the point of even increasing allocation efficiency in the spirit of second best. However, when the initial noise is sufficiently high, initial asymmetries persist as in the baseline analysis. To maximize efficiency, budget rules should be optimally designed by making the budget allocation less responsive to applications in less noisy fields. Finally, the detrimental effect of noise on selection can be dampened by pooling fields with heterogeneous noise. When pooled with noisier fields, less noisy fields obtain the lion's share of grants because their informative scores tend to be more extreme and thus end up at the top of the score distribution.

Back to the specific proportional allocation PA rule that motivated our analysis, this rule appears to be fair in treating all fields in the same way by automatically equalizing the fraction of successful projects over applications across different fields. Proportional allocation also eliminates administrative discretion and political meddling in funding allocation, given that the budget allocation is determined automatically only on the basis of relative demand from applications across fields. As another important virtue, the proportional allocation scheme has the merit of flexibly responding to demand-side signals. In spite of its simplicity, we argue that formula-based funding—as well as a general class of *sub-proportional allocation* rules SPA we characterize—has important pitfalls when fields are heterogeneous, as they typically are.

Our analysis of proportional allocation immediately applies also to large research fellowships programs, such as the EU-wide Marie Skłodowska-Curie Action (MSCA) scheme that assigns its total budget (6.16 billion euros for the period 2014-2020) in proportion to applications across all disciplines.[44] The drawbacks our analysis highlights are particularly severe for mechanisms that link the budget across very heterogeneous fields, as is the case for the ERC and MSCA, but perhaps less problematic for funders (like the NIH) that focus on research in a single domain (like medicine, even though NIH study sections cover a wide variety of disciplines, methodologies, and topics).[45]

The bottom-up formula-based approach to funding apportionment analyzed here can be contrasted to alternative top-down approaches, such as those prevailing at the NSF, in the UK, and Australia, where legislators discretionally allocate the budget across programs, following a yearly consultation process and a detailed proposal by the directors of the research funding organizations. Even at agencies that adopt proportional allocation, success rates for different programs and across fields are regularly published and closely monitored. While differences in success rates across fields in non-proportional

[44] The Canadian SSHRC Doctoral Fellowships program (covering all humanities and social sciences) also follows PA.

[45] While the great majority of NIH institutes/centers adopt the payline system and publish paylines, it is only understandable that some institutes/centers at the NIH prefer not to publish their paylines, thus retaining some flexibility when treating proposals from different panels.

systems persist over time, there is an implicit pressure to reduce the budget for fields with higher success rates in favor of fields with lower success rates.

General-interest academic journals are subject to a similar pressure to allocate space to different subfields in proportion to submissions. When co-editors are given a common target acceptance rate, fields with less accurate (or consensual) evaluation will attract more submissions.[46] Similarly, university admission boards are tempted to admit students to different programs in proportion to applications—or to increase slots available in areas that attract more applications. Giving in to this temptation may spark a race to the bottom in terms of quality of admitted students.

11 References

Achen, Alexandra C., and Paul N. Courant. 2009. "What Are Grades Made of?" *Journal of Economic Perspectives*, 23(3), 77–92.

Akerlof, George A. 1970. "The Market for 'Lemons': Quality Uncertainty and the Market Mechanism." *Quarterly Journal of Economics*, 84(3): 488–500.

Akerlof, George A. 2020. "Sins of Omission and the Practice of Economics." *Journal of Economic Literature*, 58(2), 405–418.

Alonso, Ricardo. 2018. "Recruiting and Selecting for Fit." Working Paper, London School of Economics.

Arrow, Kenneth J. 1962. "Economic Welfare and the Allocation of Resources for Invention." In: *The Rate and Direction of Inventive Activity: Economic and Social Factors* (Richard R. Nelson, ed.), Princeton University Press.

Azoulay, Pierre, and Danielle Li. 2020. "Scientific Grant Funding." NBER Working Paper No. 26889.

Azoulay, Pierre, Joshua S. Graff Zivin, Danielle Li, and Bhaven N. Sampat. 2019. "Public R&D Investments and Private-sector Patenting: Evidence from NIH Funding Rules." *Review of Economic Studies*, 86(1): 117–152.

Bagnoli, Mark, and Ted Bergstrom. 2005. "Log-Concave Probability and Its Applications." *Economic Theory*, 26(2), 445–469.

Bardhi, Arjada, Yingni Guo, and Bruno Strulovici. 2020. "Early-Career Discrimination: Spiraling or Self-Correcting?" Working Paper, Duke and Northwestern.

Bermon, Abraham, and Robert J. Plemmons (1994). *Nonnegative Matrices in the Mathematical Sciences*. Philadelphia: Society for Industrial and Applied Mathematics.

Biagioli, Mario. 1990. "Galileo's System of Patronage." *History of Science*, 28: 1–79.

[46]See also Akerlof's (2020) discussion of how a bias toward "hardness" can arise in science. Our analysis suggests a mechanism through which hardness prevails within a discipline, even though it is detrimental in the competition across disciplines. In our model, individual disciplines tend to be dominated by harder subfields and investigations with more accurate evaluation. When elements of proportionality are present in the allocation of resources across disciplines, disciplines with more accurate evaluation are destined to obtain less resources and thus become less attractive.

Blackwell, David. 1951. "Comparison of Experiments." in: *Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability*, 1: 93–102.

Board, Simon, Moritz Meyer-ter-Vehn, and Tomasz Sadzik. 2020. "Recruiting Talent." Working Paper, UCLA.

Boudreau, Kevin J., Eva C. Guinan, Karim R. Lakhani, and Christoph Riedl. 2016. "Looking Across and Looking Beyond the Knowledge Frontier: Intellectual Distance, Novelty, and Resource Allocation in Science." *Management Science,* 62(10): 2765–2783.

Bush, Vannevar. 1945. *Science: The Endless Frontier*. Washington, DC: US General Printing Office.

Caves, Richard. 2000. *Creative Industries: Contracts Between Art and Commerce*. Harvard University Press.

Che, Yeon-Koo, and Ian Gale. 2003. "Optimal Design of Research Contests." *American Economic Review*, 93(3): 646–671.

Che, Yeon-Koo, Wouter Dessein, and Navin Kartik. 2013. "Pandering to Persuade." *American Economic Review*, 103(1): 47–79.

