Productivity Beliefs and Efficiency in Science

Fabio Bertolotti
Bank of Italy

Kyle R. Myers
Harvard & NBER

Wei Yang Tham
University of Toronto

March 2025

We develop a method to estimate producers' productivity beliefs in settings where output quantities and input prices are unobservable, and we use it to evaluate allocative efficiency in the market for science. Our model of researchers' labor supply shows that their willingness to pay for their two key inputs, funding and time, reveals their underlying productivity beliefs. We estimate the model's parameters using data from a nationally representative survey of research-active professors from all major fields of science. We find that the distribution of research productivity is highly skewed. Using these estimates, we assess the market's allocative efficiency by comparing actual input allocations to optimal allocations given various objectives. Overall, the market for science is moderately efficient at maximizing output and researchers' utility: actual input levels are positively correlated with the optimal levels implied by the model. However, the wedge between researchers' actual and optimal input levels is often significant and difficult to predict. Our estimates imply that total budgets would need to increase by roughly 40% under actual allocations in order to achieve the same growth in scientific output that we predict under alternative allocations of the current budget. Scaling to the population level, this equates to billions of dollars in funding — there are substantial gains from developing new ways of identifying and supporting productive scientists.

Bertolotti: fabiobertolotti@hotmail.it. Myers: kmyers@hbs.edu. Tham: weiyang.tham@rotman.utoronto.ca.

This project received financial support from The Alfred P. Sloan Foundation and the Harvard Business School. There are no financial relationships or other potential conflicts of interest that apply to any of the authors. This project would not have been possible without the excellent work of Rachel Mural, Nina Cohodes, and Yilun Xu as well as input from Marie Thursby, Jerry Thursby, and Karim Lakhani. We have received helpful comments from Judy Chevalier, Matt Clancy, Shane Greenstein, Daniel Gross, Jorge Guzman, Bronwyn Hall, Bruce Kogut, Petra Moser, David Rivers, Tim Simcoe, as well as seminar participants at Columbia Business School, JOMT, Massachusetts Institute of Technology, the Workshop on the Organisation Economics and Policy of Scientific Research, the Banff Productivity Innovation and Economic Growth Conference, the NBER Productivity Lunch, Boston University TPRI, and the Conference on the Economics of Innovation in Memory of Zvi Griliches.

1. Introduction

Studies of industrial markets continue to find that a substantial amount of the differences in output across sectors and regions can be attributed to differences in allocative efficiency.[1] Yet, virtually no such analysis exists for the market for scientific research.[2] A long line of work has shown that frictions in science exist, whereby certain researchers are denied resources for reasons plausibly unrelated to their productivity (Merton 1973; Zuckerman 1988; Shapin 1995; Azoulay, Fons-Rosen, and Graff Zivin 2019; Hager, Schwarz, and Waldinger 2024). But we have yet to quantify how much these frictions are slowing down scientific progress. This is unfortunate given the large role that science plays in driving economic growth.[3]

Our limited view of efficiency in science stems from the challenges of estimating individual researchers' productivity. Most notably, the output of science, "*knowledge*" (Arrow 1962), is difficult to quantify. In the century since Lotka (1926) first used publication counts to proxy for units of knowledge, few advances have been made outside of bibliometric data. This is despite Adams and Griliches's (1998) note that, "*what constitutes a scientific paper makes for an elastic yardstick of scientific achievement*." In this paper, our goal is to avoid proxies altogether and, instead, estimate researchers' productivity without ever observing the quantity of knowledge they produce.

How might we estimate productivity in the absence of output data? The factor-share approach provides one way forward (i.e., De Loecker and Syverson 2021). It is assumed that producers are cost minimizers and face no adjustment costs when choosing their input levels. This implies a factor's input cost share reflects its output elasticity. However, the two scientific inputs that we are interested in, researchers' funding and their time spent on their research, are not allocated via price mechanisms and likely involve adjustment costs. Thus, the construction of factor shares and the mapping to productivity parameters is not straightforward.

To overcome these challenges, we develop a methodology for estimating productivity that does not require output quantities, does not require inputs to be explicitly priced, and directly incorporates adjustment costs. Our approach allows us to recover the distribution of researchers' productivity beliefs, which are very skewed, and quantify the value from more efficiently allocating inputs, which is very large. While we tailor our method to estimating productivity in academic science, the general approach provides a path forward for productivity estimation in settings with many heterogeneous producers (e.g., individual workers) who are obtaining inputs that are not explicitly priced and generating output that is difficult to observe.

[1] For example, see: Hsieh and Klenow (2009); Syverson (2011); Restuccia and Rogerson (2013); Hopenhayn (2014).

[2] The closest examples we are aware include Adams and Griliches (1998); Wang and Huang (2007); Rosenbloom et al. (2015), and Qiu (2023), but these evaluations of efficiency are quite indirect.

[3] For theoretical motivation, see Jones (2005); and for empirical work, see, for example: Jaffe (1989); Adams (1990); Mansfield (1998); Popp (2017); Valero and Van Reenen (2019); Hausman (2022); Andrews (2023).

The logic of our method is the same logic as the factor-share approach: producers' input demand reveals information about their productivity. However, the specific way in which we apply this logic is unconventional. The following simple example illustrates our method:

> Profit maximizing producers are indexed by i. Output (Y) is produced using a single variable input (X) per the production function: $Y_i = \alpha_i X_i$, where α_i is the focal productivity parameter. Producers are price-takers and face a common output price ($p > 0$). Input costs are heterogeneous and convex as given by: $X_i^{c_i}$, where $c_i > 1$.
>
> Consider the period before producers have obtained their inputs inside the market; call this the *pre-production* period. In the pre-production period, assume we can observe or estimate producers' expectations about the price of output (p). We cannot observe their productivity or cost parameters (α_i, c_i), and we can never observe output (Y_i) even after production occurs. Our goal is to estimate productivity (α_i).[4]
>
> Our solution is to solicit producers' willingness to pay (WTP_i) to obtain a fixed quantity of input in the pre-production period from outside the market (e.g., from us, the experimenters). Consider an experiment where producers are asked to report their WTP to obtain 1 unit of X in this way. Producers' WTP will equate their profits inside the market whether or not they purchase the 1 unit of X in the pre-production period. Formally, their WTP_i should make the following equality hold: $\max_{X_i} [p\alpha_i X_i - X_i^{c_i}] = \max_{X_i} [p\alpha_i (X_i + 1) - X_i^{c_i} - \text{WTP}_i]$.
>
> This yields a simple expression for identifying producers' productivity that does not require output quantities or input prices: $\alpha_i = \text{WTP}_i / p$.

The canonical setting of this example is one of manufacturing firms using labor and capital to produce physical goods that generate revenue via prices in the product market. But the underlying concept applies to science as well: researchers use their funding and time to produce knowledge that generates utility via the incentive structures of science. And just as a firm's willingness to pay for inputs reveals their belief about how productively they can generate physical goods, a researcher's willingness to pay for scientific inputs reveals their belief about how productively they can generate knowledge.

In order to apply this method to the market for science, we first develop a new model of researchers' production and consumption decisions. Researchers derive utility from three sources: their salary, their scientific output (i.e., the quantity of knowledge they produce), and their leisure time. Researchers choose how to allocate their time across fundraising (e.g., writing grants), research, and leisure given their beliefs, budgets, and constraints.[5] Each researcher has

[4] Note that, even if we could also observe or estimate producers' expected input choice, X_i^*, then the first order condition, $X_i^* = (p\alpha_i / c_i)^{1/(c_i - 1)}$, still does not separately identify the productivity and cost parameters (α_i, c_i).

[5] To capture heterogeneity in researchers' payoffs and preferences, we allow some functional forms to depend on

a unique production function with constant returns to scale, defined by two researcher-specific parameters: (*i*) a funding-intensity parameter that determines the relative weight of the two key inputs, funding and time, and (*ii*) a total factor productivity (TFP) parameter that describes the efficiency with which researchers produce scientific output using their inputs.

Using the model, we can write a researcher's willingness to trade off their salary for inputs as a function of their productivity beliefs. This trade-off mirrors real decisions researchers make in the job market, and it forms the basis of the experiments that provide the variation necessary to identify the model's parameters. By identifying researchers' input demand via their WTP, we can handle the fact that inputs are not explicitly priced. Moreover, since the survey also solicits researchers' actual input levels, we can still estimate researchers' scientific output despite the fact that we never need to observe the knowledge they expect to produce. Throughout the paper, we highlight several important limitations to our general approach and the specific model.

In order to generate the data necessary to estimate the model, we make use of a nationally representative survey of research-active professors across all major fields of science at roughly 150 major institutions of higher education in the US; for details, see Myers et al. (2023).[6] The survey solicits researchers' salaries, time allocations, and access to inputs. Importantly, the survey also includes a series of hypothetical experiments that solicit researchers' willingness to trade off their salary for more research funding or fewer administrative duties.

Researchers' willingness to pay for inputs are of plausible magnitudes and behavior. The median researcher is willing to pay 10 cents for $1 of research funding and $68 for one less hour of administrative duties. The components of the model explain a large fraction of the variation in researchers' responses (82–96%) and, for the most part, responses do not appear to reflect sample selection or systemic noise in respondents' willingness to pay for any hypothetical good. Furthermore, researchers' willingness to pay for free time is strongly correlated with their implied hourly wages as expected.[7]

Our estimates of productivity beliefs vary widely across researchers, even after accounting for outliers. Within major fields of study, the ratio of the 90^{th} and 10^{th} percentile of TFP is approximately 52. When we look in narrower fields of study and attempt to control for other sources of heterogeneity, we estimate 90–10 TFP ratios to be approximately 29. This represents a high degree of dispersion compared to what is observed in commercial markets at the firm-level (e.g., Syverson 2011); however, this scale of dispersion in productivity across individual workers has been observed in some high-skilled settings (e.g., Sackman, Erikson, and Grant 1968). As a

individuals' observable features.

[6] Evidence is provided in Myers et al. (2023) suggesting the presence of non-response bias in the sample is very low on observable dimensions such as institutional rank and researcher-level variables such as grant funding and publication rates; some of this is reprinted in the Appendix. The survey recruitment included randomized incentives and reminders, which we use to test for sample selection per Heckman (1979) and find no evidence for it.

[7] We also implement Dizon-Ross and Jayachandran's (2022) approach of using a "benchmark good" as a part of our willingness-to-pay elicitation to test whether respondents exhibit systematic noise in their stated preferences.

sign of face validity, our productivity estimates are positively correlated with common metrics of knowledge production (e.g., publications, citations, grant funding) and exhibits a similar degree of dispersion as those metrics. Notably, our estimates indicate that approximately half of the variance in scientific output across individual researchers is due to variance in their productivity.[8] Together, these results suggest that it may be hard for the market for science to facilitate positive selection on productivity.

To evaluate allocative efficiency, we compare inputs and outputs under actual allocations to those under alternative allocations. Specifically, we consider three alternative objectives: (*i*) maximize total scientific output, (*ii*) maximize researchers' private utility, or (*iii*) maximize social welfare given an assumption about the externalities of science. Using the model, we can solve for the allocation of inputs that achieves an objective while allowing for researchers' behavioral responses as they re-optimize their choices.[9]

Overall, we find evidence of a moderate degree of efficiency given our proposed objectives. The correlations between researchers' actual input levels and the optima implied by our model generally span 0.4–0.8; more productive researchers acquire more inputs on average. However, are still significant gains from alternative allocations. Our counterfactuals suggest that total annual scientific output could be increased by approximately 160%. The private value to researchers of this additional output is on the scale of 5%. Estimating the social value of this growth requires assumptions about externalities, which we explore.

To provide another way of characterizing the gains from reallocation, we ask the following: how much would funding levels need to increase under actual allocations to produce the same growth in output as our alternative allocations that hold input levels fixed? We find that total annual funding levels would need to increase roughly 40% in order to achieve the same 160% growth in output our alternative allocations can achieve. Even conservative approaches to scaling these estimates up to the size of the population can imply gains from reallocation that are equivalent to multi-billion dollar increases in annual research funding.

We also evaluate the degree to which differences in aggregate output across major fields of study are due to differences in the number of researchers, their productivities, their input levels, or the fields' allocative efficiency. Overall, differences in allocative efficiency are the largest determinant of differences in aggregate output. At efficient allocations, the gaps in output between fields shrink by 20–50%.

Lastly, we unpack the counterfactuals to explore the following questions: Is the efficient allocation of inputs implied by the model more or less concentrated than the actual allocation?

[8] We are unable to distinguish the degree to which the "productivity" we estimate reflects an individual's fixed capabilities as a researcher versus any cumulative advantage they have acquired (c.f., Hall and Mairesse 2024).

[9] We use a number of constraints in order to keep our counterfactuals within the support of the data we observe and plausible in nature.

How does the reallocation of each input (i.e., funding and time) independently change the results? How much does the efficient allocation change under different objectives? What is the distribution of input wedges (i.e., the difference between their optimal and actual input levels) across researchers? Are researchers' input wedges predictable given their observable features?

Our estimates come with many caveats due to the limitations of our method and setting. First, we face a set of challenges common to all existing productivity estimation techniques: (*i*) noise due to mis-measurement or mis-specification will lead us to overstate the degree of misallocation; (*ii*) the presence of unmeasured inputs will inflate the productivity estimates; and (*iii*) the presence of unmodeled heterogeneity in output prices (or preferences) will distort the productivity estimates. Second, we face a set of limitations unique to our methodology and data: (*i*) our method provides only ex-ante productivity beliefs and has no way of identifying ex-post differences in production; and (*ii*) the identifying variation in our data is based on stated preferences from hypothetical experiments, which may suffer from a range of biases. Throughout the paper we engage with these limitations by providing multiple robustness tests of our assumptions and face validity tests of the data. Overall, the results from these tests give us confidence in our approach. Still, we interpret our results as plausible upper bounds on the productivity dispersion and gains from reallocation in science. Given the lack of quantitative evidence on these points, we view this work as an important step towards understanding resource allocation in science.

The structure of the paper is as follows: Section 2 details the survey data; Section 3 describes our model of researchers' production and consumption; Section 4 describes the survey experiments where we ask scientists to trade off their salary for inputs and reports the results of those experiments; Section 5 provides the productivity estimates; Section 6 contains our counterfactual allocation exercises and explorations thereof; and Section 7 concludes with a discussion about our results specifically and the potential usefulness of our methodology more generally.

2. Survey and Data Overview

2.1. Survey Design

We use the National Survey of Academic Researchers (Myers et al. 2023) and provide below a brief overview of the survey methodology here. The population target is US professors who conduct research at major institutions of higher education. To construct the sampling frame, information on professors was collected from the 158 largest institutions in the US by total R&D funding using the National Science Foundation's 2019 Higher Education R&D survey (HERD; National Science Foundation 2023).

The population consisted of 264,036 unique e-mails. A total of 131,672 individuals were e-mailed

and 4,388 (3.33%) completed the survey.[10] We then restrict the sample to the 4,003 individuals (91.2% of respondents) who reported being a professor, spending a non-zero amount of time on research, and having a non-zero salary from their primary institution.

During recruitment, incentives and reminders were randomly assigned. The four incentive arms were: (*i*) no incentive, (*ii*) entry into a lottery to win a gift card, (*iii*) the ability to vote for a set of charities to receive a donation, and (*iv*) both the second and third incentives. The reminder arms were zero, one, or two reminders. Each email was randomly assigned to one incentive arm and one reminder arm with equal probability, resulting in twelve possible combinations.

The randomized incentives and reminders provide us with instruments that we can use to implement a sample selection correction (i.e., Heckman 1979). The validity of this approach relies on having variables that cause entry into the sample (i.e., completing the survey) but do not affect the outcomes of interest. This allows us to adjust for unobservable differences between the population and our sample. Appendix Table A1 reports the results from a regression of an indicator for survey completion on the different incentive and reminder arms, and it shows that all arms had a statistically significant positive effect on researchers' propensity to complete the survey.

In addition to adjusting for unobservable differences, Myers et al. (2023) also check the representativeness of the respondent sample by comparing it to the invited sample on observable characteristics. First, the authors explore a series of observable characteristics at the researcher-level by comparing the grant and publication histories of respondents and non-respondents using the Dimensions database (Digital Science 2018), which collects and disambiguates scientific metrics for researchers worldwide.[11] Appendix Figure A1 (replicated from Myers et al. (2023)) shows invite-respondent overlap on various measures of scientific inputs and outputs. Overall, there is little difference in the respondents and non-respondents both economically and statistically speaking.

Looking at the institutional level, Appendix Figure A2 (replicated from Myers et al. (2023)) shows invite-respondent overlap on various measures of institution funding derived from the HERD survey. As in the case of the researcher-level comparison, the distributions overlap substantially. In this case, there is some statistically significant difference; on average, respondents come from institutions with slightly less research funding (4–6%).

[10] The IRB approval permitted e-mailing only 50% of the population. The response rate is more than twice what has been obtained from sourcing academic researcher contacts from the corresponding author data contained within the publication record (e.g., Myers et al. 2020).

[11] Using a fuzzy name matching process, Myers et al. (2023) are able to confidently match 87,000 (66%) of the researchers to their records in Dimensions.

2.2. Summary Statistics: Researchers and their Inputs

Table 1 reports the summary statistics of the key covariates used in our analyses.

Fields of Study. Using the description of the professor's department, we assign them first to a narrow set of twenty "minor" fields of study and then aggregate those fields into five broader "major" fields: humanities and related; engineering, math, and related; medicine and medical sciences; natural sciences; and social sciences. Appendix Table A2 reports the distribution across the twenty minor fields. Unless otherwise noted, our counterfactual analyses will constrain the reallocation of inputs to only occur within the major fields.

Salaries and Research Inputs. The survey solicits a range of details regarding researchers' salary, their guaranteed funding (e.g., from prior grants or institutional guarantees), expected funds they will raise over the coming five years, and their time allocations. Importantly, most variables are elicited as expectations over the coming five years to ensure the responses span the same time horizon as the thought experiments described below. In the Appendix, we replicate a test of respondents' self-reporting by comparing their self-reported salaries to the publicly-reported salaries we are able to locate for a subset of researchers. Overall, there is a high degree of alignment (see Appendix Figure A3). Appendix Figure A4 plots full distributions of work time across each task, and Appendix Figure A5 plots full distributions of researching funding across the major fields for further detail. While there certainly is more research funding in traditionally capital-intensive fields (e.g., the natural sciences), there is a significant proportion of researchers in each major field with substantial research funding levels.

Position and Socio-demographics. Appendix Tables A3–A4 provides a full summary of all other major features collected regarding researchers' positions (e.g., rank and tenure status) and their socio-demographics (e.g., gender, race/ethnicity, citizenship). See Myers et al. (2023) for a more detailed investigation of these summary statistics.

2.3. Summary Statistics: Subjective Output Measures

Our methodology for estimating productivity is explicitly designed to avoid the need to quantify the output produced by researchers' efforts. Still, it is useful to have a more qualitative understanding of what researchers are producing. Fortunately, the survey includes a number of subjective questions regarding researchers' output. Table 2 summarizes answers to these questions, which asked the frequency with which researchers intended to produce outputs of certain types (e.g., articles, books, materials, products) for certain types of audiences (e.g., peers, policymakers, businesses, general public) as well as other measures of riskiness and the degree to which the researcher focused on asking new questions or answering existing questions.

The traditional proxy for scientific output is peer-reviewed journal articles, and the data indicate that this is the most common output type that researchers intend to produce. However, there is a considerable amount of variation and a significant amount of attention focused on producing output of types and for audiences that may never be codified in a journal article. More importantly, these measures are often negatively correlated in a way that suggests strategic substitution (see Appendix Table A5). For instance, researchers who focus more on publishing articles for their academic peers are less likely to focus on publishing books or developing products intended for non-academic audiences. This highlights the limitation of observable output proxies.

3. Model of Science

In this section, we model researchers' labor supply, which describes their utility from production and consumption. First, we present the environment and researchers' optimal decisions (Subsection 3.1). Then we incorporate externalities in production to define social welfare (Subsection 3.2) and highlight the connection between the model and the survey thought experiments that solicit researchers' willingness to pay (WTP) for different factors (Subsection 3.3). Lastly, we outline our estimation approach (Subsection 3.4). Additional details regarding the model and estimation are contained in Appendix B.

3.1. Researchers' Labor Supply

There are N researchers indexed by i who each choose how much total time to work (H_i) and how to allocate their time between research R_i and fundraising F_i over a fixed horizon. Their choices maximize their utility conditional on a contract from their primary institution, which is a triplet of state variables $\mathbf{S}_i = (M_i, G_i, D_i)$: salary M_i (\$), guaranteed funding G_i (\$), and administrative and teaching duties D_i (hours).[12] Researchers' indirect utility is given by:

$$\mathcal{V}(\mathbf{S}_i\,,\,\theta_i\,,\,\mu_i) = \max_{R_i\,,F_i\,,H_i} U_{1i}(M_i) + U_{2i}(Y_i) - U_{3i}(R_i\,,\,F_i\,,\,D_i) \tag{1a}$$

subject to

$$B_i = B_{\min} + G_i + \phi_i F_i \tag{1b}$$

$$R_i + F_i + D_i = H_i \tag{1c}$$

$$Y_i = \alpha_i B_i^{\gamma_i} R_i^{1-\gamma_i}\,, \tag{1d}$$

where $\theta_i = (\alpha_i\,,\,\gamma_i\,,\,\phi_i)$ is the vector of individual-specific attributes related to scientific activity and μ_i is a vector of parameters that govern the shape of the U functions.

[12] We use the term "*contract*" loosely, since, for example, a researcher's guaranteed funding may come in part from future flows of funding from grants obtained outside the institution.

The quantity of knowledge Y_i that researchers produce is a Cobb-Douglas function of total funding B_i and research time R_i, which includes researcher-specific productivities (α_i) and output elasticities (per γ_i). Additionally, we assume a minimum funding amount $B_{\min}$, and that all work hours are non-negative and have an upper bound $H_{\max}$.

Utility from salary, output, and effort are given by the following functional forms:

$$U_{1i}(\cdot) = \omega \frac{(M_i)^{1-\sigma_i}}{1-\sigma_i} \tag{2a}$$

$$U_{2i}(\cdot) = \frac{(Y_i)^{1-\eta_i}}{1-\eta_i} \tag{2b}$$

$$U_{3i}(\cdot) = \psi \frac{(R_i + F_i + D_i^{\xi_i})^{1+\zeta_i}}{1+\zeta_i} \,. \tag{2c}$$

The vector $\boldsymbol{\mu}_i$ collects the parameters $(\omega, \sigma_i, \eta_i, \psi, \xi_i, \zeta_i)$.[13]

The policy functions $\mathcal{R}(\mathbf{S}_i, \boldsymbol{\theta}_i, \boldsymbol{\mu}_i)$, $\mathcal{F}(\cdot)$, and $\mathcal{H}(\cdot)$ characterize the solutions to Equation (1). For each individual researcher, these policies determine optimal (R_i^*, F_i^*, H_i^*) as a function of states, attributes, and parameters. The derivations of these policy functions are described in Appendix B.1.

Ideally, we would have enough variation to estimate all researcher-specific attributes including the production-related parameters $(\alpha_i, \gamma_i, \phi_i)$ as well as all of the consumption-related parameters that govern researchers' utility from salary, output, and effort $(\sigma_i, \eta_i, \xi_i, \zeta_i)$. However, as shown below, we only have enough variation in the survey experiments to uniquely identify the production-related parameters $(\alpha_i, \gamma_i, \phi_i)$. Thus, we choose to specify each of the consumption-related parameters to be a parametric function of researchers' observable features.

Fortunately, we have a large vector of observable features ($\mathbf{X}_i$) that describe researchers' positions, their backgrounds, and their subjective descriptions of their scientific output (i.e., the features summarized in Table 2 and Appendix Tables A3–A4). This is useful because it allows us to incorporate more heterogeneity into the consumption components of the model and limit the degree to which variation in the data might otherwise cause us to overstate the heterogeneity in the production parameters.

Unfortunately, the computational demands of estimating the model limit the flexibility with which we can incorporate the dozens of features available. Thus, to balance the benefits of allowing for heterogeneity in consumption with the benefits of simpler estimation, we use k-means clustering to reduce the full set of observable features ($\mathbf{X}_i$) into a one-dimensional index. We assume there are two clusters of researcher types ($k = 2$) and estimate each researcher's

[13] The ξ_i parameter allows for additional disutility from duty-related work (e.g., administration or teaching), which improves the model's fit to the data. We also assume the following: $\eta_i \in (0, 1)$, $\psi > 0$, $\xi_i > 0$, and $\zeta_i > 0$.

distance from these clusters per their Euclidean similarity score. This distance provides a one-dimensional description of all of the different ways researchers responded to questions that could plausibly be driven by heterogeneous preferences.

We refer to this resulting index as describing researchers' type, T_i, Appendix B.2 reports the results of the k-means estimation showing the distribution of researcher type T in the sample as well as a view of the mean differences in the features of researchers per their type.

We then specify each of the consumption-related parameters $(\sigma_i, \eta_i, \xi_i, \zeta_i)$ to be simple, univariate functions of researcher's type. For instance, the parameter that governs the utility from private consumption is parameterized as: $\sigma_i = \exp(\delta_0 + T_i\delta_1)$. We similarly parameterize ζ_i and ξ_i with exponential functions with intercept and slope, respectively. The curvature of utility in scientific output, η_i, is modeled with a logistic function bounded in the interval $(0, 1)$. Therefore, we express individual-specific parameters $\boldsymbol{\mu}_i(T_i, \tilde{\boldsymbol{\mu}})$ as a function of researcher's type T_i and of the vector of common parameters to be estimated, denoted by $\tilde{\boldsymbol{\mu}} = (\omega, \psi, \boldsymbol{\delta})$, where $\boldsymbol{\delta}$ includes the deep parameters that govern the utility functions.

Allowing the $\boldsymbol{\mu}_i$ parameters to be type-specific is not a panacea, but it helps reduce the degree to which variation in the experimental data that is truly driven by heterogeneous preferences or heterogeneous demand for scientific output contaminates our productivity estimates. We further detail the estimation process below.

3.2. Social Value

To incorporate the externalities of knowledge production (e.g., Nelson 1959; Arrow 1962), which we assume to be net positive, we define the social value produced by each researcher as:

$$\mathcal{W}(\mathbf{S}_i\,,\,\theta_i\,,\,\boldsymbol{\mu}_i\,,\,\kappa) = U_{1i}(M_i) + \kappa U_{2i}(Y_i^*) - U_{3i}(R_i^*\,,\,F_i^*\,,\,D_i)\,, \tag{3}$$

where researchers' privately optimal choices and output are given by R_i^*, F_i^* and Y_i^*. κ reflects the size of the externalities associated with researchers' output. Thus, $1/\kappa$ gives the share of the social value generated by each researchers' output that they themselves capture and $1 - 1/\kappa$ is the size of the positive externality.

This formulation of externalities is an ad hoc way of implicitly modeling consumer surplus as some constant multiple of producer surplus. However, it has a clear economic interpretation as a measure of appropriation, and we can draw on prior studies that have sought to identify precisely this measure. Studies that focus on commercial innovators have found producers' value capture to be as large as 15% ($\kappa = 6.7$) and as low as 2% ($\kappa = 50$) (Nordhaus 2004; Philipson and Jena 2006; Jena and Philipson 2008; Lakdawalla et al. 2010). In our empirical analyses, our baseline assumption is $\kappa = 10$, which implies researchers capture 10% of the social value

they create.[14] Interestingly, as we will show below, the median researcher is willing to pay approximately 0.1 dollars to purchase 1 dollar of research funding, which is a magnitude that is consistent with our assumption that $\kappa = 10$.

3.3. Input Demand and Productivity

This subsection illustrates how our model can generate predictions about a researcher's willingness to pay (or accept) for more (or less) inputs than what existing allocation schemes would lead them to obtain.

Consider the period prior to production but after which researchers have formed their expectations about their time allocations and, therefore, their expectations about their utility over the 5-year horizon. Now, researchers are offered some alternative contracts by their primary institution that vary either funding guarantees ($G \to \widetilde{G}$) or administrative duties ($D \to \widetilde{D}$), but leave the salary unspecified.

For each offer indexed by j and characterized by $(\widetilde{G}_{ij}, \widetilde{D}_{ij})$, there is a salary $(\widehat{M}_{ij})$ that makes the researcher indifferent between their actual contract (M_i, G_i, D_i) and the offer such that:

$$V_i \equiv \mathcal{V}(M_i, G_i, D_i, \theta_i, ...) = \mathcal{V}(\widehat{M}_{ij}, \widetilde{G}_{ij}, \widetilde{D}_{ij}, \theta_i, ...) \,. \tag{4}$$

Importantly, $\mathcal{V}(\cdot)$ incorporates behavioral responses such that:

$$\underbrace{\mathcal{R}(\mathbf{S}_i, \theta_i, ...)}_{\substack{\text{expected research} \\ \text{time in actual} \\ \text{position}}} \gtrless \underbrace{\mathcal{R}(\widetilde{\mathbf{S}}_{ij}, \theta_i, ...)}_{\substack{\text{expected research} \\ \text{time in alternative} \\ \text{position}}}$$

where and likewise for $\mathcal{F}(\cdot)$ and $\mathcal{H}(\cdot)$; researchers evaluate their net indirect utility in the alternative position after they will have re-optimized their time allocations given the new state variables.[15]

For example, if a researcher is offered more guaranteed funding ($G_i < \widetilde{G}_{ij}$) then they will forecast how this additional funding will lead to changes in their optimal time allocations, their expected input levels, their expected output levels (per their productivity beliefs), and their expected indirect utility. The total increase in their expected indirect utility can be priced by the researcher and stated as a new, lower salary ($M_i > \widehat{M}_{ij}$).

Importantly, we can write that alternative salary $(\widehat{M}_{ij})$ that makes researchers indifferent between their actual position and the offer to be some function of known variables and the

[14] Note also that this approach implies that researchers' private utility reflects the case where $\kappa = 1$.

[15] Or put alternatively, $\widetilde{R}^*_{ij} \gtrless R^*_i$, $\widetilde{F}^*_{ij} \gtrless F^*_i$, and $\widetilde{H}^*_{ij} \gtrless H^*_i$.

unknown parameters to be estimated:

$$\widehat{M}_{ij} = \Pi(\widetilde{G}_{ij}, \widetilde{D}_{ij}, M_i, G_i, D_i, \theta_i, ...) , \tag{5}$$

where Π is determined by functional forms of, and optimality conditions implied by, Equation (1). Thus, if we know the left-hand side of Equation (5) from some experimental offers, and we also know all components of Π except for the parameters, then we can estimate those parameters.

3.4. Estimation

We infer individual attributes and parameters jointly, using a mix of calibration and estimation. First, we set minimum funding $B_{min} = \$5,000$ and fix total hours endowment H_{max} to 62 work hours per week (approximately the 90th percentile of the observed work hours). Second, we estimate individual attributes and common parameters of the utility function, including deep parameters that determine individual-specific parameters in μ_i as functions of researcher's type T_i. Specifically, we use survey information on declared hours worked, research time, fundraising time, guaranteed funding, expected additional funds raised, duties, and the salaries reported in the alternative offer experiments, which we map to model counterparts $(H_i, R_i, F_i, G_i, \phi_i F_i, D_i, \{M_{ij}\}_{j=1}^4)$, respectively.

Appendix B contains the full details on our estimation routine, which we briefly outline here. Conditional on some common parameters, we can infer individual-specific attributes θ_i from researchers' optimality conditions and model's structure. Therefore, we first describe this mapping, and then we proceed to present estimation of the common parameters. One important note is that we have separate estimation strategies for inferring individual attributes θ_i among researchers with non-zero fundraising time ($F_i > 0$) and those at a corner solution ($F_i = 0$).

For the former group, we first infer fundraising ability ϕ_i by exploiting the identity equating the observed, expected additional funding (EG_i) to $\phi_i F_i$. Therefore, $\forall i = 1, ..., N_F$:

$$\widehat{\phi}_i = \frac{EG_i}{F_i} . \tag{6}$$

Expression (6) is well-defined if and only if observed $F_i > 0$, which motivates why the inference strategy must differ for researchers with $F_i = 0$ at observed allocations.[16] As a second step, conditional on $\widehat{\phi}_i$, we infer factor shares $\widehat{\gamma}_i$ from researchers' first order condition in fundraising time, which we derive in Appendix B.1. For the chosen functional forms, this takes the intuitive

[16] For numerical stability, we constrain $\widehat{\phi}_i \geq 1$.

expression:

$$\widehat{\gamma}_i = \frac{B_i}{B_i + \phi_i R_i} = \frac{B_{\min} + G_i + \widehat{\phi}_i F_i}{B_{\min} + G_i + \widehat{\phi}_i F_i + \widehat{\phi}_i R_i} \,. \tag{7}$$

Equation (7) states that γ_i constitutes the weight of total funding over the total dollar-value of inputs used in scientific activity, with the last term of the denominator being the dollar-valued opportunity cost of research time relative to fundraising.

Finally, the optimality condition for total hours worked determines productivity $\widehat{\alpha}_i$ as a function of parameters, observed allocations, and $\widehat{\phi}_i$ and $\widehat{\gamma}_i$:

$$\eta_i(1-\widehat{\gamma}_i)\widehat{\alpha}_i^{1-\eta_i}(B_{\min} + G_i + \widehat{\phi}_i F_i)^{(1-\eta_i)\widehat{\gamma}_i} R_i^{(1-\eta_i)(1-\widehat{\gamma}_i)-1} = \psi(H_i - D_i + D_i^{\xi_i})^{\zeta_i} \,, \tag{8}$$

This equation holds exactly if $H_i < H_{\max}$. Therefore, Equations (6), (7), and (8) determine individual attributes as functions of parameters and observed allocations for researchers with positive fundraising time.

Unfortunately, for the group of researchers reporting zero fundraising time, we can neither infer $\widehat{\phi}_i$ from Equation (6), nor can we compute $\widehat{\gamma}_i$.[17] Therefore, we assume that ϕ_i and γ_i are parametric polynomial functions of state variables, and we estimate these functions using the other sub-sample of researchers with $F_i > 0$.[18]

Finally, given the estimates attributes $\widehat{\phi}_i$ and $\widehat{\gamma}_i$, the state variables, and the vector of common parameters $\tilde{\mu}$, we infer productivity beliefs $\widehat{\alpha}_i$ through Equation (8).

Given estimates of individual attributes $\widehat{\theta}_i = (\widehat{\alpha}_i, \widehat{\gamma}_i, \widehat{\phi}_i)$ as functions of parameters $\tilde{\mu} = (\omega, \psi, \delta)$, calibrated values $(B_{\min}, H_{\max})$, and observed allocations, we estimate the vector $\tilde{\mu}$ by generalized method of moments as:

$$\widehat{\tilde{\mu}} = \arg\min_{\tilde{\mu}} \mathcal{L}(\tilde{\mu}) = \arg\min_{\tilde{\mu}} \sum_{i=1}^{N} \sum_{j=1}^{4} \left(M_{ij} - \widehat{M}_{ij}(\mathbf{S}_{ij}, \widehat{\theta}_i(\mu_i(\tilde{\mu}, T_i)), \mu_i(\tilde{\mu}, T_i)) \right)^2 \,, \tag{9}$$

where M_{ij} is researcher i's answer to thought experiment j in the data and $\widehat{M}_{ij}$ is its model-based counterpart, which is itself an implicit function of the vector of counterfactual states $\mathbf{S}_i$, researcher's type T_i, and of the estimand $\tilde{\mu}$ which determines individual attributes $\widehat{\theta}_i$ and

[17] The Lagrange multiplier $\lambda_{F,i}$ is strictly positive and unknown.

[18] We use the following specifications: $\phi_i = \exp\left\{\sum_{p=1}^{3} \beta_{G,p} G_i^p + \sum_{p=1}^{3} \beta_{D,p} D_i^p + \sum_{p=1}^{3} \beta_{M,p} M_i^p\right\}$, and $\gamma_i = \left(1 + \exp\left\{\sum_{p=1}^{3} \iota_{G,p} G_i^p + \sum_{p=1}^{3} \iota_{D,p} D_i^p + \sum_{p=1}^{3} \iota_{M,p} M_i^p\right\}\right)^{-1}$, which we estimate using poisson and logistic regressions. Whenever the fitted $(\widehat{\phi}_i, \widehat{\gamma}_i)$ combination would imply, given the observed state variables and for a specific vector of parameters μ_i, that the optimal (observed) time allocation to fundraising would be strictly positive, we re-scale $\widehat{\phi}_i$ below the individual-specific lower bound below which the optimal fundraising time at observed states is indeed null.

individual-specific parameters μ_i.[19] Parameter estimates $\widehat{\mu}$ solve (9) conditional on parameter restrictions specified in Section 3.1. Appendix B.3 provides additional details on the search algorithm.

4. Survey Experiment

Here, we describe the thought experiments in the survey (Subsection 4.1) and how they connect to the model (Subsection 4.2). We then report the distribution of responses and conduct tests for face validity, sample selection, and other potential concerns (Subsection 4.3).

4.1. Soliciting Willingness to Pay

WTP for Funding and Time. Respondents are presented with four hypothetical scenarios, each offering different trade-offs between salary and inputs. They are asked to imagine that their primary institution has offered them: (*i*) an increase of $250,000 in guaranteed funding in exchange for a lower salary, (*ii*) an increase of $1,000,000 in guaranteed funding in exchange for a lower salary, (*iii*) an elimination of all administrative duties in exchange for a lower salary, and (*iv*) an increase in duties by 20 hours per month over a 5-year period in exchange for a higher salary.[20] All of these hypotheticals are posed over a 5-year span to fix the time horizon for all respondents. In order to solicit the salary at which they are indifferent between the current position and each offer, respondents are asked to report the lowest offered salary at which they would be willing to accept the offer.[21] Importantly, the survey only solicits researchers' WTP for more funding. This matters because counterfactual reallocations will involve reducing some researchers' funding. In the case of time, the survey solicits both WTP (for more free time) and willingness to accept (WTA; for less free time) and we treat these as symmetric (after filtered through the model of the production function). Appendix Figure C1 displays examples of how these thought experiments appeared to the survey respondents.

WTP for a Benchmark Good. When soliciting willingness to pay, especially in an un-incentivized manner, it is always possible that respondents under- or overstate their price sensitivity. In order to test for this, we follow Dizon-Ross and Jayachandran (2022) and explore respondents' willingness to pay for a "benchmark good." Dizon-Ross and Jayachandran (2022) note that, if one solicits a subject's willingness to pay for a good whose value (to the subject) is plausibly uncorrelated with the value of the focal good, then any correlation between the two willingness-to-pay values can be attributed to systematic noise. The survey asks respondents to report the

[19] To save notation, we omit that $\widehat{M}_{ij}$ are also a function of calibrated $(B_{\min}, H_{\max})$.

[20] Researchers who report no administrative duties are not shown scenario (*iii*).

[21] Pilot tests indicated that researchers could more easily report the lowest salary that could make them take the offer as opposed to the literal salary at indifference, and since indifference is a vanishingly small amount less than this reported value, we treat their answer as the amount at indifference.

maximum amount they would be willing to pay per month for high-speed internet access at their primary residence. We use this as our benchmark good since it is plausibly uncorrelated with productivity but of non-zero value to most respondents. In Appendix Table C1, we show that respondents' stated WTP for research funding and free time are rarely correlated with their stated WTP, which suggests a small amount of variation in researchers' answers is attributable to systematic noise.