Cohen, Jacob. 1960. "A Coefficient of Agreement for Nominal Scales." *Educational and Psychological Measurement,* 20(1), 37–46.

Cohen, Jacob. 1968. "Weighted Kappa: Nominal Scale Agreement with Provision for Scaled Disagreement or Partial Credit." *Psychological Bulletin,* 70(4), 213–220.

Cole, Jonathan, and Stephen Cole. 1981. *Peer Review in the National Science Foundation: Phase Two of a Study*. National Academy Press.

Cornell, Bradford, and Ivo Welch. 1996. "Culture, Information, and Screening Discrimination." *Journal of Political Economy*, 104(3), 542–571.

Cotton, Christopher. 2013. "Submission Fees and Response Times in Academic Publishing." *American Economic Review,* 103(1): 501–509.

Cullen, Julie Berry, Mark C Long, and Randall Reback. 2013. "Jockeying for Position: Strategic High School Choice under Texas' Top Ten Percent Plan." *Journal of Public Economics,* 97: 32-48

Crosland, Maurice, and Antonio Gálvez. 1989. "The Emergence of Research Grants within the Prize System of the French Academy of Sciences, 1795-1914." *Social Studies of Science*, 19(1): 71-100.

Davis, Otto A., Michael A. H. Dempster, and Aaron Wildavsky. 1966. "A Theory of the Budgetary Process." *American Political Science Review*, 60(3): 529–547.

Devlin, Jacob, Ming-Wei Chang, Kenton Lee, Kristina Toutanova. 2018. "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding." *arXiv preprint arXiv:1810.04805*.

European Commission. 2007. *ERC Work Programme 2008*. European Commission C (2007) 5746, 29 November 2007.

Fang, Dawei, Thomas Noe, and Philipp Strack. 2020. "Turning Up the Heat: The Discouraging Effect of Competition in Contests." *Journal of Political Economy*, 128(5): 1940–1975.

Fang, Hanming, and Andrea Moro. 2011. "Theories of Statistical Discrimination and Affirmative Action: A Survey." In: *Handbook of Social Economics*, Vol. 1, 133–200. Elsevier.

Frankel, Alexander. 2021. “Selecting Applicants.” *Econometrica*, 89(2): 615–645.

Gans, Joshua S., and Fiona E. Murray. 2012. “Funding Scientific Knowledge: Selection Disclosure and the Public-Private Portfolio.” Chapter 1 in: *The Rate and Direction of Inventive Activity Revisited* (Joshua Lerner and Scott Stern, eds.), National Bureau of Economic Research, Cambridge, MA.

Gross, Kevin, and Carl T. Bergstrom. 2019. “Contest Models Highlight Inherent Inefficiencies of Scientific Funding Competitions.” *PLoS Biology*, 17(1): e3000065.

Gwet, Kilem Li. 2014. *Handbook of Inter-Rater Reliability: The Definitive Guide to Measuring the Extent of Agreement Among Raters*. Advanced Analytics.

Hale, Douglas, George Lady, John Maybee, and James Quirk. 2014. *Nonparametric Comparative Statics and Stability*. Princeton University Press.

Johnson, Valen E. 2013. *Grade Inflation: A Crisis in College Education.* Springer-Verlag.

Johnson, Charles R., Ronald L. Smith, and Michael J. Tsatsomeros. 2020. *Matrix Positivity*. Cambridge University Press.

Keynes, J. Maynard. 1945. “The Arts Council: Its Policy and Hopes.” *The Listener*, 34(861): 31.

Koppenol, Willem H., Patricia L. Bounds, and Chi V. Dang. 2011. “Otto Warburg’s Contributions to Current Concepts of Cancer Metabolism.” *Nature Reviews Cancer*, 11(5): 325–337.

Lazear, Edward P. 1997. “Incentives in Basic Research.” *Journal of Labor Economics*, 15(1), S167-S197.

Lazear, Edward P., and Sherwin Rosen. 1981. “Rank-Order Tournaments as Optimum Labor Contracts.” *Journal of Political Economy,* 89(5): 841–864.

Lehmann, Erich L. 1988. “Comparing Location Experiments.” *Annals of Statistics*, 16(2): 521–533.

Leat, Diana. 2016. *Philanthropic Foundations, Public Good and Public Policy*. Palgrave MacMillan.

Leslie, Derek. 2005. “Are Delays in Academic Publishing Necessary?” *American Economic Review*, 95(1): 407-413.

MacLeod, Roy M. 1971. “The Royal Society and the Government Grant: Notes on the Administration of Scientific Research, 1849-1914.” *Historical Journal*, 14(2): 323–358.

Mandel, Richard. 1996. *Half a Century of Peer Review, 1946-1996*. Division of Research Grants, National Institutes of Health.

Marshall, Albert W., and Ingram Olkin. 2007. *Life Distributions*. Springer Science.

Martin, Michael R., Teresa Lindquist, and Theodore A. Kotchen. 2008. “Why Are Peer Review Outcomes Less Favorable for Clinical Science than for Basic Science Grant Applications?” *American Journal of Medicine,* 121(7): 637–641.

Moisson, Paul-Henri, and Jean Tirole. 2020. “Cooptation: Meritocracy vs. Homophily in Organizations.” University of Toulouse.

Moldovanu, Benny and Aner Sela. 2001. “The Optimal Allocation of Prizes in Contests.” *American Economic Review*, 91(3), 542–558.

Morgan, John, Dana Sisak, and Felix Várdy. 2018. "The Ponds Dilemma." *Economic Journal*, 128(611): 1634–1682.

Morris, Stephen, and Hyun Song Shin. 2003. "Global Games: Theory and Applications," In: Advances in Economics and Econometrics (Proceedings of the Eighth World Congress of the Econometric Society), edited by Matthias Dewatripont, Lars Hansen, and Stephen Turnovsky, 1: 56-114. Cambridge University Press.

Murphy, Richard and Felix Weinhardt. 2020. "Top of the Class: The Importance of Ordinal Rank." *Review of Economic Studies*, 87(6), 2777–2826.

Nichols, Albert L., and Richard J. Zeckhauser. 1982. "Targeting Transfers through Restrictions on Recipients." *American Economic Review*, 72(2), 372–377.

O'Keeffe, Mary, W. Kip Viscusi, and Richard J. Zeckhauser. 1984. "Economic Contests: Comparative Reward Schemes." *Journal of Labor Economics*, 2(1): 27–56.

Park, Hyunwoo, Jeongsik Leeb and Byung-Cheol Kim. 2015. "Project Selection in NIH: A Natural Experiment from ARRA." *Research Policy*, 44(6): 1145–1159

Peirce, Charles S. 1879. "Note on the Theory of the Economy of Research." *United States Coast Survey*, reprinted in 1967, *Operations Research*, 15(4): 643–648.

Persico, Nicola. 2000. "Information Acquisition in Auctions." *Econometrica*, 68(1), 135–148.

Phelps, Edmond. 1972. "The Statistical Theory of Racism and Sexism." *American Economic Review*, 62(4): 659–661.

Quah, John K.-H. and Bruno Strulovici. 2009. "Comparative Statics, Informativeness, and the Interval Dominance Order." *Econometrica*, 77(6): 1949–1992.

Rashdall, Hastings. 1895. *The Universities of Europe in the Middle Ages: Volume I*. Oxford: Clarendon Press.

Roy, Andrew D. 1951. "Some Thoughts on the Distribution of Earnings." *Oxford Economic Papers*, 3(2), 135–146.

Samuelson, Paul A. 1947. *Foundations of Economic Analysis*. Harvard University Press.

Savage, James D. 1999. *Funding Science in America: Congress, Universities, and the Politics of the Academic Pork Barrel*. Cambridge University Press.

Scotchmer, Suzanne. 2004. *Innovation and Incentives.* MIT Press.

Seglen, Per O. 1992. "The Skewness of Science." *Journal of the American Society for Information Science*, 43(9): 628–638.

Siegel, Ron. 2009. "All-Pay Contests." *Econometrica*, 77(1), 71–92.

Stephan, Paula. 2012. *How Economics Shapes Science*. Harvard University Press.

Taylor, Curtis R. and Huseyin Yildirim. 2011. "Subjective Performance and the Value of Blind Evaluation." *Review of Economic Studies*, 78(2): 762–794.

Thomson, William. 2019. *How to Divide When There Isn't Enough: From Aristotle, the Talmud, and Maimonides to the Axiomatics of Resource Allocation.* Cambridge University Press.

Upchurch, Anna Rossler. 2016. *The Origins of the Arts Council Movement: Philanthropy and Policy*. London: Palgrave MacMillan.

van Zwet, William R. 1964. *Convex Transformations of Random Variables*. Amsterdam: Mathematisch Centrum.

von Hippel, Ted, and Courtney von Hippel. 2015. "To Apply or Not to Apply: A Survey Analysis of Grant Writing Costs and Benefits." *PLoS ONE*, 10(3): e0118494.

Weinstein, Milton and Richard Zeckhauser. 1973. "Critical Ratios and Efficient Allocation." *Journal of Public Economics*, 2(2): 147–157.

Westfall, Richard S. 1985. "Science and Patronage: Galileo and the Telescope." *Isis*, 76(1): 11–30.

Weisbrod, Burton. 1961. *Economics of Public Health*. University of Pennsylvania Press.

Wildavsky, Aaron. 1964. *The Politics of the Budgetary Process*. Little, Brown and Company.

Zeckhauser, Richard. 1967. *Some Thoughts on the Allocation of Resources to BioMedical Research*. Occasional Paper No. 4. U.S. Department of Health, Education and Welfare: Office of Planning and Evaluation.

Zuckerman, Harriet, and Robert K. Merton. 1971. "Patterns of Evaluation in Science: Institutionalisation, Structure and Functions of the Referee System." *Minerva*, 9(1): 66-100.

Zunz, Olivia. 2012. *Philanthropy in America: A History*. Princeton University Press.

A Proofs

This appendix establishes two simple lemmas that play a key role in Propositions 1 and 2, which are proved in the text. Proposition 3 is proved in Appendix B.

Lemma 1 (Lehmann Equivalence) *Noise increase in the sense of Lehmann*

$$F_{\bar{\sigma}}^{-1}\left(F_{\sigma}\left(x\,|\,\theta\right)\,|\,\theta\right) \text{ decreases in } \theta \text{ for any } \bar{\sigma} > \sigma \text{ and for any signal realization } x \tag{23}$$

if and only if

$$F_{\sigma}\left(F_{\sigma}^{-1}\left(q\,|\,\theta\right)\,|\,\theta'\right) \text{ increases in } \sigma \text{ for any } \theta' > \theta \text{ and any percentile } q \in [0,1]$$

according to (3).

Proof of Lemma 1. Fixing $\bar{\sigma} > \sigma$, $\theta' > \theta$ and $q \in [0,1]$, condition 3 gives

$$F_{\sigma}\left(F_{\sigma}^{-1}\left(q\,|\,\theta\right)\,|\,\theta'\right) < F_{\bar{\sigma}}\left(F_{\bar{\sigma}}^{-1}\left(q\,|\,\theta\right)\,|\,\theta'\right).$$

Define $x = F_{\sigma}^{-1}\left(q\,|\,\theta\right)$ and $y = F_{\bar{\sigma}}^{-1}(q\,|\,\theta)$, we can rewrite

$$F_{\sigma}\left(x\,|\,\theta'\right) < F_{\bar{\sigma}}\left(y\,|\,\theta'\right).$$

Applying $F_{\sigma}^{-1}(\cdot\,|\,\theta')$ on both sides, substituting $y = F_{\sigma}^{-1}(F_{\sigma}(x|\theta)\,|\theta)$ in terms of x, we have

$$F_{\sigma}^{-1}\left(F_{\sigma}\left(x\,|\,\theta'\right)\,|\,\theta'\right) < F_{\sigma}^{-1}(F_{\sigma}(x\,|\,\theta)\,|\,\theta).$$

Since q is arbitrary, any signal realization can be obtained as $x = F_{\sigma}^{-1}(q\,|\,\theta)$, so that condition (23) holds. The other direction proceeds along similar lines.