4.2. Connection to Model: WTP and Compensating Variation

Four survey thought experiments elicit the compensating variation of individual researchers in relation to (*i*) an increase of \$250,000 in guaranteed funding G_i, (*ii*) an increase of \$1,000,000 in guaranteed funding G_i, (*iii*) a reduction of duties D_i to 0, and (*iv*) an increase in duties by 20 hours per month over a 5-year period. In other words, we ask for income levels $\widehat{M}_{ij}$, where $j = \{1, 2, 3, 4\}$ indexes the four experiments, that would make the researcher's utility in the counterfactual scenario equal to their indirect utility V_i^* at current allocations. Formally, counterfactual guaranteed funding and duties in the four experiments are:

$$
\begin{aligned}
(\widetilde{G}_{i1}, \widetilde{D}_{i1}) &= \left((G_i + \$250{,}000), D_i\right) \\
(\widetilde{G}_{i2}, \widetilde{D}_{i2}) &= \left((G_i + \$1{,}000{,}000), D_i\right) \\
(\widetilde{G}_{i3}, \widetilde{D}_{i3}) &= \left(G_i, 0\right) \\
(\widetilde{G}_{i4}, \widetilde{D}_{i4}) &= \left(G_i, (D_i + 20 \text{ hours/month})\right),
\end{aligned}
$$

where G_i and D_i represent actual states.

As noted above, the values of $\widehat{M}_{ij}$ reported in these four experiments make the researcher indifferent between all possibilities (per Eq. 4), and therefore we can write these reported values to be a known function of observable data and the unknown parameters to be estimated (per Eq. 5).

4.3. Survey Experiment Results

Figure 1 plots the distribution of WTP for guaranteed research funding and free time (in the form of WTP for less duties or WTA for more duties). The values shown are averaged over the two thought experiments for either factor and converted to a per-dollar basis. As evidenced by Figure 1, there is considerable variation in WTP responses across researchers. Some of this variation reflects different preferences and constraints, but some also reflects heterogeneous productivity across researchers. The model described above allows us to separate these two forces.

Validating these WTP responses is a difficult but important task. They are the key ingredient of our entire exercise, but they may be driven in part by sample selection effects or behavioral

biases or noise due to a simple lack of effort on the part of respondents. Although these are practical trade-offs that professors face throughout their careers, systematic data has not been collected in this way before to our knowledge. Thus, it is not clear what plausible variation would look like here. Still, we can conduct some tests motivated by economic and statistical theory to explore how reasonable these distributions of WTP are.

Motivated by the notion of opportunity costs, we test how WTP varies with researchers' implied hourly wage (per their annual salary divided by their total hours of work). Assuming this implied wage rate is a proxy for researchers' opportunity costs, it should be the case that researchers with higher hourly wages are willing to pay more for their free time. In Appendix Figure C2A, we see that this is indeed the case. Furthermore, it should be the case that, conditional on opportunity costs of time, researchers who expect to spend more time fundraising should have a higher WTP for guaranteed funding. In Appendix Figure C2B, we find this to be true.

As another test of the WTP estimates, Appendix Tables C1 and C2 report a series of analyses where we regress researchers' WTP responses on the state and choice variables of the model, and possibly also (*i*) the large vector of observable features describing each researcher ($\mathbf{X}_i$), (*ii*) an inverse Mills ratio constructed using the randomized survey participation incentives following Heckman (1979), and/or (*iii*) the researchers' WTP for the benchmark good (high-speed internet) following Dizon-Ross and Jayachandran (2022).

We find the variables of the model can explain 80–90% of the variation in WTP responses (see Appendix Table C1). This supports the model's ingredients as being key determinants of researchers' answers. Secondly, we find that the inverse Mills ratio is not a statistically significant predictor of responses (see Appendix Table C1). Under the assumptions outlined in Heckman (1979), this supports the notion that respondents did not differentially select into our sample as a function of their WTP for inputs and suggests some generalizability of our responses. Finally, we find that researchers' WTP for the benchmark good (high-speed internet at their house) is correlated with their WTP for inputs. Under the assumption that researchers' demand for home internet is truly uncorrelated with their productivity, this provides some evidence of systematic noise in how respondents are reporting their WTP in these experiments. But, the magnitudes of this relationship are quite small and the benchmark good WTP is included in our vector of features we use to make the researcher type index (that determines heterogeneity in the consumption parameters), which gives us some confidence that this systematic noise is small and the model will adjust for it to some degree. Overall, despite being un-incentivized experiments, researchers' responses appear to behave as expected and be of plausible magnitudes.

5. Estimates for the Model of Science

5.1. A View of Researchers' Utility Functions

The common parameter estimates for the model are reported in Appendix Table D1. Appendix Figure D1 illustrates the model's fit by comparing the distribution of researchers' responses with the model's predictions and errors. As evidence of fit, the median absolute difference between a researcher's stated WTP and the model's prediction ranges from 6–12 percentage points across the four questions.

To illustrate the utility function in a more intuitive way, Figure 2 shows how an average researcher's utility depends on the levels of the three state variables (i.e., salary, administrative duties, guaranteed research funding) and research output. The figure shows the percent change in a researcher's utility levels as the variable is increased from the 10^{th} percentile level to the 90^{th} percentile level while all other variables and parameters are held fixed at the sample averages.[22]

In terms of magnitude, salary and administrative duties are the most important to researchers' utility. Shifting the average researcher's salary from the 10^{th} percentile to the 90^{th} percentile increases their utility by roughly 60%. An equivalent relative increase in administrative duties reduces utility by about 20%. In contrast, similarly scaled increases in guaranteed funding or research output raise utility by only a few percentage points.

5.2. Productivity Distributions

Figure 3 shows the unconditional distributions of the production function parameters γ_i (funding intensity) and α_i (TFP). Figure 3A displays the distribution of the γ_i parameter, which describes the relative weight of funding versus research time in researchers' production functions. The distribution highlights a significant degree of heterogeneity across researchers, with roughly 20% of our sample having a funding intensity either larger than 0.6 or smaller than 0.2.

As both an interesting exercise and test of face validity, Appendix Figure D2 plots the average funding intensity parameters (γ_i) for the twenty minor fields represented in the sample. We find γ_i is highest in chemistry and engineering, where research is often capital intensive and involves the use of expensive lab equipment. In contrast, and in line with intuition, the social sciences (e.g., economics, political science) display the lowest funding intensity on average. Notably, Appendix Figure D2 highlights the heterogeneity in funding intensity across researchers even within these narrower fields of study.

Figure 3B displays the distribution of the α_i parameter, our measure of researchers' TFP. Our

[22] For simplicity, behavioral responses where researchers re-optimize are not incorporated into Figure 2.

estimates reveal a large skew in researchers' beliefs about their productivity.

As one test of face validity, Appendix Table D2 reports regression results that test for associations between researchers' productivity (α_i) and their performance per traditional metrics of scientific productivity, which are often more simply just output levels (e.g., publications, citations). Under some reasonable assumptions, we expect a positive association between these metrics, and we find it. Researchers with higher productivity beliefs are more likely to have more actual publication output whether that output is measured using recent publication counts or citation-weighted counts. This provides a signal that our estimates do in fact reflect real productivity differences across researchers.

To understand how variation in TFP affects output, we decompose the variance in log output into components attributable to TFP, input levels, and factor intensity. Note that the variance in log output due to the variance in log TFP is given by: $\text{Var}(\log(\alpha_i)) + 2\text{Cov}(\log(\alpha_i)\,,\, \gamma_i \log(B_i)) + 2\text{Cov}(\log(\alpha_i)\,,\, (1-\gamma_i)\log(R_i))$. Using this relationship we estimate that 46% of the variance in output across researchers is due to the variance in TFP. Without adjusting for the covariances between TFP and the input levels and funding intensity parameter, we obtain an estimate of 80%. This finding indicates that more productive researchers obtain more inputs that are well-suited to their production functions, and it motivates questions related to allocative efficiency that we return to in the next section.

To get some sense as to the (private) value of these productivity differences, Appendix Figure D4 shows the results of an exercise where we solve for researchers' WTP for productivity differences based on their willingness to trade utility from salary for utility from productivity. This allows for a comparison in dollar equivalents. Per this conversion, this heterogeneity in productivity appears economically meaningful, as researchers are often willing to pay thousands of dollars per year for higher productivity.

While Figure 3B shows a substantial degree of heterogeneity, some fraction of this variation may be due to differences in the demand for their output or the nature of science within their fields of study; for a similar reason, most studies on the industrial organization of firms report productivity distributions based only on within-industry variation. Thus, Figure 4 shows the distribution of researcher productivity (α_i) after various controls for field-specific (or "industry-specific") variation are introduced.

First, Figure 4 shows the raw distribution of productivity levels, which mirrors the distribution shown in Figure 3B but now on a logarithmic scale. Next, we regress researchers' TFP estimates on a set of major-field fixed effects, and we report the distribution of residual productivity levels. Lastly, we also condition productivity on the full vector of covariates used to generate the researcher type index.

Figure 4 also reports ratio of the 90^{th} and 10^{th} percentiles, the "90–10 TFP ratio" measure of

variance commonly reported in traditional, firm-level studies. The 90–10 TFP ratio of the raw productivity estimates is substantial, but, even after conditioning on the additional controls, we still find a large dispersion. The 90^{th} percentile researcher believes they are roughly 30-times as productive as the 10^{th} percentile researcher.

It is difficult to benchmark this dispersion. Firm-level estimates are primarily from manufacturing sectors, which typically involve 90–10 TFP ratios on the scale of 1× to 5× (Syverson 2011). Worker-level productivity estimates tend to be similarly dispersed (Hoffman and Stanton 2024); however, those estimates do not come from occupations as complex and creatively-oriented as science. One notable example is Sackman, Erikson, and Grant (1968) who used direct observation to estimate software engineers' productivity and found that some individuals were as much as an order of magnitude faster on coding tasks than others. Interestingly, this is the same approximate scale that we identify with our sample of researchers.[23] As another test of face validity, Appendix Figure D3 plots the TFP distribution on the same scale as the field-normalized citation distribution for the researchers in our sample that we can match to their publication records. Notably, the 90–10 ratio for citations is 56, which is very similar to the dispersion we observe in productivity. Overall, the degree of dispersion suggests substantial scope for misallocation. We investigate this next.

6. Counterfactuals and Allocative Efficiency

6.1. Overview

In this section, we use our productivity estimates to analyze allocative efficiency through the lens of our model. Specifically, we treat guaranteed funding (G) and duties (D) as policy levers a planner can adjust to maximize a given objective. Importantly, we search for the allocations of guaranteed funding and duties that maximize an objective after accounting for researchers' endogenous behavioral responses to the planner's decisions. The margins for behavioral responses are total hours worked, time spent fundraising (to obtain additional funds) and time spent on research (to directly produce output). In reality, researchers may endogenously adjust along many other margins (e.g., the types of projects they pursue), but these complexities are beyond the scope of our model. Still, adjustments to time allocations are likely a first-order response, and ignoring other margins allows us to keep the problem tractable.

We estimate optimal allocations given two different objectives. First, estimate the allocation that maximizes total output. Second, we consider a utilitarian objective of maximizing researchers' aggregate utility. To aid interpretation, we conduct an exercise where we estimate how much total research funding must be increased *using actual allocations* to achieve the same growth in scientific output that we are able to achieve *using actual funding levels* in alternative allocations.

[23] An interesting counter-example is Gabaix and Landier (2008) who estimate relatively small differences in talent (in the value-added sense) among CEOs despite their substantial differences in compensation.

Beyond this, we explore a range of alternative constraints on how inputs are reallocated to draw broader conclusions about the gains from reallocation. Appendix B contains further details on how we specify the optimization problems and constraints.

The externalities of science loom large in these counterfactuals. As noted, our approach assumes all researchers' production involves the same relative amount of externalities (per the common parameter κ that multiplies their private value from output into social value). In practice, externalities likely vary greatly across fields, which is why we primarily focus on reallocations *within* the five major fields of researchers in our sample (i.e., Engineering, math and related; Humanities and related; Medical and health sciences; Natural sciences; Social sciences). We also report results where we condition researchers' actual and counterfactual outcomes (i.e., output, utility) on the large vector of covariates used in the researcher type index. This approach allows us to isolate gains from reallocation between researchers with similar observable features.

6.2. Summary of Counterfactuals

Table 3 details the changes we estimate after reallocating inputs to maximize total output (Cols. 2–3) or researchers' total utility (Cols. 4–5); Column (1) describes the actual allocation of inputs for reference. In all cases of Table 3, we hold the total amount of funding in each major field fixed, we only reallocate within fields, and social value is evaluated where $\kappa = 10$ (implying that scientists capture 10% of the value of their output). Given our limited ability to incorporate heterogeneous preferences into the model, Columns (3) and (5) report changes in output and welfare after removing variation in those metrics correlated with the covariates used to construct the researcher type index.[24]

Input reallocation. The first four rows of Table 3 report how reallocation changes the equilibrium distribution of inputs. For these rows, Columns (2–3) are identical since they are based on the same counterfactual and we do not include any controls for these statistics; likewise for Columns (4–5). Compared to the status quo, the counterfactual allocations lead to more research time on average (approx. 20%), they increase the variance in research time (approx. 40%), and they decrease the variance in research budgets (approx. –20%).

In the Appendix, we illustrate the actual and optimal input distributions as well as Lorenz curves (see Figure D5) to show how more or less unequal the distributions are under actual and optimal allocations. In general, the model opts to make the distribution of duties more unequal, while the inequality of the distribution of guaranteed funding is relatively unchanged.

[24] In those cases, we regress researcher-level output or welfare metrics on the model variables and the full vector of covariates used in the type index, and then we subtract out variation in the metric predicted by the covariates conditional on the model variables. We include the model variables as controls because the productivity parameters are (partially) determined by them and so we do not want to remove variation in outcomes due to productivity differences.

After researchers' behavioral responses to these reallocations, there is not much change in the inequality of the distribution of research time and total funding (see Figure D5). However, there is significant reallocation of inputs across researchers (see Figure D6).

Gains from reallocations. Overall, the model suggests that there are large gains in output to be had from alternative allocations. Whether the objective is to maximize output or researchers' private utility, the new allocations yield roughly 130–160% more output. Given the changes in research time, this translates into significant welfare gains on the scale of 3–4% for researchers' and 5–15% for society. The finding that both objectives yield qualitatively similar gains is primarily driven by the fact that researchers have control over their time. Whatever the planner hopes to maximize, the approach is to ensure that the most productive researchers are incentivized to spend more time working on their science.

To contextualize these gains, we estimate how much more funding under actual allocations would be necessary to achieve the same growth in scientific output that our counterfactual allocations achieve using current funding levels. Table 4 reports these estimates under both a "*Mechanical*" scenario that does not incorporate researchers' behavioral response and a "*Behavioral*" scenario that incorporates researchers' responses. This exercise indicates that research budgets would need to increase roughly 40% for the actual allocations to achieve the same growth in output that we observe in the counterfactuals. On an annual basis this amounts to roughly $55,000 per researcher. In aggregate, this sums to roughly $220 million per year in our sample. Assuming there is no sample selection, this implies that the gains from a more efficient allocation are equivalent to a funding increase of roughly $13 billion per year if we scale these results from our sample to the entire population of researchers targeted by the survey.[25]

Additional results. In the appendix, we report results from counterfactual exercises using alternative constraints and specifications (Tables D3 and D4). There are a few findings of note. First, it appears that the allocation of duties is only meaningfully relevant for influencing researchers' utility, whereas the allocation of funding is what has the main influence on output.[26] Another important result from these alternative scenarios is that allowing for reallocation across major fields does not yield much additional output or value compared to the scenarios where reallocation is constrained to occur only within major fields (as is done in the counterfactuals reported in Table 3). This finding suggests that the degree of misallocation within fields is at least as large within fields as it is across fields. Similarly, we do not see much gain from allowing the total research budget to be unconstrained, and allowing only researchers' fundraising decisions to constrain the size of the budget. Interestingly, in this scenario, the total research budget

[25] For reference, total annual R&D spending at all US institutes of higher education was roughly $100 billion in recent years (Gibbons and NCSES 2024).

[26] When only duties are reallocated, researchers' utility increases by roughly 3%, but output only increases by roughly 1%. When only funding is reallocated, researchers utility increases by only 1%, but output increases by 150%.

grows only about 15%. While we caution from interpreting this result too seriously, it suggests that the total research budget is not far from the optimum given the (fixed) size of the research workforce. Of course, this ignores dynamic concerns which may certainly be relevant.

6.3. Input Wedges: Actual Versus Optimal

The results summarized thus far suggest that reallocating researchers' time constraints and funding can yield significant gains in output and welfare. In order to better understand how these gains are achieved, we now focus on a single counterfactual specification and compare the actual and optimized input levels. For simplicity, we focus for the remainder of this section on the scenario where duties and funding are jointly optimized to maximize output, which corresponds to the counterfactual described in Columns (2–3) of Table 3.

To better understand how actual and optimal input levels correlate at the individual level, we estimate a series of regressions of the following form:

$$\text{Actual Input Level}_i = a + \beta \text{Optimal Input Level}_i + \delta Z_i + \epsilon_i \,, \tag{10}$$

which relates researchers' actual and optimal input levels possibly conditioning on one (or more) covariate Z. If allocations were perfectly efficient, such a regression would yield an estimate $\widehat{\beta} = 1$ (with no standard error) since actual levels would equal optimal levels. If allocations are not efficient, then $\widehat{\beta} < 1$.

Focusing first on estimating Equation 10 with no covariates, Figure 5 reports estimates of β on a binned scatterplot of the data. For both inputs, there is a clear positive relationship between researchers' actual and optimal levels. For every hour a researcher *should* commit to research, they actually commit 0.3 hours on average. Interestingly, for every one dollar in total funding the researcher *should* have, they actually have nearly one dollar (0.987) on average. The figure also reports the mean absolute differences between actual and optimal levels, which are approximately 10 hours of research time and $70,000 in funding. For reference, these values are both roughly half the sample means.[27]

Appendix Figure D7 illustrates the same relationship between actual and optimal inputs, but instead focuses on the policy levers which are allocated pre-behavioral response (i.e., administrative duties and guaranteed research funding). In contrast to Figure 5 , here we find effectively no relationship between actual and optimal levels. Thus, while the optimum implied by the model is an equilibrium allocation of inputs that is relatively similar to what is actually observed on average, the model presents a very different way of achieving it.

Appendix Table D5 Column (1) replicates these univariate regressions, and then Columns (2–5) include additional sets of covariates, which are all subsets of the vector of covariates used

[27] Appendix Figure D8 shows the distribution of the researcher-level input wedges (actual minus optimal levels).

in the researcher type index: researchers' professional positions (e.g., rank, tenure status); their subjective measures of their research output (i.e., as reported in Table 2); and their socio-demographic features (e.g., age, gender, household structure).

Examining the R^2 statistics reported in Table D5 Columns (2–5) as we include different sets of covariates generally reveals small increases in R^2 compared to Column (1). At most, the inclusion of the full vector of covariates increases the R^2 by 4–14 percentage points. We interpret this pattern as indicating that misallocation using these sorts of observable features is a challenging exercise, and that the degree of misspecification in the model (along these specific dimensions) is relatively small. The latter gives us confidence in our results, and the former has the interesting implication that predicting which researchers are over- or under-resourced based only on their observables may be a challenging exercise.

6.4. Case Studies of Potential Frictions

When estimating regressions of the form in Equation 10, it is important to be careful when interpreting the δ coefficient. A covariate Z may be a significant predictor of actual levels conditional on optimal levels (i.e., $\widehat{\delta} \neq 0$) for two reasons: (*i*) the feature truly is a predictor of misallocation as implied by the model, in which case a positive (negative) association indicates that the feature is predictive of a researcher being over-resourced (under-resourced) due to a friction; (*ii*) there is misspecification in the model and the feature describes some heterogeneous preferences or demand variation that we have failed to capture, which has led to bias in our productivity estimates. We cannot separate these two possibilities.

Here, we embrace the interpretation that our estimates of δ are indicative of a friction (and not misspecification) only for two features for which misspecification is plausibly not a large concern, but these results should still be interpreted cautiously.

First, we focus on a highly scrutinized feature: gender. A large body of work has documented a wide range of biases and frictions facing female researchers when it comes to the acquisition of inputs (or credit) for their science.[28] But while the vast majority of this work rejects null hypotheses and finds female researchers are under-resourced, they typically cannot formally quantify *how much* female researchers are under-resourced. Our model and approach allow us to do just that.