Lemma 2 (Top Percentile Dilation) *With additive noise, $F_{\sigma}(x|\theta) = F\left(\frac{x-\theta}{\sigma}\right)$, the top-percentile dilation ratio $\rho(w)$ defined in (15) is increasing, decreasing or constant in w whenever the type distribution G respectively has increasing, decreasing or constant hazard rate.*

Proof of Lemma 2. Fix $a' > a$. From the definition (14) of the percentile $t(\hat{x}(a)|w)$, we obtain

$$\rho(w) = \frac{1 - G(\hat{x}(a') - \sigma F^{-1}(1-w))}{1 - G(\hat{x}(a) - \sigma F^{-1}(1-w))} = \frac{1 - G(\theta(\hat{x}(a')|w))}{1 - G(\theta(\hat{x}(a)|w))}$$

where we define $\theta(\hat{x}(a)|w) = \hat{x}(a) - \sigma F^{-1}(1-w)$ and $\theta(\hat{x}(a')|w) = \hat{x}(a') - \sigma F^{-1}(1-w)$. Given that the acceptance standard $\hat{x}$ is decreasing in applications a by the demand condition (6) where $\hat{\theta}\,(a) = G^{-1}\,(1-a)$, for any w we have $\theta(\hat{x}(a)|w) > \theta(\hat{x}(a')|w)$. By the location structure we have $\frac{d}{dw}\theta(\hat{x}(a)|w) = \frac{d}{dw}\theta(\hat{x}(a')|w) > 0$, given that $\theta(\hat{x}(a)|w)$ is the inverse of the winning probability, which increases with θ for any level of applications a.

Denoting the hazard rate of G by $h(\theta) = \frac{g(\theta)}{1-G(\theta)}$, we can write

$$1 - G(\theta) = \exp\left(-\int_{0}^{\theta} h(x)dx\right).$$

Exploiting the fact that the exponential transforms quotients into differences, the additivity of the integral and $\theta(\hat{x}(a)|w) > \theta(\hat{x}(a')|w)$, we obtain

$$\rho(w) = \exp\left(\int_{\theta(\hat{x}(a')|w)}^{\theta(\hat{x}(a)|w)} h(x)dx\right).$$

By Leibniz integral rule, collecting the term $\delta(w) := \frac{d}{dw}\theta(\hat{x}(a)|w) = \frac{d}{dw}\theta(\hat{x}(a')|w)$, we have

$$\frac{d}{dw}\rho(w) = \rho(w)\delta(w)[h\,(\theta(\hat{x}(a)|w)) - h(\theta(\hat{x}(a')|w))].$$

As $\theta(\hat{x}(a)|w) > \theta(\hat{x}(a')|w)$ and $\delta(w) > 0$, this quantity is positive/constant/negative if and only if h is increasing/constant/decreasing.

B Equilibrium with Multiple Fields

A general equilibrium $a^E = \left(a_1^E, \ldots, a_N^E\right)$ solves in every field i,

$$\frac{A_i(a_i, \sigma_i)}{a_i} = \frac{B_i\left(a_i, a_{-i}\right)}{a_i}$$

equating the average awards and the average budget per applicant. By the implicit function theorem, the comparative statics with respect to noise are given by

$$\frac{\partial a}{\partial \sigma} = -\left[\frac{\partial\left(A/a\right)}{\partial a} - \frac{\partial\left(B/a\right)}{\partial a}\right]^{-1} \frac{\partial\left(A/a\right)}{\partial \sigma}$$

where $J = \partial\left(A/a\right)/\partial a - \partial\left(B/a\right)/\partial a$ is the Jacobian matrix of the system with respect to applications. By Proposition 1.d, Lehmann informativeness guarantees that $\partial\left(A/a\right)/\partial\sigma$ is a negative diagonal matrix and therefore the comparative statics have the same sign pattern as the inverse of the Jacobian, J^{-1}.

An equilibrium can also be interpreted as the steady state of a dynamic adjustment process: if excess average grants are awarded in field i, $A(a_i;\sigma)/a_i > B\left(a_i\right)/a_i$, then the grantmaker raises the acceptance standard, which induces fewer applications and reduces the average grants awarded; and vice versa. Formally, this is modelled as the differential equation

$$\frac{da_i}{dt} = -H_i\left(\frac{A_i(a_i;\sigma_i)}{a_i} - \frac{B_i\left(a_i, a_{-i}\right)}{a_i}\right)$$

where the speed of adjustment H increases in the distance from the equilibrium, $H_i' > 0$, and is zero at equilibrium, $H(0) = 0$. In the neighborhood of an equilibrium, the linear approximation of the adjustment process is

$$\begin{aligned}\frac{da_i}{dt} \quad \approx \quad & -H_i'\left(\frac{A_i\left(a_i^E;\sigma_i\right)}{a_i} - \frac{B_i\left(a_i^E, a_{-i}^E\right)}{a_i}\right) \\ & \left\{\left(\frac{\partial\left(A_i/a_i\right)}{\partial a_i} - \frac{\partial\left(B_i/a_i\right)}{\partial a_i}\right)\left(a_i - a_i^E\right) - \sum_{j\neq i}\frac{\partial\left(B_i/a_i\right)}{\partial a_j}\left(a_j - a_j^E\right)\right\}\end{aligned}$$

where the partial derivatives are evaluated at the equilibrium. This can be expressed in matrix form $-DJ\left(a - a^E\right)$, where D is a positive diagonal matrix with element $d_{ii} \equiv H_i'\left(\frac{A_i\left(a_i^E;\sigma_i\right)}{a_i} - \frac{B_i\left(a_i^E, a_{-i}^E\right)}{a_i}\right) > 0$ and J is the Jacobian of the equilibrium system. In order for the system to be dynamically stable, $-DJ$ must be a stable matrix: every eigenvalue of $-DJ$ must have a negative real part. Equivalently, every eigenvalue of DJ must have a positive real part.

Proof of Proposition 3.b. First, note that under the proportional budget rule, PA, the Jacobian $J = \partial (A/a)/\partial a - \partial (B/a)/\partial a$ is symmetric, and therefore has real eigenvalues. If the system is stable, all the eigenvalues have positive real parts. Combined with symmetry, this implies that all the eigenvalues are real and positive, and therefore the matrix is positive definite. Second, recall that the inverse of a positive definite matrix is also positive definite, and that positive definite matrices have positive diagonal elements. Therefore $\left[\frac{\partial(A/a)}{\partial a} - \frac{\partial(B/a)}{\partial a}\right]^{-1}$ has a positive diagonal. Thus, applications in field i increase in noise dispersion in that field.

Next, we define sub-proportionality for a multivariate budget rule.

Definition 1 (Sub-Proportional Budget) *A budget function $B(a_1, \ldots, a_n)$ is* sub-proportional *if the negation of the Jacobian of the average budget per applicant*

$$-\left[\frac{\partial (B/a)}{a}\right] = -\begin{bmatrix} \frac{\partial(B_1/a_1)}{\partial a_1} & \frac{\partial(B_1/a_1)}{\partial a_2} & \cdots & \frac{\partial(B_1/a_1)}{\partial a_N} \\ \frac{\partial(B_2/a_2)}{\partial a_1} & \frac{\partial(B_2/a_2)}{\partial a_2} & \cdots & \frac{\partial(B_2/a_2)}{\partial a_N} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial(B_N/a_N)}{\partial a_1} & \frac{\partial(B_N/a_N)}{\partial a_2} & \cdots & \frac{\partial(B_N/a_N)}{\partial a_N} \end{bmatrix} \tag{SPA}$$

has:

1. *non-negative principal minors*

$$\det\left(-\left[\frac{\partial (B/a)}{a}\right]^{\iota}\right) \geq 0 \text{ for } \iota \in \mathscr{P}(\mathscr{N}), \tag{SPA1}$$

where $P(\mathscr{N})$ is the set of all subsets of $\mathscr{N} = \{1, ..., N\}$ and $-\left[\frac{\partial(B/a)}{a}\right]^{\iota}$ is the submatrix obtained by eliminating any set $\iota \in P(\mathscr{N})$ of rows and corresponding columns, with strict inequality for minors of order 1, and

2. *non-positive cofactors*

$$C^{\iota}_{ji} = (-1)^{|\iota^-|}(-1)^{(i+j)} \det\left(-\left[\frac{\partial (B/a)}{a}\right]^{\iota}_{ji}\right) \leq 0 \text{ for } \iota \in \mathscr{P}(\mathscr{N}), \tag{SPA2}$$

of the submatrices obtained by eliminating the row and the column containing element j,i from principal submatrices $-\left[\frac{\partial(B/a)}{a}\right]^{\iota}$, where the cardinality $|\iota^-|$ of the set

$$\iota^- = \{k \in \iota : \min\langle i, j\rangle < k < \max\langle i, j\rangle\} \tag{24}$$

counts the number of diagonal elements with indices in ι that have been displaced off the diagonal once row j and column i are deleted from the original matrix.

Condition SPA is a multivariate generalization of condition (13) analyzed above for the case with $N = 1$ field. Condition SPA1 requires that the determinants of principal submatrices $-\mathbf{dp}^{\iota}$, obtained by eliminating any set $\iota \in P(\mathscr{N})$ of rows and corresponding columns, are non negative (or equivalently that $-\mathbf{dp}$ is a P_0 matrix; see Johnson, Smith, and Tsatsomeros 2020). Condition SPA2 imposes a restriction on the cofactors of the off-diagonal elements, C^{ι}_{ji}, which are the signed determinants of the submatrices obtained by eliminating the row and the column containing element j,i from principal submatrices $-\mathbf{dp}^{\iota}$.

The proof of Proposition 3.a relies on the properties of M-matrices. A matrix A is a non-singular M-matrix if it can be expressed in the form

$$A = sI - B$$

where B is a non-negative matrix and s is greater than the spectral radius (the maximum of the absolute values of the eigenvalues) of A. An M-matrix has non-positive off-diagonal elements and there is a large number of equivalent conditions for a matrix with non-positive off-diagonal elements to be an M-matrix. We will use the following two equivalences; see Chapter 6 of Berman and Plemmons (1994).

Theorem 2 *Let A be a real square matrix with non-positive off-diagonal elements. Then the following two statements are equivalent to A being an M-matrix:*

1. *All the eigenvalues of A have positive real parts.*
2. *A is inverse positive:* $A^{-1} \geq 0$.

Proof of Proposition 3.a. Fix any field i. By IHR, $A_i(a_i)/a_i$ is strictly increasing in a_i and, by the strict inequality of SPA1, $B_i(a_i, a_{-i})/a_i$ is strictly decreasing in a_i Hence, the general equilibrium is unique.

Next we show that the inverse of the Jacobian J^{-1} is an M-matrix. As the off-diagonal elements of an M-matrix are non-positive and the diagonal elements are positive, this yields the desired comparative statics. Furthermore, since M-matrices are closed under positive diagonal multiplication, for any positive diagonal matrix D, $J^{-1}D^{-1}$ is also an M-matrix and hence all its eigenvalues have positive real parts. It follows that all the eigenvalues of DJ have positive real parts, and therefore the equilibrium system is dynamically stable.

First, we show that the Jacobian is invertible and hence the equilibrium is generic and the comparative statics are well-defined. Computing the determinant of the Jacobian, we obtain

$$\det(J) = \sum_{|\iota|=0}^{N} \sum_{\iota \in \mathscr{P}_{|\iota|}(\mathscr{N})} \underbrace{\det\left(-\left[\frac{\partial (B/a)}{\partial a}\right]^{\iota}\right)}_{\geq 0} \prod_{i \in \iota} \underbrace{\frac{\partial (A_i/a_i)}{\partial a_i}}_{>0} > 0$$

where the first sign follows from SPA1 and the second sign from IHR. To see that the inequality is strict, observe that $-\frac{\partial B_i(a_1,..,a_n)/a_i}{\partial a_i} = \det(-\mathbf{d}(B/a)^{\iota}) > 0$ for $\iota = \mathscr{N} \setminus \{i\}$, by the strict inequality part of SPA1, hence the sum involves strictly positive addends.