In Appendix Table D5, Columns (4–5) report the results from estimating regressions of the form shown in Equation 10, where Z_i is an indicator for researchers who self-report as female. Whether we only include other socio-demographic covariates (Col. 4) or the full set of covariates (Col. 5), we estimate a statistically significant negative association indicating female researchers

[28] See, for example, Bornmann, Mutz, and Daniel (2007); European Commission (2015); Van der Lee and Ellemers (2015); Witteman et al. (2019); Kim and Moser (2021).

are under-resourced. They spend approximately 1 fewer hours on their research per week and have roughly $10,000 less in total research funding annually. Both of these wedges are approximately 10% of the sample mean.

Next, we focus on another question that has received much attention by meta-science scholars: the Matthew effect. In general, the Matthew effect posits that researchers amass resources beyond what their productivity warrants due to their social status (Merton (1968)). The specific version of this effect that we can test for is the degree to which the use of grant dollars and publication outcomes as productivity proxies distorts the allocation of inputs (e.g., Lee et al. 2013; Penner et al. 2013; Gralka, Wohlrabe, and Bornmann 2019). Again, we run regressions of the form shown in Equation 10, now including a vector variables describing researchers' recent grant funding and publication or citation output. As reported in Appendix Table D6, we find some evidence of statistically significant distortions whereby a one standard deviation increase in these traditional proxies leads to an over-resourcing on the scale of roughly 0.1–0.2 standard deviations (approximately 5–10% relative to the sample means).

These two cases reveal statistically significant predictors of the input wedges that are plausibly due to frictions in the allocation process. Researchers are more likely to be under-resourced if they are female, and they are more likely to be over-resourced if their observable input and output measures are higher. Still, the R^2 statistics shown in Table D5 convey a seemingly novel point—these observable features, which have received so much attention thus far, explain a very small amount of the misallocation implied by the model. It appears difficult to use standard observable features to predict which researchers are over or under-resourced.

6.5. Output Differences across Major Fields

Our final exercise seeks to understand the determinants of scientific output differences across the major fields of researchers. Doing so takes a strong stance on the comparability of output across fields, which is debatable. Thus, we treat this exercise as more speculative.

In general, the aggregate output of a market depends on input levels (i.e., the number of researchers, their funding, and their time spent on research), productivity levels (i.e., researchers' TFP), and the market's allocative efficiency. Here, we explore the relative importance of these three dimensions.

First, we divide all fields' total output by the number of researchers in that field in order to compare the per capita output only. Next, we use the field with the most output, Medicine, as a benchmark and consider the other fields' output in percentage terms relative to Medicine's actual total (per capita) output.

The gray bars in Figure 6 plot the total scientific output implied by the data and our productivity estimates across the five major fields of study. The four comparison fields have aggregate output

levels that are roughly 25–75% that of Medicine.

Medicine is the most resourced field, in terms of research funding and time, so our first test is to equate input levels across fields. In this case, aggregate output in the Social Sciences and the Humanities more than doubles. But, except for the Social Sciences most of the gap in aggregate output between each field and Medicine remains.

Average TFP levels across the fields are relatively similar such that equating productivity has little impact on differences in aggregate output. In fact, the average TFP level in Medicine is slightly lower than that of all other fields, which may be related to the size of the field.

Lastly, we use our estimates of output from the counterfactual where we maximize output within each of the five major fields to estimate aggregate output gaps in the scenario where all fields are equal in their allocative efficiency (i.e., it is maximized).[29] Here, we see a reordering of fields in terms of their per capita output. The Natural and Social sciences would produce more output than Medicine, and the other two fields would rise to producing nearly 75% the output of Medicine. In this sense, differences in allocative efficiency appears to be the most important determinant of differences in aggregate output across fields of science.

We can only speculate as to why, based on our sample of medical researchers, the field of medicine appears to be the best at allocating resources efficiently. One might think that medical researchers spend more time engaged in fundraising efforts and that process facilitates positive selection on productivity; however, medical researchers exhibit quite average levels of fundraising effort.[30]

Appendix Figure D9 shows that our sample of medical researchers have a markedly different distribution of duties (e.g., teaching, administration, clinical, etc.) than the other researchers in our sample. The mean duty level is slightly lower relative to other fields, and the variance is much larger. This sort of allocation is consistent with the logic of our model and results — in a world with highly dispersed productivity, the planner should move non-research duties on to as few, low productivity (in the scientific sense) researchers. Of course there are both practical constraints and unmodeled objectives. For example, high-productivity researchers could plausibly have the largest externalities from their other duties (e.g., teaching) and the allocations we observe are balancing social value from research with social value from those other duties. Further work on understanding across-field differences in allocation mechanisms in science seems warranted.

[29] Since this leads to output growth in all fields, we re-scale aggregate levels so that the level observed in the field of Medicine remains the benchmark at 100%.

[30] The field-specific averages for fundraising hours per week are as follows: Engineering & Math = 6.1; Humanities = 2.7; Medicine = 4.9; Natural sciences = 6.7; Social Sciences = 2.5.

7. Discussion

Our methodology allows us to estimate researchers' productivity beliefs without observing output quantities or input prices. With these estimates, we provide a novel view of productivity and allocative efficiency in science. We find actual input allocations to be positively correlated with those that maximize plausible objectives, but we also find large gains to be had from more efficient allocations. In the counterfactuals we explore, total scientific output per research-hour or per research-dollar could be more than doubled.

Our model abstracts away from many important considerations that likely drive actual allocations (e.g., dynamic returns), and our data is based on stated preferences. These and other limitations of our approach motivate us to interpret these magnitudes as plausible upper bounds on the gains from more efficient use of scientific inputs. These large potential gains provide a glimmer of hope amid declining R&D productivity across most sectors of the economy (Bloom et al. 2020) and the persistently growing burden of knowledge that raises the cost of conducting frontier science (Jones 2009). Furthermore, our approach to identifying productive scientists may prove useful in talent selection processes more generally (e.g., Agarwal and Gaule 2020).

Our approach has deep roots in the economics of labor (i.e., surveys of time use and work-leisure trade-offs), industrial organization (i.e., production function estimation), marketing (i.e., using surveys to identify demand functions), and macroeconomics (i.e., models of factor misallocation). By drawing on insights from these fields, our methodology allows us to overcome many of the challenges that have long plagued our understanding of productivity and efficiency in science. However, this approach introduces new assumptions and challenges to grapple with.

Our analyses reveal researchers' *beliefs* about their productivity and therefore their *beliefs* about how well inputs are allocated. This is a crucial limitation, since there are clearly many potential biases affecting these beliefs. Still, science is inherently about forecasting uncertain outcomes, and, therefore, the optimal mechanisms for identifying productive researchers and allocating them more inputs will need to tackle this challenge of engaging with researchers' forecasts of their productivity. Specifically, one important next step in this line of work will be to ensure that the producers being studied report their willingness-to-pay for inputs truthfully. Of course, the theoretical underpinnings of how to elicit true willingness-to-pay estimates have long been established (Becker, DeGroot, and Marschak 1964). However, the magnitudes involved in our setting involve trade-offs on the scale of tens or hundreds of thousands of dollars. Thus, developing the practical details of incentivizing truthful responses from producers at this scale would be a fruitful endeavor.

More broadly, our approach may prove useful in other settings that face similar challenges in estimating productivity and efficiency. There are many markets populated by a large number of

producers acquiring inputs in a highly decentralized way to produce outputs that are not easy to observe. For example, in developing economies, accurate producer-level data on outputs can be difficult to obtain (e.g., Tybout 2000); in entrepreneurship, many organizations never produce observable output before exiting (e.g., Decker et al. 2014); and in non-profit sectors, the output may be so high-dimensional that reaching consensus on a suitable proxy is challenging (e.g., Philipson and Lakdawalla 2001). In each of these examples, our methodology could provide a way forward to better understanding the distribution and determinants of productivity and efficiency.

Tables & Figures

TABLE 1
Summary Statistics—Salary, Funding, Time, and Fields

	mean	*s.d.*	*p*50
Salary and research funds, $/year			
Total salary	159,028.23	74,516.13	140,000.00
Guaranteed & existing funding	52,223.08	83,420.47	5,000.00
Fundraising expectations	93,909.19	144,013.69	20,000.00
Work time, hrs./week			
Total work	48.55	10.00	48.00
Research	18.51	9.60	17.40
Fundraising	4.47	5.09	2.65
Administration	7.48	6.17	5.80
Teaching and other work	18.09	9.78	17.20
Major field, {0,1}			
Engineering, math, & related	0.17	0.37	
Humanities & related	0.19	0.39	
Medical & health sciences	0.28	0.45	
Natural sciences	0.16	0.37	
Social sciences	0.21	0.40	

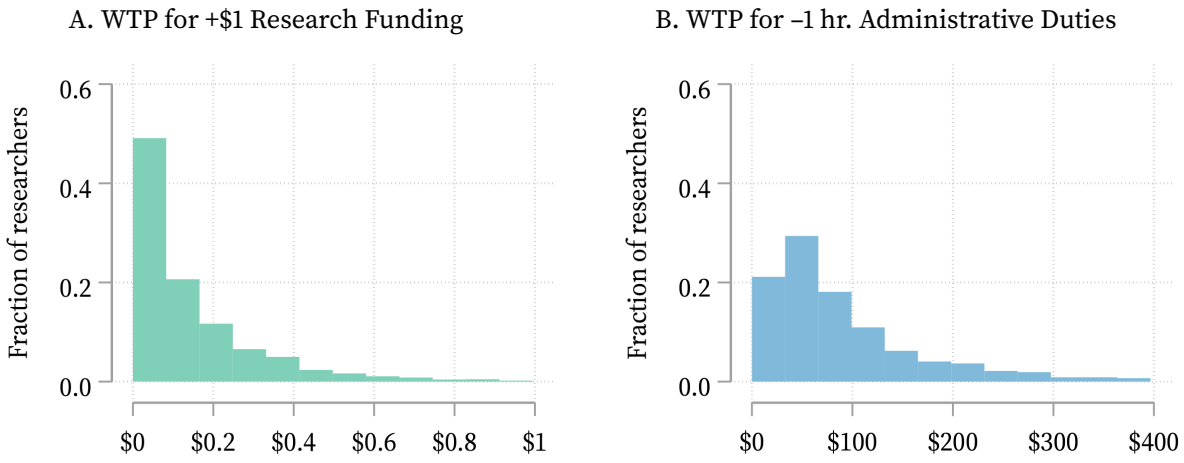
Note: Reports summary statistics for 4,003 researcher-level observations. Unless otherwise noted, all variables are continuous and bound below by zero. *{0,1}* indicates binary variables.

TABLE 2
Summary Statistics—Subjective Output

	mean	*s.d.*
Intended research outputs, {0,1,2}		
Journal articles	1.87	0.37
Books	0.52	0.68
Materials or methods	0.70	0.68
Products	0.47	0.64
Intended research audience, {0,1,2}		
Academic peers	1.88	0.36
Policymakers	0.84	0.69
Businesses	0.54	0.62
General public	0.83	0.62
Riskiness of own research, [0,10]		
Own belief	4.63	2.34
Belief of peers' beliefs	4.57	2.37
Theoretical vs. empirical, [0,10]		
Ask [0] or answer [10] questions	4.89	2.54

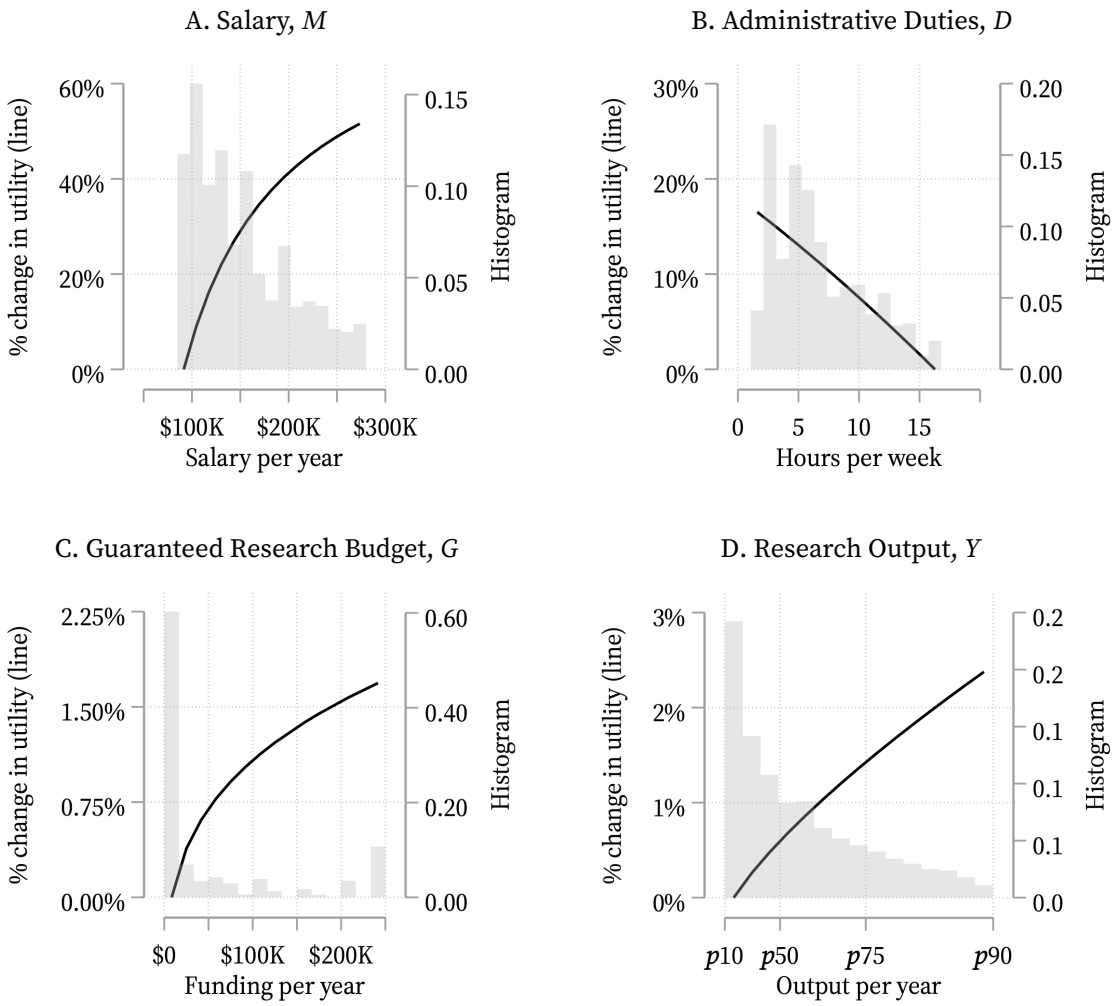
Note: Reports summary statistics for 4,003 researcher-level observations. The intended research outputs and audience variables are coded responses to questions of the form *How often are the following the intended output / audience of your research,* where responses are coded as follows: *Rarely*=0, *Sometimes*=1, *Very often*=2. The question about risk used a scale where 0 indicated no risk and 10 indicated very high risk. The question about theory versus empirics used a scale where 0 indicated that the researcher focused on asking new questions and 10 indicated that the researcher focsed on answering existing questions.

FIGURE 1
Willingness to Pay Experiment Responses



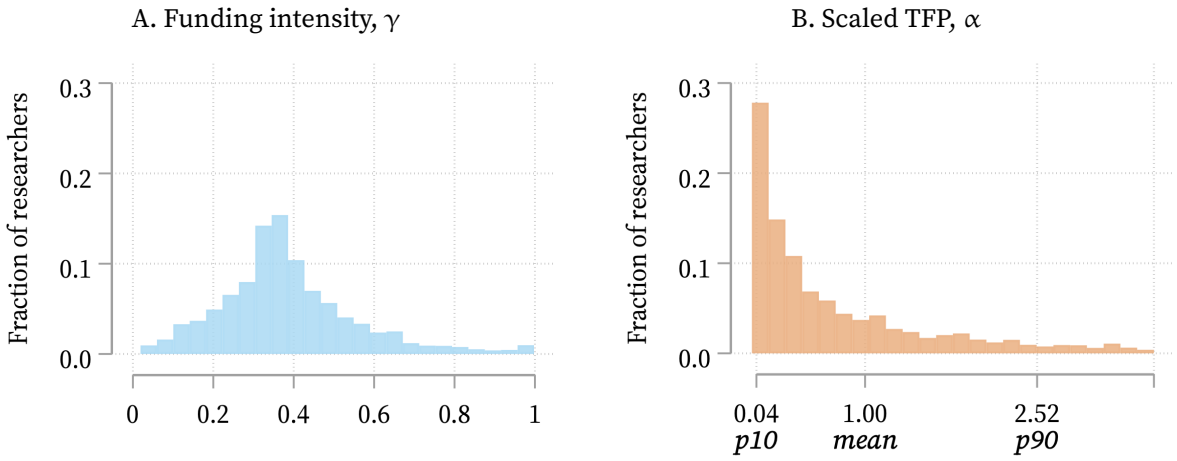
Note: Shows the distribution of researchers' stated WTP for $1 more of additional research funds (Panel A) and 1 hour less of administrative duties (Panel B). Approximately 1% of the upper tails have been trimmed for visibility.

FIGURE 2
Utility Function Visualization



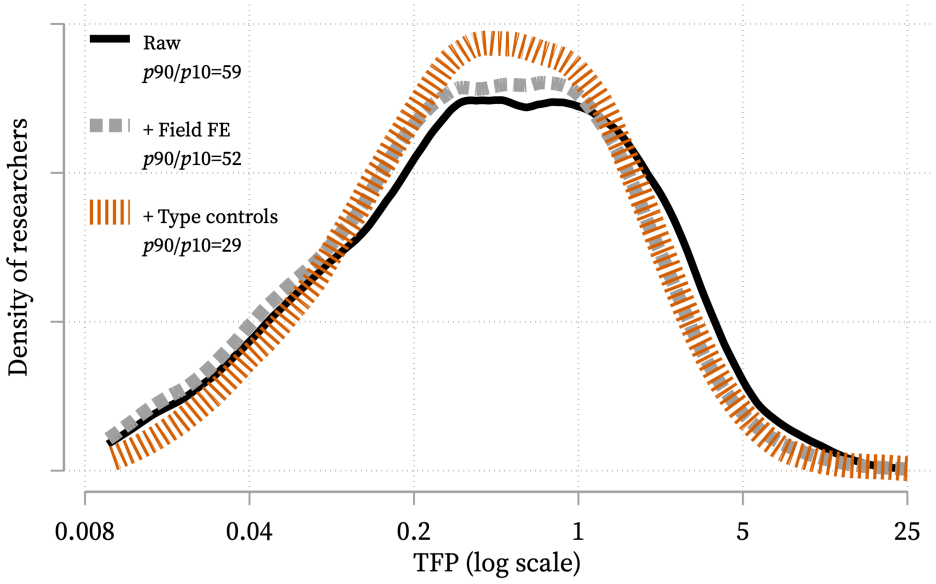
Note: Shows the percent change in utility in four of the key model variables (Panels A–D) holding all other variables and parameters fixed at the sample means. The black line (left *y* axis) shows the percent change in utility as the focal variable increases from the 10th percentile to the 90th percentile. The histogram (right *y* axis) shows the distribution of the focal variable in the sample. Note the different scales of all *y* axes.

FIGURE 3
Production Function Parameters



Note: Shows the distribution of researcher-specific estimates of the production function parameters γ_i (funding intensity; Panel A) and α_i (productivity; Panel B). In Panel (A, larger values indicate more funding intensity and $\gamma \in [0, 1]$. In Panel (B), the top 5% of productivities have been trimmed for visibility and estimates are scaled into units where the mean is normalized to equal 1 (with the 10th and 90th percentiles of the distribution noted on the x axis).

FIGURE 4
Conditional TFP Distributions



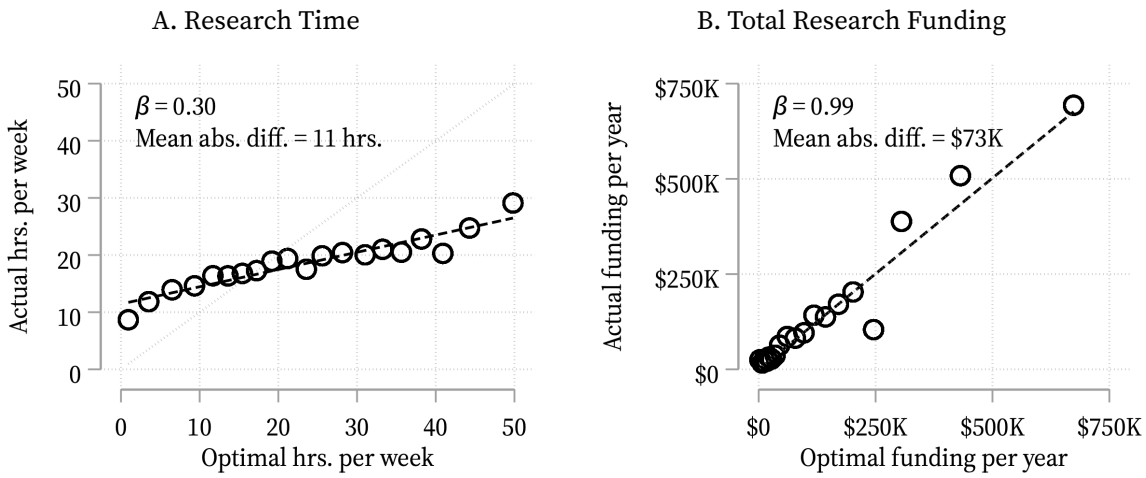
Note: Shows the distribution of researchers' TFP (α_i) where units have been rescaled so the mean equals 1; note the log scale. In addition to (*i*) the distribution of the raw values, productivities are also shown after removing residual variation due to (*ii*) major field fixed effects, and (*iii*) the covariates used to construct the researcher type index. 90-10 percentile ratios are also shown.

TABLE 3
Outcomes given Actual and Optimized Allocations

	Current allocation	Optimized allocations			
	(1)	(2)	(3)	(4)	(5)
Research inputs					
Research hrs./week, avg.	18.5	+27%	+27%	+14%	+14%
Research hrs./week, s.d.	9.6	+43%	+43%	+43%	+43%
Budget $-K/year, avg.	147.1	0%	0%	0%	0%
Budget $-K/year, s.d.	206.9	–18%	–18%	–20%	–20%
Research output					
Output, avg.	*n.r.*	+160%	+140%	+160%	+130%
Output, s.d.	*n.r.*	–5.2%	–7.9%	–5.4%	–5.7%
Output per hr.	*n.r.*	+100%	+91%	+130%	+110%
Welfare					
Researcher utility, avg.	*n.r.*	+4.5%	+3.3%	+4.7%	+3.2%
Researcher utility, s.d.	*n.r.*	–5.2%	–7.9%	–5.4%	–5.7%
Researcher utility per hr.	*n.r.*	+25%	+24%	+16%	+15%
Social value, avg.	*n.r.*	+16%	+5.7%	+16%	+5.2%
Social value, s.d.	*n.r.*	+21%	–6.5%	+20%	–3.4%
Social value per hr.	*n.r.*	+34%	+27%	+26%	+19%
Objective, max		Y	Y	$\mathcal{V}$	$\mathcal{V}$
Type controls			✓		✓

Note: Reports summary statistics for inputs under actualallocations (Col. 1). The first three sets of rows in Columns 2–5 report the percentage change in research inputs (*Research inputs*), outputs (*Research outputs*), and utility (*Welfare*) under alternative allocations; estimates are rounded to aid in comparison. The bottom sets of rows outline the objective of the counterfactuals explored in Columns 2–5. The two different objectives explored are maximizing output (Y) or researchers' private utility ($\mathcal{V}$). Models with *Type controls* report output and welfare changes after removing residual variation due to covariates used to construct the researcher type index. All optimized allocations allow for researchers' behavioral responses.

FIGURE 5
Actual and Optimal Input Level Correlations



Note: Shows the binned scatterplot of actual and optimal input levels per an objective of maximizing output. Also reports the coeffificient estimate from a regression of actual on optimal input levels and the mean absolute difference between actual and optimal input levels.

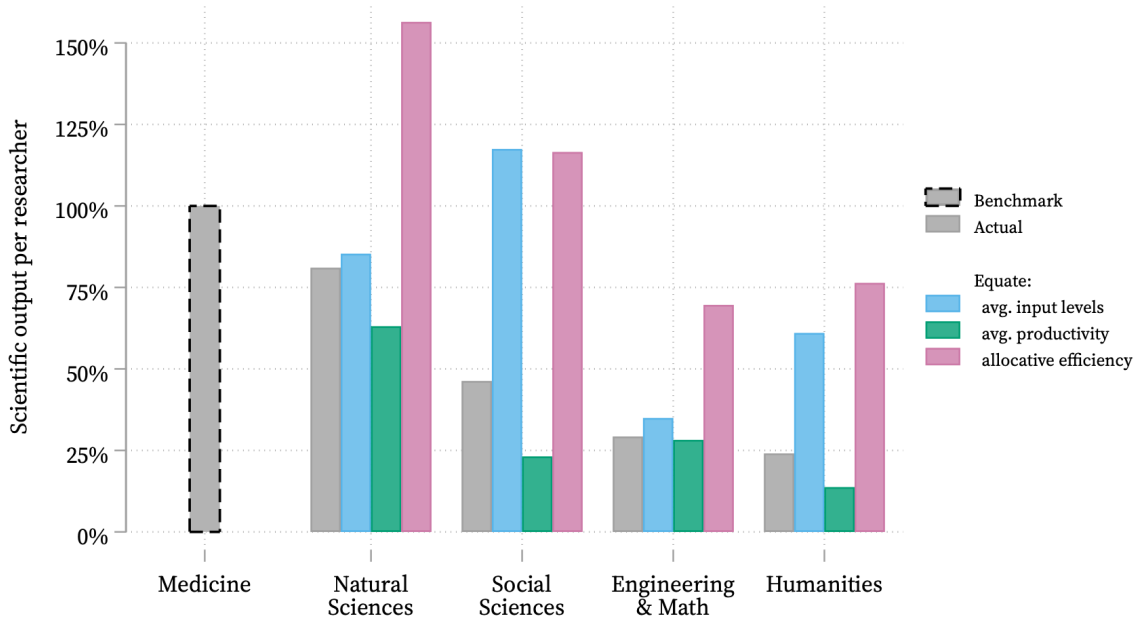
TABLE 4

Growth in Funding with Actual Allocations Needed to Produce the Same Output as Optimized Allocation of Actual Funding

	Mechanical	Behavioral
Δ *Annual guaranteed funding, G*		
Sample average	+$57K (109%)	+$55K (105%)
Δ *Annual total funding, B*		
Sample average	+$57K (42%)	+$55K (41%)
Sample total	+$228M	+$220M
Population total	+$13.7B	+$13.2B

Note: Reports the increase in guaranteed (*G*) and total (*B*) research funding under actual allocations necessary to achieve the same growth in output achieved by the reallocation of actual input levels that maximizes researchers' utility (+160%). Sample averages on a per-researcher basis are reported in addition to sample totals (summing over all in-sample researchers) and implied population totals (scaling the sample up to the population size per the survey response rate). The *Mechanical* scenario does not allow researchers to re-optimize their time allocations and the *Behavioral* scenario does.

FIGURE 6
Across Field Output Comparisons



Note: Shows the per capita scientific output of each major field benchmarked to the field with the most output, Medicine. Actual output levels (in percentage terms relative to the benchmark) are shown alongisde counterfactual output levels where all fields have the same average input levels per researcher, average productivity, and allocative efficiency.

References

Adams, James D. 1990. "Fundamental stocks of knowledge and productivity growth." *Journal of Political Economy* 98 (4): 673–702.

Adams, James D, and Zvi Griliches. 1998. "Research Productivity in a System of Universities." *Annals of Economics and Statistics* 49-50: 127–162.

Agarwal, Ruchir, and Patrick Gaule. 2020. "Invisible geniuses: Could the knowledge frontier advance faster?" *American Economic Review: Insights* 2 (4): 409–424.

Andrews, Michael J. 2023. "How do institutions of higher education affect local invention? Evidence from the establishment of US colleges." *American Economic Journal: Economic Policy* 15 (2): 1–41.

Arrow, Kenneth. 1962. "Economic Welfare and the Allocation of Resources for Invention." In *The Rate and Direction of Inventive Activity: Economic and Social Factors*, 609–626: Princeton University Press.

Azoulay, Pierre, Christian Fons-Rosen, and Joshua S Graff Zivin. 2019. "Does science advance one funeral at a time?" *American Economic Review* 109 (8): 2889–2920.

Becker, Gordon M, Morris H DeGroot, and Jacob Marschak. 1964. "Measuring utility by a single-response sequential method." *Behavioral Science* 9 (3): 226–232.

Bloom, Nicholas, Charles I Jones, John Van Reenen, and Michael Webb. 2020. "Are ideas getting harder to find?" *American Economic Review* 110 (4): 1104–1144.

Bornmann, Lutz, Rüdiger Mutz, and Hans-Dieter Daniel. 2007. "Gender differences in grant peer review: A meta-analysis." *Journal of Informetrics* 1 (3): 226–238.

De Loecker, Jan, and Chad Syverson. 2021. "An industrial organization perspective on productivity." In *Handbook of Industrial Organization*, vol. 4, 141–223: Elsevier.

Decker, Ryan, John Haltiwanger, Ron Jarmin, and Javier Miranda. 2014. "The role of entrepreneurship in US job creation and economic dynamism." *Journal of Economic Perspectives* 28 (3): 3–24.

Digital Science. 2018. "Dimensions [Software]." Available from https://app.dimensions.ai. Accessed in 2023, under licence agreement.

Dizon-Ross, Rebecca, and Seema Jayachandran. 2022. "Improving Willingness-to-Pay Elicitation by Including a Benchmark Good." *AEA Papers and Proceedings* 112: 551–555.

European Commission. 2015. "She figures.", Directorate-General for Research and Innovation. Luxembourg: Publications Office.

Gabaix, Xavier, and Augustin Landier. 2008. "Why has CEO pay increased so much?" *The Quarterly Journal of Economics* 123 (1): 49–100.

Gibbons, Michael T, and NCSES. 2024. "Higher Education R&D Expenditures Increased 11.2%, Exceeded $108 Billion in FY 2023." National Science Foundation 25-313. https://ncses.nsf.gov/pubs/nsf25313.

Gralka, Sabine, Klaus Wohlrabe, and Lutz Bornmann. 2019. "How to measure research efficiency in higher education? Research grants vs. publication output." *Journal of Higher Education Policy and Management* 41 (3): 322–341.

Hager, Sebastian, Carlo Schwarz, and Fabian Waldinger. 2024. "Measuring science: Performance metrics and the allocation of talent." *American Economic Review* 114 (12): 4052–4090.

Hall, Bronwyn H, and Jacques Mairesse. 2024. "Explorations of Cumulative Advantage Using Data on French Physicists."

Hausman, Naomi. 2022. "University innovation and local economic growth." *Review of Economics and Statistics* 104 (4): 718–735.

Heckman, James J. 1979. "Sample selection bias as a specification error." *Econometrica* 47 (1): 153–161.

Hoffman, Mitchell, and Christopher T. Stanton. 2024. "People, Practices, and Productivity: A Review of New Advances in Personnel Economics." In *Handbook of Labor Economics,*: Elsevier.

Hopenhayn, Hugo A. 2014. "Firms, misallocation, and aggregate productivity: A review." *Annual Review of Economics* 6 (1): 735–770.

Hsieh, Chang-Tai, and Peter J Klenow. 2009. "Misallocation and manufacturing TFP in China and India." *The Quarterly Journal of Economics* 124 (4): 1403–1448.

Jaffe, Adam B. 1989. "Real effects of academic research." *American Economic Review* 79 (5): 957–970.

Jena, Anupam B, and Tomas J Philipson. 2008. "Cost-effectiveness analysis and innovation." *Journal of Health Economics* 27 (5): 1224–1236.

Jones, Benjamin F. 2009. "The burden of knowledge and the "Death of the renaissance man": Is innovation getting harder?" *Review of Economic Studies* 76 (1): 283–317.

Jones, Charles I. 2005. "Growth and ideas." In *Handbook of economic growth,* vol. 1, 1063–1111: Elsevier.

Kim, Scott Daewon, and Petra Moser. 2021. "Women in science. Lessons from the Baby Boom."

Lakdawalla, Darius N, Eric C Sun, Anupam B Jena, Carolina M Reyes, Dana P Goldman, and Tomas J Philipson. 2010. "An economic evaluation of the war on cancer." *Journal of Health Economics* 29 (3): 333–346.

Lee, Carole J, Cassidy R Sugimoto, Guo Zhang, and Blaise Cronin. 2013. "Bias in peer review." *Journal of the American Society for Information Science and Technology* 64 (1): 2–17.

Van der Lee, Romy, and Naomi Ellemers. 2015. "Gender contributes to personal research funding success in The Netherlands." *Proceedings of the National Academy of Sciences* 112 (40): 12349–12353.

Lotka, Alfred J. 1926. "The frequency distribution of scientific productivity." *Journal of the Washington Academy of Sciences* 16 (12): 317–323.

Mansfield, Edwin. 1998. "Academic research and industrial innovation: An update of empirical findings." *Research Policy* 26 (7-8): 773–776.

Merton, Robert K. 1968. "The Matthew effect in science: The reward and communication systems of science are considered.." *Science* 159 (3810): 56–63.

Merton, Robert K. 1973. *The Sociology of Science: Theoretical and Empirical Investigations*. Chicago: The University of Chicago Press.

Myers, Kyle R, Wei Yang Tham, Nina Cohodes, Karim Lakhani, Rachel Mural, Jerry G Thursby, Marie C Thursby, and Yilun Xu. 2023. "New facts and data about US research professors."

Myers, Kyle R, Wei Yang Tham, Yian Yin, Nina Cohodes, Jerry G Thursby, Marie C Thursby, Peter Schiffer, Joseph T Walsh, Karim R Lakhani, and Dashun Wang. 2020. "Unequal effects of the COVID-19 pandemic on scientists." *Nature Human Behaviour* 4 (9): 880–883.

National Science Foundation. 2023. "Higher Education Research and Development Survey (HERD)." Available from https://www.nsf.gov/statistics/srvyherd/. Accessed in 2023.

Nelson, Richard. 1959. "The simple economics of basic scientific research." *Journal of Political Economy* 67 (3): 297–306.

Nordhaus, William D. 2004. "Schumpeterian profits in the American economy: Theory and measurement."

Penner, Orion, Raj K Pan, Alexander M Petersen, Kimmo Kaski, and Santo Fortunato. 2013. "On the predictability of future impact in science." *Scientific Reports* 3 (1): 3052.

Philipson, Tomas, and Anupam Jena. 2006. "Surplus appropriation from R&D and health care technology assessment procedures."

Philipson, Tomas, and Darius Lakdawalla. 2001. "Medical care output and productivity in the nonprofit sector." In *Medical Care Output and Productivity*, edited by David M Cutler and Ernst R. Berndt, 119–140:

University of Chicago Press.

Popp, David. 2017. "From science to technology: The value of knowledge from different energy research institutions." *Research Policy* 46 (9): 1580–1594.

Qiu, YJ Jeff. 2023. "The Matthew effect, research productivity, and the dynamic allocation of NIH grants." *RAND Journal of Economics* 54 (1): 135–164.

Restuccia, Diego, and Richard Rogerson. 2013. "Misallocation and productivity." *Review of Economic Dynamics* 16 (1): 1–10.

Rosenbloom, Joshua L, Donna K Ginther, Ted Juhl, and Joseph A Heppert. 2015. "The effects of Research & Development funding on scientific productivity: Academic chemistry, 1990-2009." *PLoS ONE* 10 (9): e0138176.

Sackman, Harold, Warren J Erikson, and E Eugene Grant. 1968. "Exploratory experimental studies comparing online and offline programming performance." *Communications of the ACM* 11 (1): 3–11.

Shapin, Steven. 1995. "Here and Everywhere: Sociology of Scientific Knowledge." *Annual Review of Sociology* 21 (1): 289–321.

Syverson, Chad. 2011. "What determines productivity?" *Journal of Economic Literature* 49 (2): 326–365.

Tybout, James R. 2000. "Manufacturing firms in developing countries: How well do they do, and why?" *Journal of Economic Literature* 38 (1): 11–44.

Valero, Anna, and John Van Reenen. 2019. "The economic impact of universities: Evidence from across the globe." *Economics of Education Review* 68: 53–67.

Wang, Eric C, and Weichiao Huang. 2007. "Relative efficiency of R&D activities: A cross-country study accounting for environmental factors in the DEA approach." *Research Policy* 36 (2): 260–273.

Witteman, Holly O, Michael Hendricks, Sharon Straus, and Cara Tannenbaum. 2019. "Are gender gaps due to evaluations of the applicant or the science? A natural experiment at a national funding agency." *The Lancet* 393 (10171): 531–540.

Zuckerman, H. 1988. "The Sociology of Science." In *Handbook of Sociology*, edited by NJ Smelser, 511–574: Sage Publications.

Appendix A. Additional Survey Statistics and Comparisons

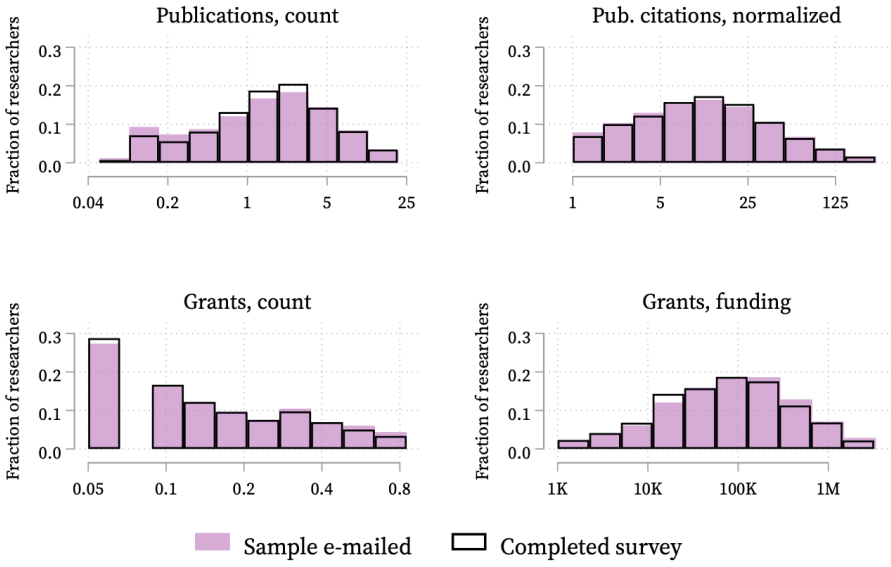
TABLE A1
Survey Completion per Randomized Treatments

	(1)
Either incentive	0.00297**
	(0.00117)
Both incentives	0.00797***
	(0.00141)
1 reminder	0.0112***
	(0.00114)
2 reminders	0.0188***
	(0.00119)
Constant	0.0197***
	(0.00107)
N obs.	131,672

Note: Reports estimates from a regression of a binary indicator of survey completion on binary indicators for the randomized incentives and reminders including observations for all researchers emailed. Robust standard errors reported; * $p < 0.10$,** $p < 0.05$,*** $p < 0.01$.