Having established the inverse exists, to prove it is an M-matrix we show it has non-positive off-diagonal entries (i.e. it is a Z-matrix) whose inverse has nonnegative entries. To sign the off-diagonal elements, since the determinant of J is positive, the sign of the off-diagonal entries of J^{-1} is the sign of the off-diagonal cofactors of J. Denoting by C_{ji} the cofactor obtained by removing the jth row and ith column, we have

$$C_{ji} = \sum_{|\iota|=0}^{N-2} \sum_{\iota \in \mathscr{P}_{|\iota|}(\mathscr{N} \setminus \{i,j\})} \underbrace{\mathscr{C}_{ji}^{\iota}}_{\leq 0} \prod_{k \in \iota} \underbrace{\frac{\partial (A_k/a_k)}{\partial a_k}}_{>0} \leq 0,$$

where the first sign follows from SPA2 and the second from IHR. To show inverse-positivity, IHR implies that $\frac{\partial (A_i/a_i)}{\partial a_i} > 0$, and SPA2 together with the strict inequality part of SPA1 ensures that $\frac{\partial (B/a)}{\partial a}$ is non-positive and so $\frac{\partial (A/a)}{\partial a} - \frac{\partial (B/a)}{\partial a}$ must be non-negative. Therefore the inverse Jacobian J^{-1} is an M-matrix, and hence the equilibrium is dynamically stable with the desired comparative statics.

Proposition 4 *Quasi-proportional allocation QPA with $\rho_i \in [0,1]$ is sub-proportional SPA.*

Proof of Proposition 4. We verify SPA1 by computing the determinants of the principal minors

$$\det(-\mathbf{dp}^{\iota}) = B^{N-|\iota|} \frac{\prod_{i \in \mathscr{N} \setminus \iota} a_i^{\rho_i - 2}}{\left(\sum_{i \in \mathscr{N}} a_i\right)^{N+1-|\iota|}} \left[\sum_{i \in \mathscr{N} \setminus \iota} a_i^{\rho_i} \prod_{\substack{j \in \mathscr{N} \setminus \iota, \\ j \neq i}} \left(1 - \rho_j\right) + \prod_{i \in \mathscr{N} \setminus \iota} (1 - \rho_i) \sum_{k \in \iota} a_k^{\rho_k} \right],$$

which are non negative whenever $\rho_i \in [0,1]$ for $i \in \mathscr{N}$. To verify SPA2, the cofactors of the off-diagonal element i, j

$$\mathscr{C}_{ij}^{\iota} = -B^{N-|\iota|-1} \frac{\rho_i}{a_i a_j} \prod_{m \in \mathscr{N} \setminus \iota \setminus \{i,j\}} (1 - \rho_m) \frac{\prod_{m \in \mathscr{N} \setminus \iota} a_m^{\rho_m - 2}}{\left(\sum_{l \in \mathscr{N}} a_l^{\rho_l}\right)^{N-|\iota|}},$$

are non positive whenever $\rho_i \in [0,1]$ for $i \in \mathscr{N}$.

C Empirical Validation: Details and Robustness

C.1 RCN Applications Classification

We describe here the machine learning procedure we set up to assign RCN applications to ERC panels, based on the text analysis of the titles and the abstracts of the applications.

Table 2: Training Characteristics and Parameters

Fields	**LS**	**PE**	**SH**
Characteristics of training sets:			
Number of Panels	9	10	6
Number of Observations	3,643	4,909	2,409
Hyperparameters for training:			
Learning rate	2e-5	5e-5	5e-5
Number of epochs	5	3	3
Prediction accuracy:			
Validation accuracy (%)	74	83	83
Test accuracy (%)	72	80	81

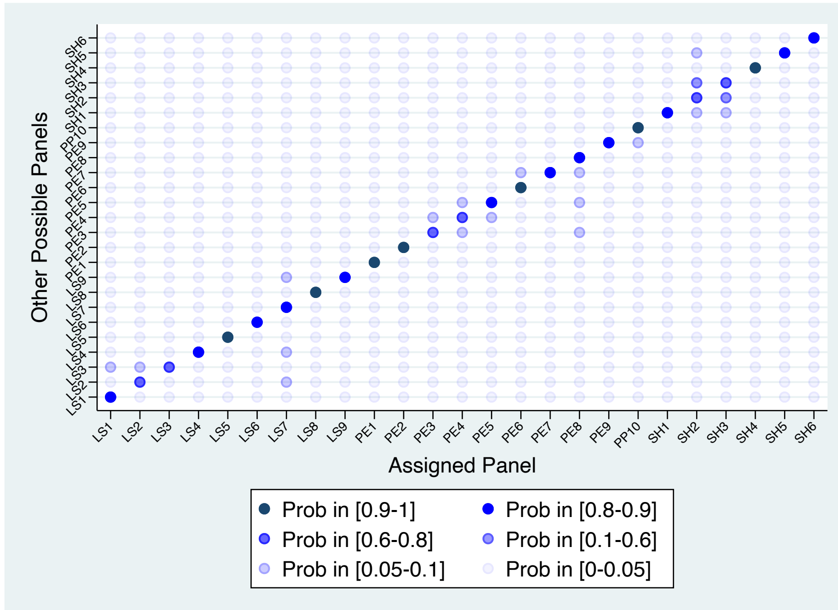
C.1.1 Datasets

Our data consists of three pairs of datasets, one for each of the ERC domains: Life Sciences (**LS**), Physical Sciences and Engineering (**PE**) and Social Sciences and Humanities (**SH**). Domains are further divided into panels: 9 in LS, 10 in PE, and 6 in SH, see Table 2. The challenge is to assign a panel to unlabeled abstracts. For each field, the training set has two columns: {Title + Abstract} and Label. We call the "prediction set" the dataset of unlabeled abstracts containing only the {Title + Abstract} column. The prediction is the output obtained by feeding this set to our trained model, as described below.

C.1.2 Training

We classified text using BERT (Biredictional Encoder Representations from Transformers), a state-of-the-art language representation model that enables us to achieve high accuracy. Following Devlin et al. (2018), we combine BERT with a neural network algorithm. More precisely, BERT is a pre-trained model in the sense that it has an underlying vocabulary that maps each text input into a single token sequence. In this sequence, each token is represented not only by its content, but also by its position in the sentence. The representation map is the by-product of prior pre-training by the Google team on a large corpus of texts. Then, BERT can be adapted to the specific task, a step called *fine-tuning*, with only one additional output layer. This step is straightforward and involves the choice of common hyperparameters for neural networks. For the learning rate, we obtain good results with hyperparameters in the recommended range in BERT's foundational paper. Our batch size is fixed at 8, constrained by our GPU power. The number of epochs, the number of complete passes through the training dataset, is also in the range of recommended values. We report the parameters chosen for each field in Table 2.

Figure 13: Predicted Probabilities of RCN Applications Assignment.



Notes: This figure shows the average probabilities for RCN applications of belonging to a particular ERC panel. The assigned ERC panel is based on the highest predicted probability. The probabilities sum to one over columns. In many cases, the probability of belonging to a particular panel is over 90 percent on average. *Source:* Own calculations combining RCN and ERC data.

C.1.3 Performance

We report two types of performance metrics: the validation and the test accuracy. The validation accuracy is obtained when all of the training set is used in the training process. We first split the data and we use the fraction p of observations in the neural network loop while the parameters are iteratively evaluated and adjusted to maximize the percentage of correct predictions on the $1-p$ remaining fraction. This value is the validation accuracy. In contrast, the test accuracy requires an additional earlier split. Only a fraction q of the data is used in the training process. The steps described above are therefore executed with the fractions qp and $q(1-p)$ of the training dataset. Test accuracy is then estimated on the remaining fraction $1-q$ of the dataset. Naturally, since the calculation of the test accuracy involves the prediction on a dataset never encountered by the neural network, test accuracy is generally smaller than validation accuracy. We report in Table 2 validation and test accuracy.

C.1.4 Predicted Panel

The outcome of the classification is a vector of probabilities for each RCN application, measuring the likelihood of belonging to a particular ERC panel. We assign each application based on the highest probability among all existing panels. The prediction is usually very sharp, in the sense that the algorithm most often picks one ERC panel with a very large probability. The highest probability of belonging to a particular ERC panel is equal to 85 percent on average, the next highest prediction is only equal to 6 percent on average. This means that few applications are marginally assigned to a panel. There is some heterogeneity across fields as depicted in Figure 13. While applications that are classified in mathematics (PE01) rarely share a similarity with applications in any other fields, it is less the case for applications that are assigned to SH2 (Institutions, Governance and Legal Systems) or SH3 (The Social World and its Diversity). This is also the case for applied life sciences such as LS07 (Prevention, Diagnosis and Treatment of Human Diseases) that shares similarities with more fundamental life science panels. Part of this variation may be due to measurement error, but it is also a fact that some applications are at the boundary between fields and that researchers in some disciplines have a choice of multiple panels in which they would stand a chance of winning. We deal with this uncertainty below by performing a bootstrap analysis.

C.2 Interrater Agreement

Once each RCN application is assigned to a particular ERC panel, we turn to the analysis of the RCN grading. On average, each application is reviewed by about 4.05 reviewers and each reviewer grades about 12.3 applications. Our data contains about 40,000 grades in total. Grading at the RCN is done on a scale of 1 to 7, although in practice 98 percent of the grades range between 3 and 7. The average grade is 5.05 and the standard deviation is equal to 1.1.

We use these data to construct inter-rating agreement measures for each ERC panel. We construct four different measures. The first one is the simple percent agreement between reviewers. Define n_{ij} as the number of applications graded by two raters assigning grades i and j. Denoting the total number of applications by n, the percentage agreement is calculated as

$$p_0 = \frac{\sum_{i=1} n_{ii}}{n}.$$

While simple, this measure tends to over-estimate agreement, as two graders may simply give the same grade by chance. There is a substantial literature on accounting for chance agreement, with several Kappa statistics such as Cohen's kappa or Fleiss' kappa, the latter taking explicitly into account the possibility of multiple reviewers. Gwet (2014) developed an Agreement Coefficient (AC) that incorporates both the number of rating categories and the frequency with which they are used by the raters. In the case of Cohen's kappa, chance agreement is computed as

$$p_e = \frac{1}{n^2} \sum_j \sum_i n_{ij} \sum_i n_{ji}$$

and the statistic is then

$$\kappa = \frac{p_0 - p_e}{1 - p_e}.$$

Table 3: Interrater Agreement by ERC Panels.

	Percent Agreement	Cohen's kappa	Fleiss' AC	Gwet's AC	Observations
	(1)	(2)	(3)	(4)	(5)
SH01	0.274***	0.04***	0.04***	0.17***	1643
	(0.011)	(0.014)	(0.014)	(0.013)	
SH02	0.274***	0.04***	0.04***	0.169***	9147
	(0.005)	(0.006)	(0.006)	(0.005)	
SH03	0.285***	0.048***	0.048***	0.182***	4900
	(0.006)	(0.008)	(0.008)	(0.007)	
SH04	0.288***	0.049***	0.047***	0.187***	1680
	(0.011)	(0.014)	(0.014)	(0.013)	
SH05	0.265***	0.04***	0.04***	0.158***	2378
	(0.009)	(0.011)	(0.011)	(0.01)	
SH06	0.261***	0.017	0.016	0.155***	1580
	(0.011)	(0.014)	(0.014)	(0.013)	
LS01	0.346***	0.064***	0.06***	0.26***	891
	(0.016)	(0.021)	(0.022)	(0.018)	
LS02	0.291***	0.046**	0.045**	0.191***	1051
	(0.014)	(0.018)	(0.018)	(0.016)	
LS03	0.315***	0.034	0.033	0.223***	569
	(0.019)	(0.026)	(0.026)	(0.022)	
LS04	0.292***	0.054***	0.053***	0.191***	2096
	(0.01)	(0.013)	(0.013)	(0.011)	
LS05	0.28***	0.019	0.019	0.179***	2023
	(0.01)	(0.013)	(0.013)	(0.012)	
LS06	0.355***	0.1***	0.1***	0.268***	1695
	(0.012)	(0.015)	(0.015)	(0.013)	
LS07	0.28***	0.044***	0.044***	0.177***	10246
	(0.004)	(0.006)	(0.006)	(0.005)	
LS08	0.295***	0.019	0.019	0.199***	2518
	(0.009)	(0.012)	(0.012)	(0.01)	
LS09	0.294***	0.023**	0.023**	0.198***	3225
	(0.008)	(0.011)	(0.011)	(0.009)	
PE01	0.406***	0.094***	0.093***	0.334***	950
	(0.016)	(0.023)	(0.023)	(0.018)	
PE02	0.361***	0.063***	0.062***	0.279***	1151
	(0.014)	(0.02)	(0.02)	(0.016)	
PE03	0.324***	0.06***	0.059***	0.232***	923
	(0.015)	(0.02)	(0.02)	(0.018)	
PE04	0.32***	0.056***	0.056***	0.227***	1103
	(0.014)	(0.019)	(0.019)	(0.016)	
PE05	0.295***	0.036**	0.035**	0.197***	1574
	(0.011)	(0.015)	(0.015)	(0.013)	
PE06	0.3***	0.061***	0.06***	0.2***	2003
	(0.01)	(0.013)	(0.013)	(0.012)	
PE07	0.287***	0.03**	0.029**	0.188***	1815
	(0.011)	(0.014)	(0.014)	(0.012)	
PE08	0.308***	0.054***	0.054***	0.211***	5401
	(0.006)	(0.008)	(0.008)	(0.007)	
PE09	0.413***	0.121***	0.12***	0.339***	681
	(0.019)	(0.027)	(0.027)	(0.022)	
PE10	0.315***	0.043***	0.043***	0.223***	7419
	(0.005)	(0.007)	(0.007)	(0.006)	