FIGURE A1
Sample Representativeness per Publication and Grant Funding Measures

A. Distribution Comparisons



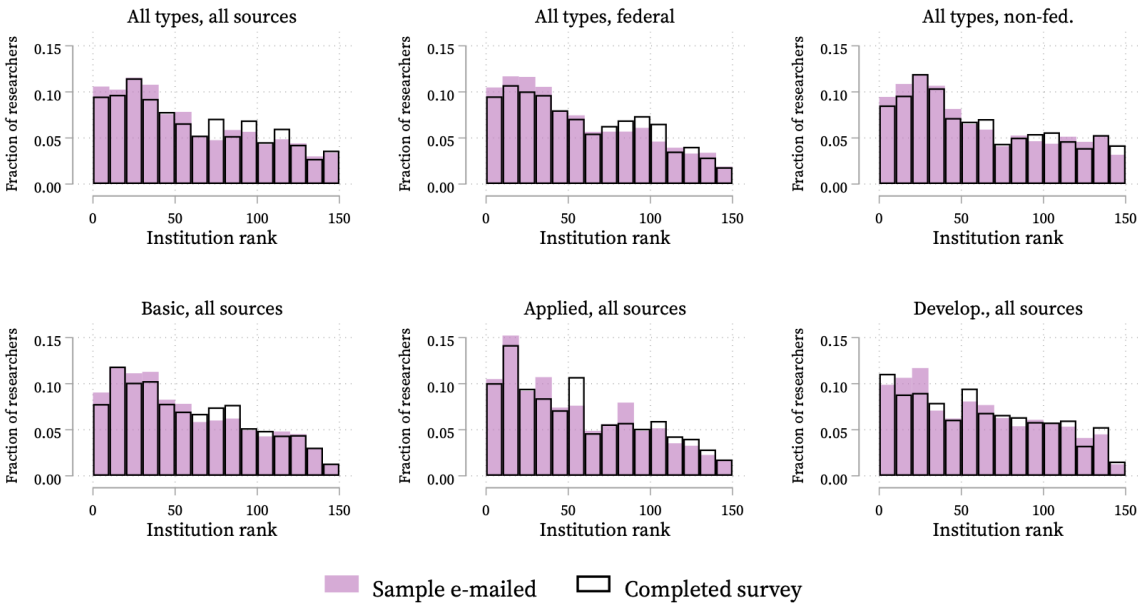
B. Regression Comparisons

	Pub., count (1)	Pub. cites, normalized (2)	Grant, count (3)	Grant, total $ (4)
Completed survey	−0.111	−0.575	0.00878**	−5568.8
	(0.0688)	(0.807)	(0.00343)	(8519.4)
Constant	2.906***	20.59***	0.117***	121272.4***
	(0.0158)	(0.185)	(0.000715)	(2093.0)
% diff.	−3.8%	−2.8%	7.5%	−4.6%
N obs.	87,000	87,000	87,000	87,000

Note: Panel (a) shows the distribution of publication and grant outcomes split by whether the researcher was emailed (i.e., the full sample) versus those who completed the survey; note the log x axes. Panel (b) reports estimates from a regression of each measure on a binary indicator for whether the researcher completed the survey; *% diff.* reports this difference in percentage terms. Robust standard errors reported; * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

FIGURE A2
Sample Representativeness per Institutional Funding Measures

A. Distribution Comparisons



B. Regression Comparisons

	All sources, by type			All types, by source		
	All	Federal	Non–fed.	Basic	Applied	Develop.
	(1)	(2)	(3)	(4)	(5)	(6)
Completed survey	–35.01***	–21.30***	–13.71***	–24.31***	–7.170***	–3.812***
	(6.953)	(4.511)	(3.062)	(4.867)	(2.492)	(1.328)
Constant	649.1***	351.7***	297.4***	413.2***	176.7***	60.46***
	(1.308)	(0.863)	(0.578)	(0.917)	(0.459)	(0.254)
% diff.	–5.4%	–6.1%	–4.6%	–5.9%	–4.1%	–6.3%
N obs.	130,785	130,785	130,785	130,785	130,735	128,169

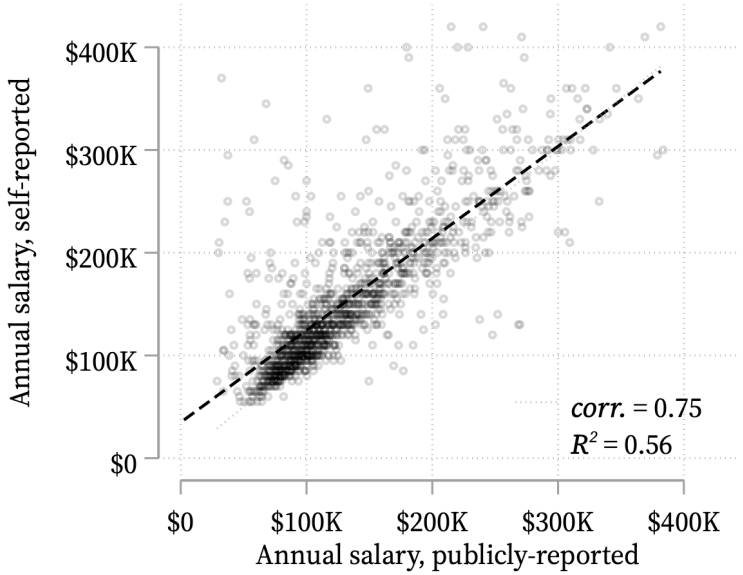
Note: Panel (a) shows the distribution of institutional funding measures split by whether the researcher was emailed (i.e., the full sample) versus those who completed the survey. Panel (b) reports estimates from a regression of each measure on a binary indicator for whether the researcher completed the survey; *% diff.* reports this difference in percentage terms. Robust standard errors reported; * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

TABLE A2
Distribution of Minor Fields of Study

	share
Agriculture	0.02
Biology	0.05
Business	0.05
Chemistry	0.02
Communication	0.03
Computer science	0.02
Economics	0.02
Education	0.04
Engineering	0.06
Geography	0.05
Humanities	0.11
Law	0.02
Mathematics	0.03
Medical school	0.21
Medicine and health	0.07
Other social sciences	0.04
Physics	0.04
Political science	0.04
Psychology	0.04
Sociology	0.04

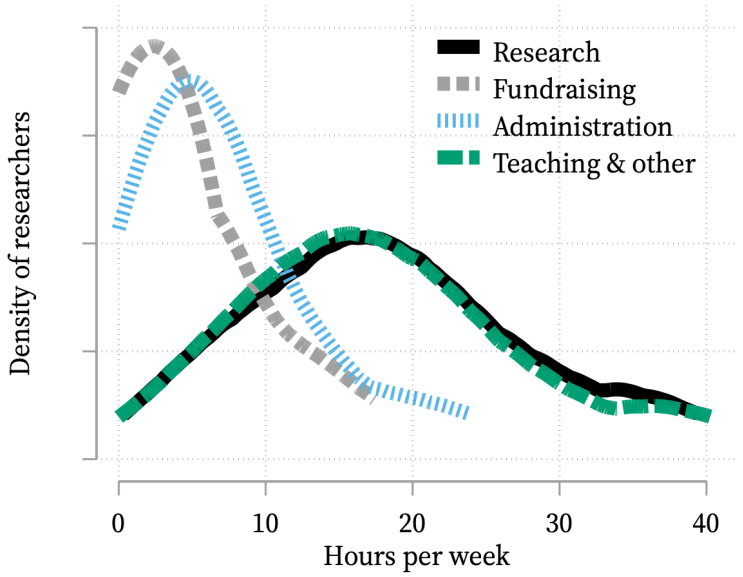
Note: Reports the share of 4,003 researcher-level observations in each minor field.

FIGURE A3
Correlation of Self- and Publicly-reported Annual Salaries



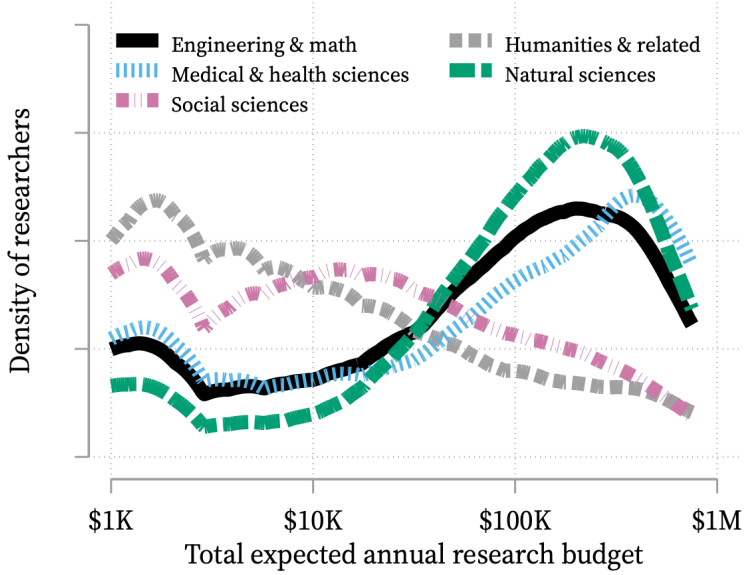
Note: Shows the publicly- and self-reported annual salaries for 1,369 in-sample researchers whose salaries were located in public reportings, and reports the pairwise correlation alongside the R^2 statistic from a regression of self-reported on publicly-reported salary.

FIGURE A4
Time Allocation Distributions



Note: Shows the distribution of researchers' hours per week spent on four categories of work tasks.

FIGURE A5
Research Funding Distributions by Major Fields of Study



Note: Shows the distribution of researchers' total expected annual research budget (including guaranteed funds plus fundraising expectations) split by the researchers' major field of study; note the log scale.

TABLE A3
Summary Statistics of Other Variables

	mean	*s.d.*
From HERD: Institution-level R&D		
Total R&D, $M	611.32	451.18
R&D per researcher, $M	0.61	0.86
Share federal gov.'t R&D, [0,1]	0.52	0.12
Share basic R&D, [0,1]	0.63	0.20
From Dimensions: Research output		
Publications per year	5.28	7.09
Citations per year	23.07	49.10
Co-authors per publication per year	9.09	68.40
Position details		
Assistant professor, {0,1}	0.26	0.44
Associate professor, {0,1}	0.26	0.44
Full professor, {0,1}	0.41	0.49
Other rank, {0,1}	0.07	0.25
Not on tenure track, {0,1}	0.19	0.39
Pre-tenure, {0,1}	0.22	0.42
Tenured, {0,1}	0.58	0.49
Years until next contract eval.	3.68	1.82
Duration of contract	2.45	1.93

Note: Reports summary statistics for 4,003 researcher-level observations. *From HERD* indicates variables from National Science Foundation (2023). *From Dimensions* indicates variables from the Digital Science (2018) dataset. *N.R.* stands for Not Reported.

TABLE A4
Summary Statistics of Other Variables (cont.'d)

	mean	*s.d.*
Gender identity, {0,1}		
Female	0.40	0.49
Male	0.55	0.50
Other or N.R	0.05	0.22
Racial/ethnic identity, {0,1}		
Asian	0.13	0.33
Black	0.03	0.18
Hispanic	0.06	0.23
White	0.77	0.42
Other or N.R	0.05	0.21
Citizenship, {0,1}		
Citizen, domestic-born	0.71	0.45
Citizen or perm resident, foreign-born	0.24	0.43
Other or N.R citizenship	0.05	0.21
1st–3rd generation in U.S.	0.30	0.46
Other or N.R generation in U.S.	0.70	0.46
Other covariates		
Age	48.78	12.05
Household total income	260,336.00	215,388.31
Married or domestic partnership, {0,1}	0.82	0.39
Single, {0,1}	0.13	0.34
Other or N.R relationship, {0,1}	0.05	0.22
Dependents in household	0.98	1.13
Risk-taking in personal life, [0,10]	5.26	2.13

Note: Reports summary statistics for 4,003 researcher-level observations. *N.R.* stands for Not Reported.

TABLE A5
Pairwise Correlations of Subjective Output Measures

	Articles	Books	Methods	Products	Academic	Policy	Business	Public
Articles								
Books	−0.13^{s}							
Methods	0.08^{s}	−0.20^{s}						
Products	−0.13^{s}	−0.12^{s}	0.23^{s}					
Academic	0.48^{s}	−0.01	0.02	−0.24^{s}				
Policy	0.01	−0.00	0.10^{s}	0.25^{s}	−0.07^{s}			
Business	−0.03	−0.07^{s}	0.20^{s}	0.38^{s}	−0.12^{s}	0.24^{s}		
Public	−0.12^{s}	0.19^{s}	0.04	0.21^{s}	−0.20^{s}	0.28^{s}	0.13^{s}	

Note: Reports pairwise correlations for the four subjective measures of researchers' intended output types (Journal articles; Books; Materials or Methods; Products) and the four subjective measures of researchers' intended audiences (Academic peers; Policymakers; Businesses and organizations; General public); ^{s} $p < 0.01$.

Appendix B. Model Details

B.1. Derivation of Policy Functions

The policy functions $\mathcal{R}(\mathbf{S}_i, \theta_i, \mu_i)$, $\mathcal{F}(\mathbf{S}_i, \theta_i, \mu_i)$ and $\mathcal{H}(\mathbf{S}_i, \theta_i, \mu_i)$ characterize the solution (R_i^*, F_i^*, H_i^*) to problem (1) for each individual scientist i as a function of states $\mathbf{S}_i$, attributes θ_i, and parameters μ_i. Moreover, they determine indirect utility $V_i^* = \mathcal{V}(\mathbf{S}_i, \theta_i, \mu_i)$ at current allocations.

Once substituting the constraints, we can express the utility-maximization problem as

$$\mathcal{V}(\cdot,\cdot,\cdot) = \max_{F_i, H_i} U_{1i}(M_i) + U_{2i}(\alpha_i (B_{\min} + G_i + \phi_i F_i)^{\gamma_i} (H_i - F_i - D_i)^{1-\gamma_i}) - U_{3i}(H_i, D_i) \tag{B1}$$

with the policy functions $\mathcal{F}(\cdot,\cdot,\cdot)$ and $\mathcal{H}(\cdot,\cdot,\cdot)$ solving the optimality conditions:

$$\frac{\partial U_{2,i}(Y_i)}{\partial Y_i}\frac{\partial Y_i}{\partial F_i} + \lambda_{F,i} = 0 \tag{B2}$$

$$\frac{\partial U_{2,i}(Y_i)}{\partial Y_i}\frac{\partial Y_i}{\partial H_i} - \frac{\partial U_{3,i}(H_i, D_i)}{\partial H_i} - \lambda_{H,i} = 0\,, \tag{B3}$$

and $\mathcal{R}(\cdot,\cdot,\cdot)$ being residually determined based on the time-constraint (1c).

We derive model's policy functions describing how scientists' optimal time allocation vary with state variables and parameters. We start from model's optimality conditions (B3) and (B3) and evaluate them using the chosen functional forms described in the main text. We obtain the conditions:

$$Y_i^{1-\eta_i}\Big[\gamma_i\phi_i(B_{\min} + G_i + \phi_i F_i)^{-1} - (1-\gamma_i)(H_i - D_i - F_i)^{-1}\Big] + \lambda_{F,i} = 0 \tag{B4}$$

$$Y_i^{1-\eta_i}(H_i - D_i - F_i)^{-1}(1-\gamma_i) - \psi(H_i - D_i + D_i^{\xi_i})^{\zeta_i} - \lambda_{H,i} = 0\,. \tag{B5}$$

We characterize the policy functions in two intervals of H_i. The first is $H_i \in (D_i, H_{i,F>0}]$ and the second is $H_i \in (H_{i,F>0}, H_{\max}]$, depending on whether the optimal time allocation features a corner solution for fundraising activity ($F_i = 0$) or an interior one ($F_i > 0$). We identify the threshold $H_{i,F>0}$ below which the optimal fundraising time is zero by solving (B5) for F_i as a function of H_i under the hypothesis that F_i is strictly positive, i.e., $\lambda_{F,i} = 0$:

$$F_i = \gamma_i(H_i - D_i) - \frac{1-\gamma_i}{\phi_i}(B_{\min} + G_i)\,. \tag{B6}$$

Therefore, (B6) implies that fundraising time is strictly positive as long as:

$$H_i > D_i + \frac{1-\gamma_i}{\gamma_i \phi_i}(B_{\min} + G_i)\ , \tag{B7}$$

which identifies the threshold $H_{i,F>0}$, which is strictly larger than D_i as long as $\gamma_i < 1$. As a consequence, we can solve for optimal hours H_i using equation (B5), which takes the piece-wise functional form:

(B8)

$$\begin{cases} (1-\gamma_i)[\alpha_i (B_{\min}+G_i)^{\gamma_i}]^{1-\eta_i}(H_i - D_i)^{(1-\gamma_i)(1-\eta_i)-1} - \psi(H_i - D_i + D_i^{\xi_i})^{\zeta_i} - \lambda_{H,i} = 0 & \text{if } H_i \le H_{i,F>0} \\ \left[\alpha_i(\phi_i\gamma_i)^{\gamma_i}(1-\gamma_i)^{1-\gamma_i}\right]^{1-\eta_i}\left(H_i - D_i\frac{B_{\min}+G_i}{\phi_i}\right)^{-\eta_i} - \psi(H_i - D_i + D_i^{\xi_i})^{\zeta_i} - \lambda_{H,i} = 0 & \text{if } H_i > H_{i,F>0} \end{cases}$$

and defines an implicit solution $\mathcal{H}_i$ for H_i as a function of all parameters and state variables. Specifically, (B8) is continuous but not differentiable at $H_{i,F>0}$. Moreover, the Lagrange multiplier $\lambda_{H,i}$ equals zero for $H_i < H_{\max}$. Finally, previous equation implies that H_i is always strictly larger than D_i because $\lim_{H_i \to D_i}$ (B8) $= +\infty$. Hence, the non-negativity constraint on research time is always met.

Given the policy function $\mathcal{H}_i$ for hours, the functions defining optimal allocation to fundraising time and research time can be derived as:

$$\mathcal{F}_i = \begin{cases} 0 & \text{if } \mathcal{H}_i \le H_{i,F>0} \\ \gamma_i(\mathcal{H}_i - D_i) - \frac{1-\gamma_i}{\phi_i}(B_{\min}+G_i) & \text{if } \mathcal{H}_i > H_{i,F>0} \end{cases} \tag{B9}$$

and

$$\mathcal{R}_i = \mathcal{H}_i - D_i - \mathcal{F}_i\ . \tag{B10}$$

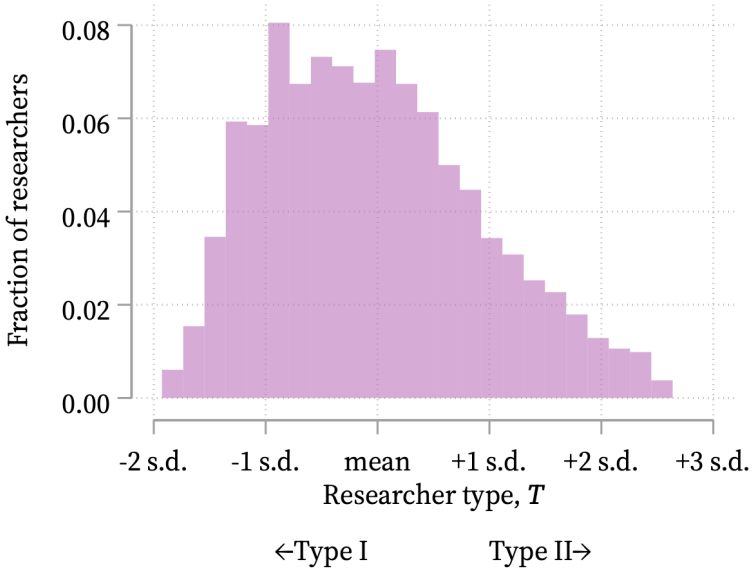
B.2. Reducing Dimensions of Heterogeneity: k-means Clustering Results

B.3. Estimation Algorithm

We proceed to describe the estimation algorithm for the common parameters $\tilde{\mu} = (\omega, \psi, \delta_{\sigma,0}, \delta_{\sigma,1}, \delta_{\eta,0}, \delta_{\eta,1}, \delta_{\xi,0}, \delta_{\xi,1}, \delta_{\zeta,0}, \delta_{\zeta,1})$.