Note: The table displays inter-rating agreement measures based on grades of funding applications at the RCN.

Table 4: Effect of Evaluation Noise on Funding Outcomes: Bootstrap Analysis.

	Effect of evaluation noise	st dev	95% interval
Requested funding (β_a)			
Percent Agreement	.178	(.025)	[.128,.227]
Cohen's kappa	.31	(.039)	[.23,.386]
Fleiss' AC	.404	(.215)	[.004,.896]
Gwet's AC	.398	(.218)	[.001,.924]
Budget shares (β_B)			
Percent Agreement	.476	(.07)	[.33,.62]
Cohen's kappa	.582	(.381)	[-.117,1.467]
Fleiss' AC	.569	(.387)	[-.135,1.507]
Gwet's AC	.476	(.064)	[.344,.604]

Note: This table shows bootstrap estimates of the effect of inter-reliability agreement measures on the trend in funding after the reform. The effects are expressed in standard deviation effects. Each row reports the statistics computed from 500 replications. The estimation also controls for time, panel*seniority fixed effects and an aggregate time trend after the reform.

We refer the reader to Gwet (2014) for further details on other measures. Table 3 displays the results for all interrater agreement measures by ERC panel. In practice, the different measures are highly correlated, with correlation coefficients ranging from 0.84 to 0.99.

C.3 Robustness Using Bootstrap Analysis

To complement Section 5 we assess the robustness of the results to relax the assumption of assigning the RCN applications based on the most likely ERC panel. We exploit the fact that we have instead a vector of probabilities of belonging to any panel, for each application. We implement a bootstrap analysis based on multiple replications where we stochastically assign each grant application to a panel, based on the vector of assignment probabilities. For each replication, we re-calculate the interrater agreement measures and re-estimate model (18). We replicate this procedure 500 times. The results are displayed in Table 4. As in the main analysis, we find that the coefficient β_a is positive and statistically significant, across all the different statistics we use to measure evaluation noise.

D Supplementary Appendix: Impact of ARRA Budget Increase

As part of the stimulus bill introduced by the US Congress in 2009 in the aftermatch of the great financial crisis, the American Recovery and Reinvestment Act (ARRA) allocated an additional \$8.97 billion to extramural research grants at the NIH in two parts:

- Part of the funds (19.3%) of the total ARRA budget appropriated to the NIH were allocated to "not ARRA solicited" applications that had been previously submitted and reviewed in recent evaluation cycles, but were marginally rejected. Park, Lee, and Kim (2015) empirically document that "not ARRA solicited" applications resulted in less high-impact articles than regular projects.

- The remainder of the funding bonanza was set aside to increase the budget for "ARRA solicited" grant competitions. In this case, potential applicants were informed of the larger budget. A second fact, documented by Stephan (2012, p. 145), is that such applications increased so much that the success rate actually decreased.

These two facts can be rationalized within the framework we developed above. First, the budget allocated to "not ARRA solicited" applications corresponds to an unanticipated increase in the budget. Compared to the pre-policy equilibrium, the policy change shifts the supply to the right. Holding fixed the amount of applications $\bar{a} = a^B$ at the pre-policy equilibrium level for the initial budget B, the model predicts that the applications funded as a result of the increase in the budget to $B' > B$ are of lower quality

$$E\left[\theta|\theta \geq 1 - G(\bar{a}), x \in [\hat{x}^{B'}(\bar{a}), \hat{x}^{B}(\bar{a})]\right] < E\left[\theta|\theta \geq 1 - G(\bar{a}), x \geq \hat{x}^{B}(\bar{a})\right],$$

by the MLRP of the signal (2).

The impact of the anticipated increase in the budget on the success rate depends on the monotonicity of the hazard rate. By Proposition 2.d , applications must increase more than proportionally with the budget for the success rate to decrease as the budget increases—and this occurs in equilibrium if and only if the type distribution has decreasing hazard rate. This is exactly what happened as a result of the "ARRA solicited" part of budget increase in 2009. This observation is consistent with a type distribution with decreasing hazard rate at the top, as is natural to expect for the talent of scientists and artists; see Seglen (1992) and Caves (2000).[47]

[47] Distributions with decreasing hazard rate are more right skewed than the exponential distribution. They can be obtained stretching an exponential distribution toward the top tail through a convex transformation. A distribution has decreasing hazard rate whenever it is larger than the exponential distribution in van Zwet (1964) convex transform order. Given two distributions G and H, van Zwet (1964) defines G to be smaller than H in the convex transform order, denoted $G \prec_c H$, whenever $H^{-1}(G(\cdot))$ is convex. As shown by van Zwet (1964), a distribution G with increasing (or decreasing) hazard rate can be obtained through an increasing and concave (or convex) transformation $G^{-1}(G_{\text{Exp}}(\cdot))$ of a random variable with exponential distribution. To gain intuition, visualize the random variable G^{-1} on the vertical axis as an increasing transformation of an exponential random variable G^{-1}_{Exp} on the horizontal axis through a Q–Q plot. Concavity (or convexity) of $G^{-1}(G_{\text{Exp}}(\cdot))$ contracts (or stretches) the top tail and makes it thinner (or thicker) than the top tail of an exponential.