We first define a multi-dimensional grid of parameter values at which we perform a preliminary evaluation of the loss function in problem (9). The grid is defined by the Cartesian product of

FIGURE B1
Researcher Heterogeneity



Note: Shows the distribution of the log-transformed and standardized euclidean similarity scores from the *k*-means cluster estimation with two *k*=I,II.

TABLE B1
Comparison of *k*-means Clusters

	mean I	*mean II*	*mean diff.*
Major field, {0,1}			
Engineering, math, & related	0.17	0.17	0.00
Humanities & related	0.15	0.21	0.06***
Medical & health sciences	0.35	0.24	–0.12***
Natural sciences	0.13	0.17	0.04***
Social sciences	0.20	0.21	0.02
Position details			
Full professor {0,1}	0.05	0.61	0.57***
Not tenure-track {0,1}	0.41	0.07	–0.33***
Has tenure {0,1}	0.00	0.91	0.91***
Socio-demographics			
Age [30-80]	41.17	53.02	11.85***
Female {0,1}	0.54	0.40	–0.14***
Married {0,1}	0.78	0.84	0.06***
Number dependents [0,5+]	0.99	0.98	–0.01
White race/ethnicity {0,1}	0.74	0.79	0.05***
US-born {0,1}	0.73	0.70	–0.02*
Personal risk-taking [0,10]	5.22	5.28	0.05
Rest-of-household earnings, $(0,\infty)$	96.83	103.80	6.97
Intended research outputs, {0,1,2}			
Journal articles	1.84	1.89	0.05***
Books	0.39	0.59	0.20***
Materials or methods	0.74	0.67	–0.07***
Products	0.51	0.45	–0.06***
Intended research audience, {0,1,2}			
Academic peers	1.84	1.90	0.06***
Policymakers	0.90	0.81	–0.09***
Businesses	0.57	0.52	–0.05**
General public	0.84	0.82	–0.02
Riskiness of own research, [0,10]			
Own belief	4.28	4.82	0.54***
Belief of peers' beliefs	4.24	4.75	0.51***
Theoretical vs. empirical, [0,10]			
Ask [0] or answer [10] questions	5.17	4.74	–0.43***

Note: Reports a t-test of the difference in means between the two predicted clusters following the *k*-means cluster estimation with k=2; * $p < 0.10$,** $p < 0.05$,*** $p < 0.01$.

the following vectors:

$$
\begin{aligned}
\omega^{(0)} &= [0.1, \quad 1, \quad 10] \\
\psi^{(0)} &= [0.00001, \quad 1, \quad 10] \\
\delta_{\sigma,0}^{(0)} &= [-100, \quad 0] \\
\delta_{\sigma,1}^{(0)} &= [0] \\
\delta_{\eta,0}^{(0)} &= [\ln(0.8) - \ln(0.2), \quad 0, \quad \ln(0.2 - \ln(0.8)] \\
\delta_{\eta,1}^{(0)} &= [0] \\
\delta_{\xi,0}^{(0)} &= [-11.5, \quad \ln(2)] \\
\delta_{\xi,1}^{(0)} &= [0] \\
\delta_{\zeta,0}^{(0)} &= [-11.5, \quad 0] \\
\delta_{\zeta,1}^{(0)} &= [0]
\end{aligned}
$$

thus including 216 grid-points.

Secondly, we select the grid-points where the value of the loss function is within 0.5% of the identified minimum value and we use them as initial points for the numerical solution to problem (9). We perform a preliminary search using a "Nelder-Mead simplex direct search" method with a high tolerance on the loss function and standard tolerance on parameter values. We then select the parameters vectors where the value of the loss function is within 1% of the identified minimum value and we use them as initial points for the numerical solution to problem (9) with the same search algorithm and with lower loss function tolerance. We repeat this step twice until we are left with a single candidate optimal parameter vector.

B.4. Objectives and Constraints for Counterfactuals

We first analyze a setting where social planner's objective is to maximize field-specific aggregate scientist's utility, i.e., the sum of field-f scientists' individual utility, by reallocating guaranteed funding G_i and administrative duties D_i within the field, conditional on their total amounts and the total value of additional funding being fixed to the observed endowments.

To this end, we define the output function

$$
\mathcal{Y}_i(\tilde{G}_i, \tilde{D}_i, \pi_f) = \alpha_i \left(B_{\min} + G_i + \phi_i \mathcal{F}(\tilde{G}_i, \tilde{D}_i, \pi_f) \right)^{\gamma_i} \mathcal{R}(\tilde{G}_i, \tilde{D}_i, \pi_f)^{1-\gamma_i}
$$

where the adjustment factor π_f to fundraising ability in field f enforces the constraint that additional funding in the new, counterfactual allocation must be equal to observed additional funding. For convenience, in this section we omit from the notation the dependence of the

policy functions on attributes $\widehat{\theta}_i$ and parameters $\widehat{\mu}_i$.

Therefore, the problem in field f is:

$$\max_{(\tilde{G}_i,\tilde{D}_i)_{i=1}^{N_f}} \sum_{i=1}^{N_f}\left(\kappa\frac{\mathcal{Y}_i(\tilde{G}_i,\tilde{D}_i,\pi_f)^{1-\eta_i}}{1-\eta_i} - \psi\frac{(\mathcal{H}_i(\tilde{G}_i,\tilde{D}_i,\pi_f)-\tilde{D}_i+\tilde{D}_i^{\xi_i})^{1+\zeta_i}}{1+\zeta_i}\right)$$

subject to

$$\begin{aligned}
&\sum_{i=1}^{N_f}\tilde{G}_i = \widehat{G}^{tot} \quad [\text{Multiplier } \lambda_{G,f}]\\
&\sum_{i=1}^{N_f}\tilde{D}_i = \widehat{D}^{tot} \quad [\text{Multiplier } \lambda_{D,f}]\\
&\sum_{i=1}^{N_f} EG_i = \pi_f \sum_{i=1}^{N_f}\widehat{\phi}_i\mathcal{F}_i(\tilde{G}_i,\tilde{D}_i,\pi_f)\ ,
\end{aligned} \tag{B11}$$

where N_f is the number of scientists in field f. We also impose non-negativity constraints on individual $\tilde{G}_i$ and $\tilde{D}_i$ allocations. The individual research output function $\mathcal{Y}_i = \widehat{\alpha}_i\mathcal{B}_i^{\widehat{\gamma}_i}\mathcal{R}_i^{1-\widehat{\gamma}_i}$ and hours $\mathcal{H}_i$ make explicit that both are functions of the considered instruments $(\tilde{G}_i,\tilde{D}_i)$. The last constraint requires that total additional funding in the observed allocations, i.e., the term on the left hand side, equals additional funding in the optimal allocations conditional on the field-specific adjustment factor π_f. Therefore, all the policy functions—i.e., researchers' behavioral responses to the planner's decisions—not only vary with policy levers $(\tilde{G}_i,\tilde{D}_i)$ but also with π_f, which can be interpreted as an endogenous adjustment to fundraising probability. When the new allocation violates the constraint, fundraising probability uniformly shrinks, thus reducing additional funding both directly and indirectly through the behavioral decline in fundraising hours.

In the optimal allocations, the planner seeks to equate the marginal utility of guaranteed funding and duties across individuals, conditional on the non-negativity constraint not being hit and the additional funding constraint being met. Therefore, the solution of the problem solves, for each $i = 1, \ldots, N_f$ the following equations:

$$\frac{d\mathcal{V}_i}{d\tilde{G}_i} = \kappa\mathcal{Y}_i^{-\eta_i}\frac{d\mathcal{Y}_i}{d\tilde{G}_i} - \psi(\mathcal{H}_i-\tilde{D}_i+\tilde{D}_i^{\xi_i})^{\zeta_i}\frac{\partial\mathcal{H}_i}{\partial\tilde{G}_i} = \lambda_{G,f}-\lambda_{G^*,i} \tag{B12}$$

$$\frac{d\mathcal{V}_i}{d\tilde{D}_i} = \kappa\mathcal{Y}_i^{-\eta_i}\frac{d\mathcal{Y}_i}{d\tilde{D}_i} - \psi(\mathcal{H}_i-\tilde{D}_i+D_i^{\xi_i})^{\zeta_i}\left(\frac{\partial\mathcal{H}_i}{\partial\tilde{D}_i}-1+\xi_i\tilde{D}_i^{\xi_i-1}\right) = \lambda_{D,f}-\lambda_{D^*,i}\ , \tag{B13}$$

where the marginal product of guaranteed funds $\frac{d\mathcal{Y}_i}{d\tilde{G}_i}$ and the (negative) marginal product of

duties $\frac{d\mathcal{Y}_i}{d\tilde{D}_i}$ are determined by equations (B14) and (B15), respectively:

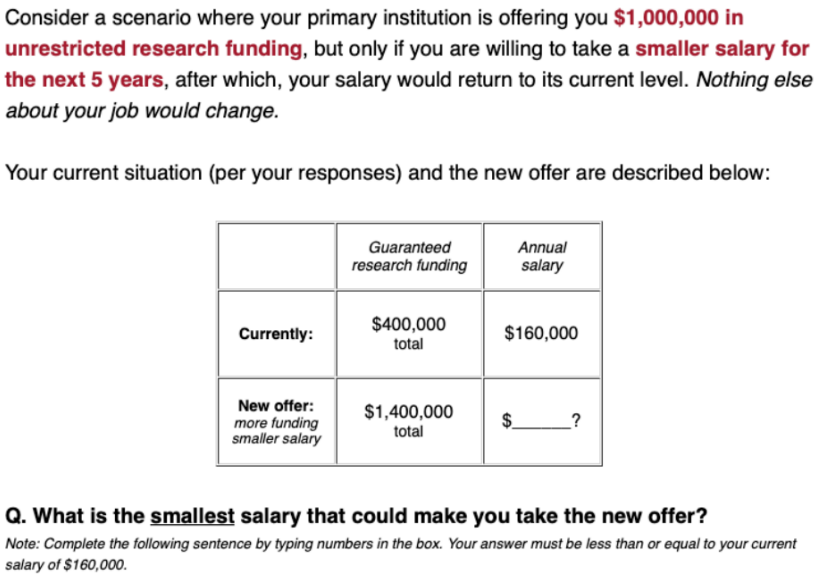
$$\frac{d\mathcal{Y}_i}{d\tilde{G}_i} = \widehat{\alpha}_i \mathcal{B}_i^{\widehat{\gamma}_i} \mathcal{R}_i^{1-\widehat{\gamma}_i} \left(\widehat{\gamma}_i \mathcal{B}_i^{-1} \frac{\partial \mathcal{B}_i}{\partial \tilde{G}_i} + (1-\widehat{\gamma}_i) \mathcal{R}_i^{-1} \frac{\partial \mathcal{R}_i}{\partial \tilde{G}_i} \right) \tag{B14}$$

$$\frac{d\mathcal{Y}_i}{d\tilde{D}_i} = \widehat{\alpha}_i \mathcal{B}_i^{\widehat{\gamma}_i} \mathcal{R}_i^{1-\widehat{\gamma}_i} \left(\widehat{\gamma}_i \mathcal{B}_i^{-1} \frac{\partial \mathcal{B}_i}{\partial \tilde{D}_i} + (1-\widehat{\gamma}_i) \mathcal{R}_i^{-1} \frac{\partial \mathcal{R}_i}{\partial \tilde{D}_i} \right). \tag{B15}$$

Total funding as a function of state variables is defined by $\mathcal{B}_i(G_i, D_i, \pi_f) = B_{\min} + G_i + \pi_f \widehat{\Phi}_i \mathcal{F}_i(G_i, D_i, \pi_f)$.

Appendix C. Additional Experiment Details

FIGURE C1
Example Image of Survey Experiment



Consider a scenario where your primary institution is offering you **$1,000,000 in unrestricted research funding**, but only if you are willing to take a **smaller salary for the next 5 years**, after which, your salary would return to its current level. *Nothing else about your job would change.*

Your current situation (per your responses) and the new offer are described below:

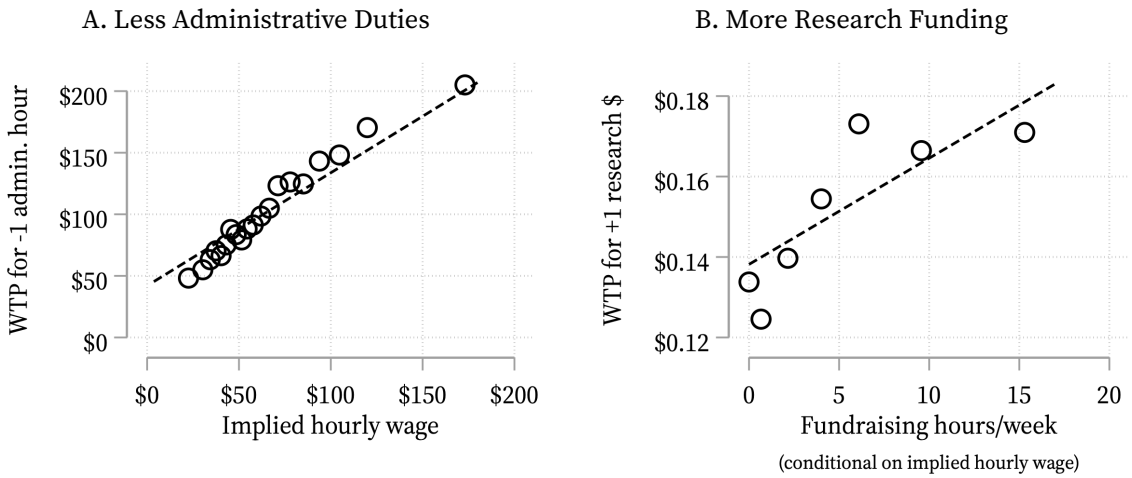
	Guaranteed research funding	*Annual salary*
Currently:	$400,000 total	$160,000
New offer: *more funding smaller salary*	$1,400,000 total	$_____?

Q. What is the smallest salary that could make you take the new offer?

Note: Complete the following sentence by typing numbers in the box. Your answer must be less than or equal to your current salary of $160,000.

Note: Shows a screenshot of the survey experiment designed to solicit researchers' willingness to trade off their salary for additional guaranteed funding.

FIGURE C2
Correlations of Implied Valuations



Note: Shows a binned scatterplot and line-of-fit for the relationship between researchers' implied hourly wage and their willingness to pay for 1 less hour of administrative duties (Panel A), and the relationship between how many hours per week a researcher spends on fundraising and their willingness to pay for $1 more of additional research funds conditional on their implied hourly wage (Panel B).

TABLE C1
Proportion of Response Variation Correlated with Model and Non-model Variables

	WTP +$250K		WTP +$1M		WTP +admin.		WTP –admin.	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
R^2	0.95	0.96	0.87	0.89	0.82	0.86	0.96	0.97
Model vars.	✓	✓	✓	✓	✓	✓	✓	✓
X		✓		✓		✓		✓
N obs.	4,003	4,003	4,003	4,003	4,003	4,003	4,003	4,003

Note: Reports the R^2 statistics from regressions of researchers' responses to the four experiments (i.e., their willingness to pay for the alternative scenarios) on different combinations of variables: *Model vars.* includes the state and choice variables of the model; **X** includes the full vector of variables that comprise the type index.

TABLE C2
Potential Role of Non-response Bias and Stated Preference WTP Bias

	WTP +$250K		WTP +$1M		WTP +admin.		WTP –admin.	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
IMR		–0.000858		–0.00391		–0.00466		–0.000835
		(0.00379)		(0.00593)		(0.00684)		(0.00327)
Benchmark		–0.0126***		–0.0258***		0.0134*		–0.0102***
		(0.00411)		(0.00644)		(0.00742)		(0.00355)
R^2	0.95	0.95	0.87	0.87	0.82	0.82	0.96	0.96
Model vars.	✓	✓	✓	✓	✓	✓	✓	✓
N obs.	4,003	4,003	4,003	4,003	4,003	4,003	4,003	4,003

Note: Reports the estimates from regressions of researchers' responses to the four experiments (i.e., their willingness to pay for the alternative scenarios) on a control for non-response bias in the form of the Inverse Mills Ratio (IMR) and a control for bias in individuals' stated WTP in the form of their WTP for the benchmark good (i.e., high-speed internet at home); all variables are standardized. *Model vars.* includes the state and choice variables of the model. Robust standard errors reported; * $p < 0.10$,** $p < 0.05$,*** $p < 0.01$.

Appendix D. Additional Productivity and Efficiency Results

TABLE D1
Utility Function Parameter Estimates

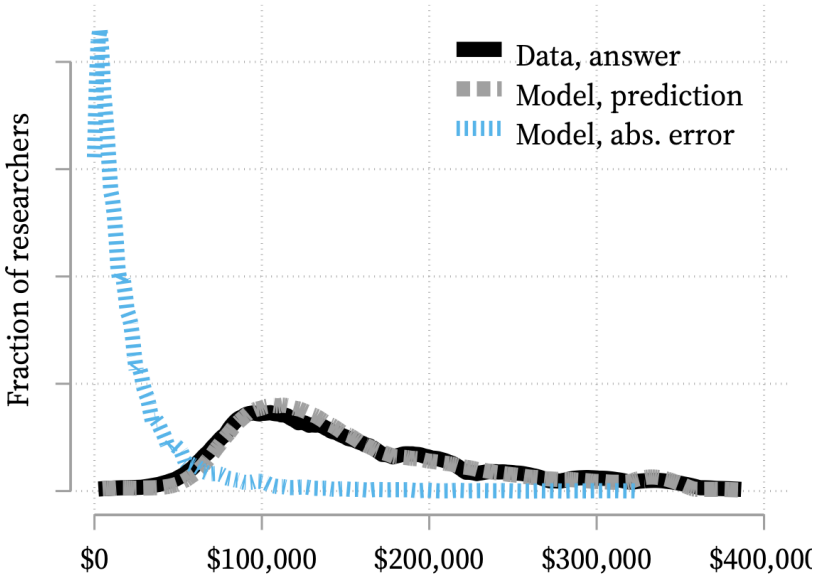
	mean	*s.d.*	*p*10	*p*50	*p*90
ω	4.2e-04				
σ_i	1.835	0.0338	1.795	1.831	1.882
η_i	0.257	0.0303	0.2217	0.2529	0.2994
ψ	5.9e-15				
ξ_i	1.428	0.0484	1.36	1.432	1.487
ζ_i	3.9e-06	4.0e-07	3.4e-06	3.9e-06	4.4e-06

Note: Reports summary statistics for the parameter estimates; if only means are reported, the parameter is assumed to be homogeneous. Parameter definitions are as follows:

parameter	component
ω	utility from income
σ	utility from income, curvature
η	utility from output, curvature
ψ	disutility from total work time
ξ	disutility from admin., teaching, and other, add.'l curvature
ζ	disutility from total work time, curvature

FIGURE D1
Model Fit

A. Answer, Prediction, and Error Distributions

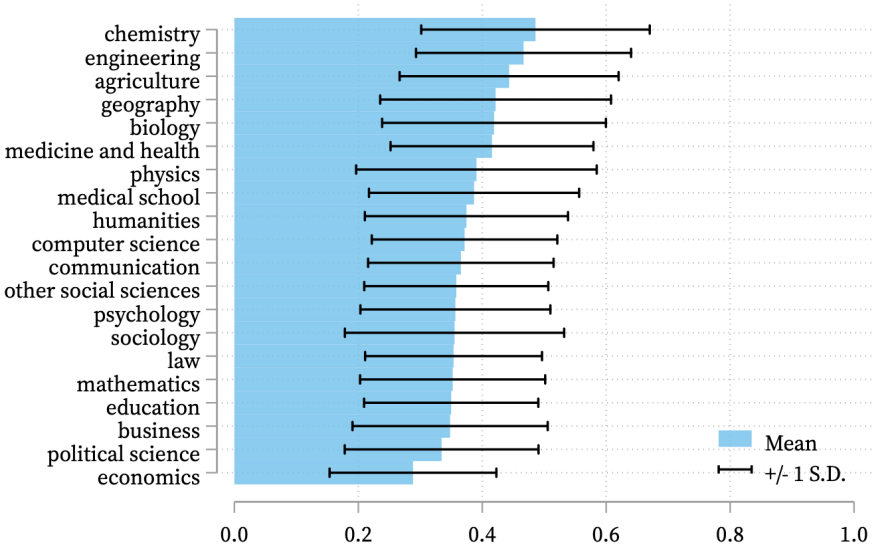


B. Distributions of Question-specific Errors

	*p*10	*p*25	*p*50	*p*75	*p*90
G +$250,000	0.11	0.88	6.18	15.54	29.58
G +$1,000,000	0.63	3.42	12.25	27.07	53.71
A +20 hr./month	2.19	5.56	11.40	21.01	35.66
A –all hr./month	0.26	2.41	6.17	12.75	22.01

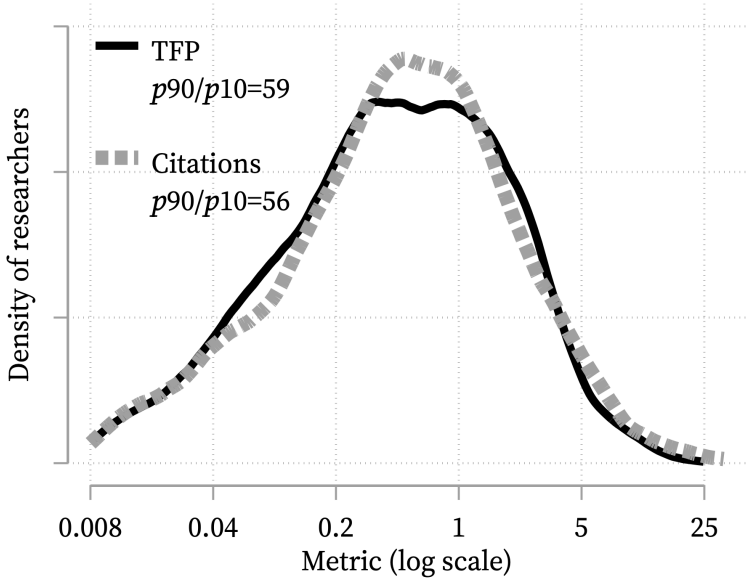
Note: Panel (A) shows the distribution of researchers' responses to the four experiments (i.e., their stated annual salary at indifference) alongside the distribution of the model's predictions and absolute errors. Panel (B) reports summary statistics for the distribution of the absolute errors for each of the four experiments where the model's errors are expressed in percentage points of the actual responses.

FIGURE D2
Research Funding Intensity (γ) by Minor Field of Study



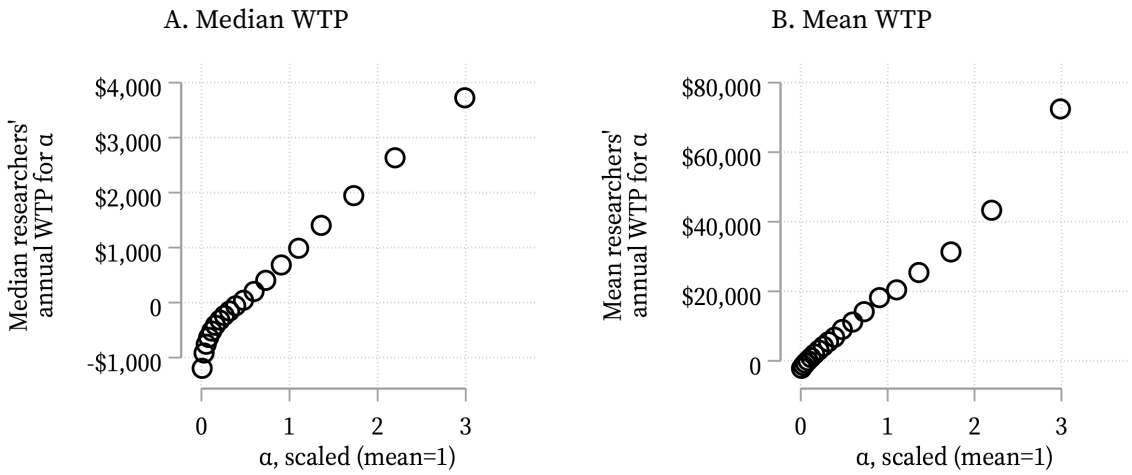
Note: Shows the mean and standard deviations of researcher-specific estimates of the production function parameter γ_i (funding intensity) split by minor field of study.

FIGURE D3
Productivity and Citation Distribution Comparison



Note: Shows the distribution of researchers' TFP (α_i) and their field-normalized citations, where both metrics have been rescaled so the mean equals 1; note the log scale. 90-10 percentile ratios are also shown.

FIGURE D4
Dollar Conversions of Research Productivity



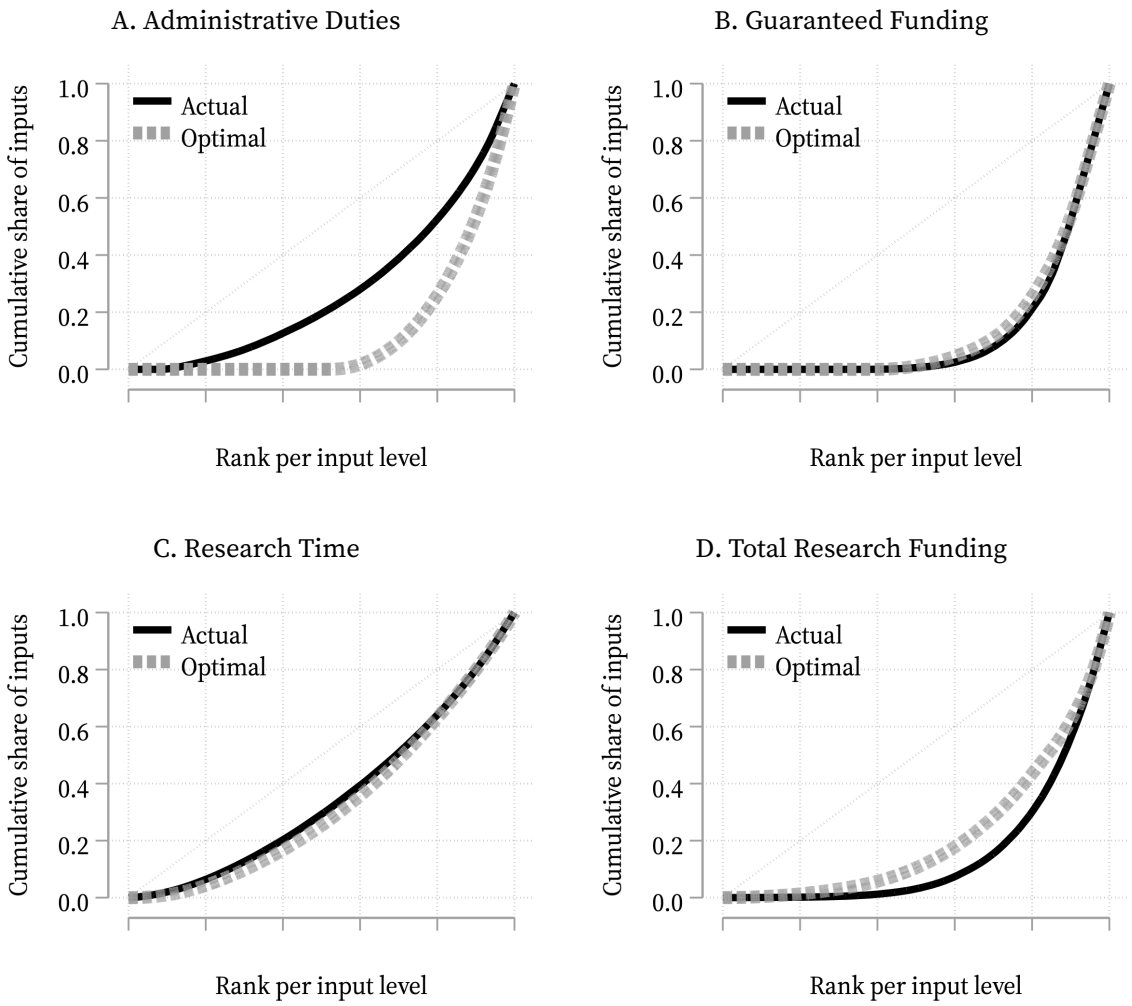
Note: Shows the implied dollar value of productivity in terms of the median (Panel A) or mean (Panel B) of in-sample researchers' willingness to pay for a given productivity level, where productivity (x axis) is scaled so that the sample mean equals one. For example, Panel (A) indicates that the median amount that researchers would be willing to pay to have average productivity is approximately $1,000 per year. Note, negative values imply that researchers must be paid to receive a given productivity level (because it is lower than their current productivity).

TABLE D2
Observable Output Correlations

	Recent publications, count		Recent publications, cite-weighted	
	(1)	(2)	(3)	(4)
$\log(\alpha)$	0.201***	0.176***	0.276***	0.254***
	(0.0406)	(0.0397)	(0.0648)	(0.0648)
log(recent grant $)		0.0220***		0.0174***
		(0.00350)		(0.00525)
Field–FE, γ, ϕ, **X**–index	✓	✓	✓	✓
R^2	0.18	0.20	0.15	0.16
N obs.	2,703	2,703	2,703	2,703

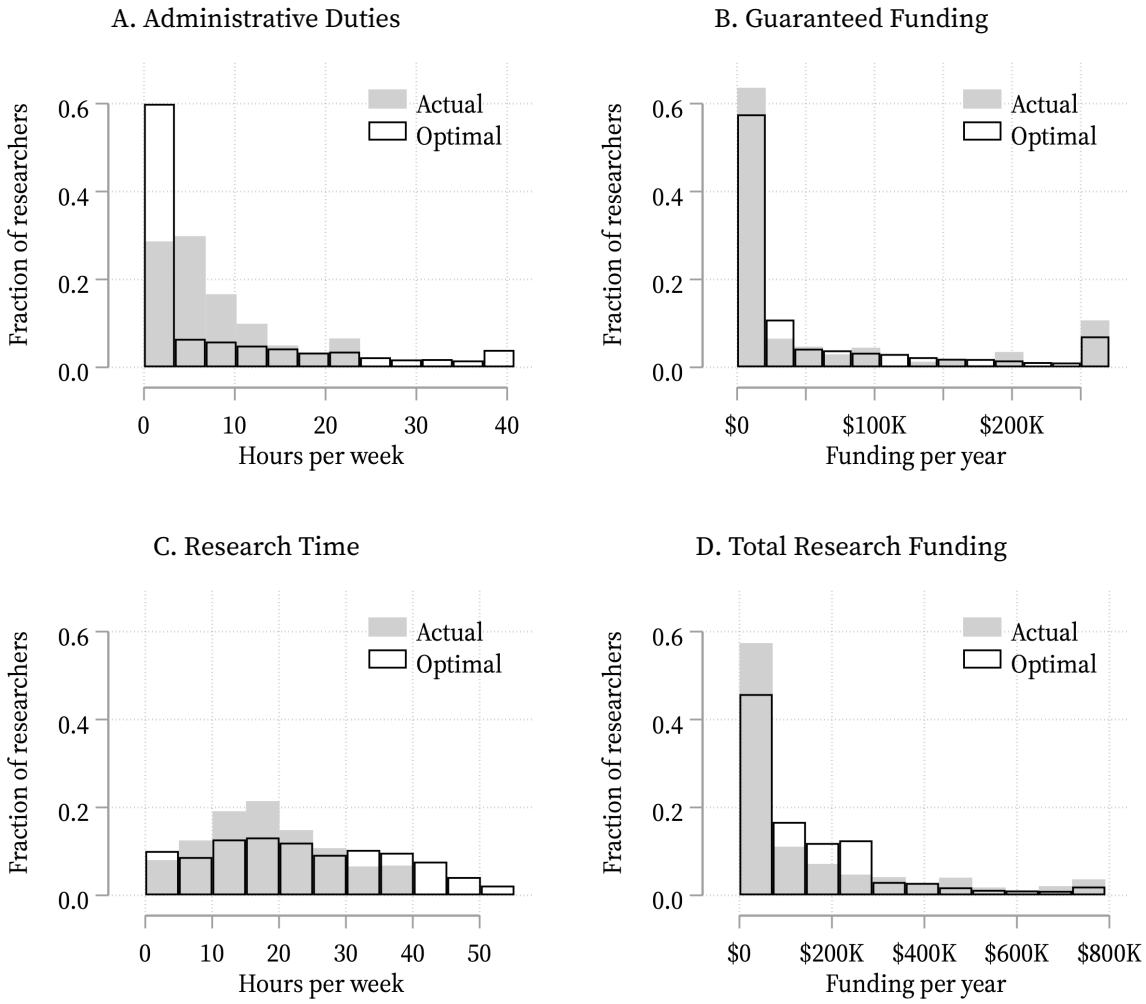
Note: Reports the estimates from regressions of researchers' publication measures (including publications from 2018–2022) on their estimated research productivity (α) as well as vector of controls that includes major field fixed effects (*Field-FE*) and the researchers' funding intensity (γ), fundraising efficiency (ϕ), their type (**X**-*index*), and, in some specifications, a control for their publicly observable research grant funding over the same period. Robust standard errors reported; * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

FIGURE D5
Lorenz Curves for Actual and Optimal Input Levels



Note: Shows Lorenz curves for actual and optimal input levels. Researchers are sorted on the x axis per their ranking in terms of how much of each input they have, and the lines plot the cumulative share of total inputs (per the y axis) summing from the lowest- to highest-ranked researcher. Thus, plots closer to the 45^o line indicate input allocations closer to equality.

FIGURE D6
Actual Versus Optimal Input Levels



Note: Shows the distribution of actual and optimal input levels. Panels (a–b) show the distribution of the inputs assumed to be directly under the planners' control. Panels (c–d) show the distribution of inputs after researchers' choose their time allocations in response to the planner.

TABLE D3
Production with Actual and Optimized Allocations—Alternative Counterfactuals

	Current allocation	Optimized allocations				
	(1)	(2)	(3)	(4)	(5)	(6)
Research inputs						
Research hrs./week, avg.	18.5	–11%	+18%	+14%	+25%	+16%
Research hrs./week, s.d.	9.6	+6.0%	+18%	+43%	+39%	+41%
Budget $-K/year, avg.	147.1	0%	0%	0%	0%	+14%
Budget $-K/year, s.d.	206.9	+14%	–24%	–20%	–20%	–3.9%
Research output						
Output, avg.	*n.r.*	+1.3%	+150%	+160%	+160%	+160%
Output, s.d.	*n.r.*	+0.7%	+43%	+44%	+44%	+44%
Output per hr.	*n.r.*	+14%	+110%	+130%	+110%	+120%
Output per $	*n.r.*	+1.4%	+150%	+160%	+160%	+130%
Welfare						
Researcher utility, avg.	*n.r.*	+3.6%	+1.0%	+4.7%	+2.1%	+4.7%
Researcher utility, s.d.	*n.r.*	–5.7%	+0.3%	–5.4%	–2.4%	–5.4%
Researcher utility per hr.	*n.r.*	–8.3%	+16%	+16%	+21%	+18%
Researcher utility per $	*n.r.*	+3.6%	+1.0%	+4.8%	+2.3%	+16%
Social value, avg.	*n.r.*	+3.4%	+12%	+16%	+13%	+16%
Social value, s.d.	*n.r.*	–4.0%	+23%	+20%	+21%	+20%
Social value per hr.	*n.r.*	–8.5%	+25%	+26%	+30%	+28%
Social value per $	*n.r.*	+3.4%	+12%	+16%	+13%	+26%
Input reallocation						
Research hrs./week		9%	19%	27%	24%	25%
Budget $-K/year		14%	27%	28%	25%	25%
Objective, max		$\mathcal{V}$	$\mathcal{V}$	$\mathcal{V}$	Y	$\mathcal{V}$
Reallocate D		✓		✓	✓	✓
Reallocate G			✓	✓	✓	✓
Unconstrained B						✓

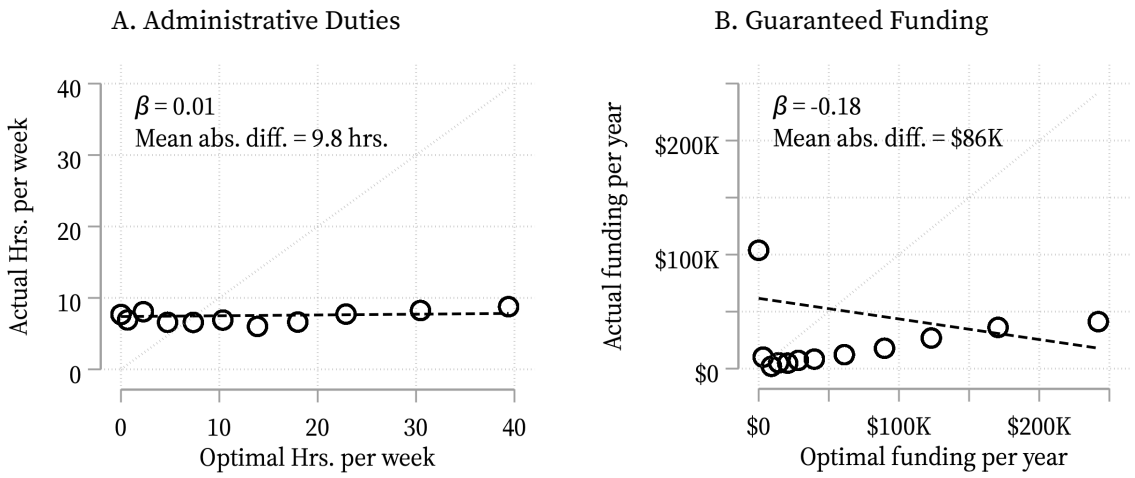
Note: Reports summary statistics for inputs under actual allocations (Col. 1). The first three sets of rows in Columns 2–6 report the percentage change in research inputs (*Research inputs*), outputs (*Research outputs*), and utility (*Welfare*) under alternative allocations; estimates are rounded to aid in comparison. The *Input Reallocation* rows report the amount of inputs reallocated expressed as a percentage of the total level of the input (e.g., 50% implies that half of all dollars are moved from one researcher to another). The bottom sets of rows outline the objective and constraints of the five different counterfactual allocations explored in Columns 2–6. The two different objectives explored are maximizing researchers' private utility ($\mathcal{V}$) or output (Y). D refers to administrative duties, and G refers to guaranteed research funding. *Unconstrained B* indicates the scenario when the total research budget is left unconstrained and so the total amount of funding in the market is limited only by researchers' fundraising choices. All optimized allocations allow for researchers' behavioral responses after D and/or G have been reallocated.

TABLE D4
Production with Actual and Optimized Allocations—Alternative Specifications

	Current allocation	Optimized allocations			
	(1)	(2)	(3)	(4)	(5)
Research inputs					
Research hrs./week, avg.	18.5	+14%	+30%	+6.5%	+14%
Research hrs./week, s.d.	9.6	+43%	+60%	+27%	+44%
Budget $-K/year, avg.	147.1	0%	0%	0%	0%
Budget $-K/year, s.d.	206.9	–20%	–22%	–22%	–24%
Research output					
Output, avg.	*n.r.*	+160%	+160%	+250%	+170%
Output, s.d.	*n.r.*	+44%	+45%	+180%	+44%
Output per hr.	*n.r.*	+130%	+100%	+230%	+130%
Welfare					
Researcher utility, avg.	*n.r.*	+4.7%	+4.8%	+250%	+4.8%
Researcher utility, s.d.	*n.r.*	–5.4%	–5.1%	+180%	–5.5%
Researcher utility per hr.	*n.r.*	+16%	+27%	+230%	+17%
Social value, avg.	*n.r.*	+16%	+17%	+250%	+17%
Social value, s.d.	*n.r.*	+20%	+20%	+180%	+20%
Social value per hr.	*n.r.*	+26%	+36%	+230%	+27%
Input reallocation					
Research hrs./week		27%	32%	27%	28%
Budget $-K/year		28%	28%	31%	31%
Absolute input cap			✓		
Homog. param.				✓	
Across-field *B*					✓

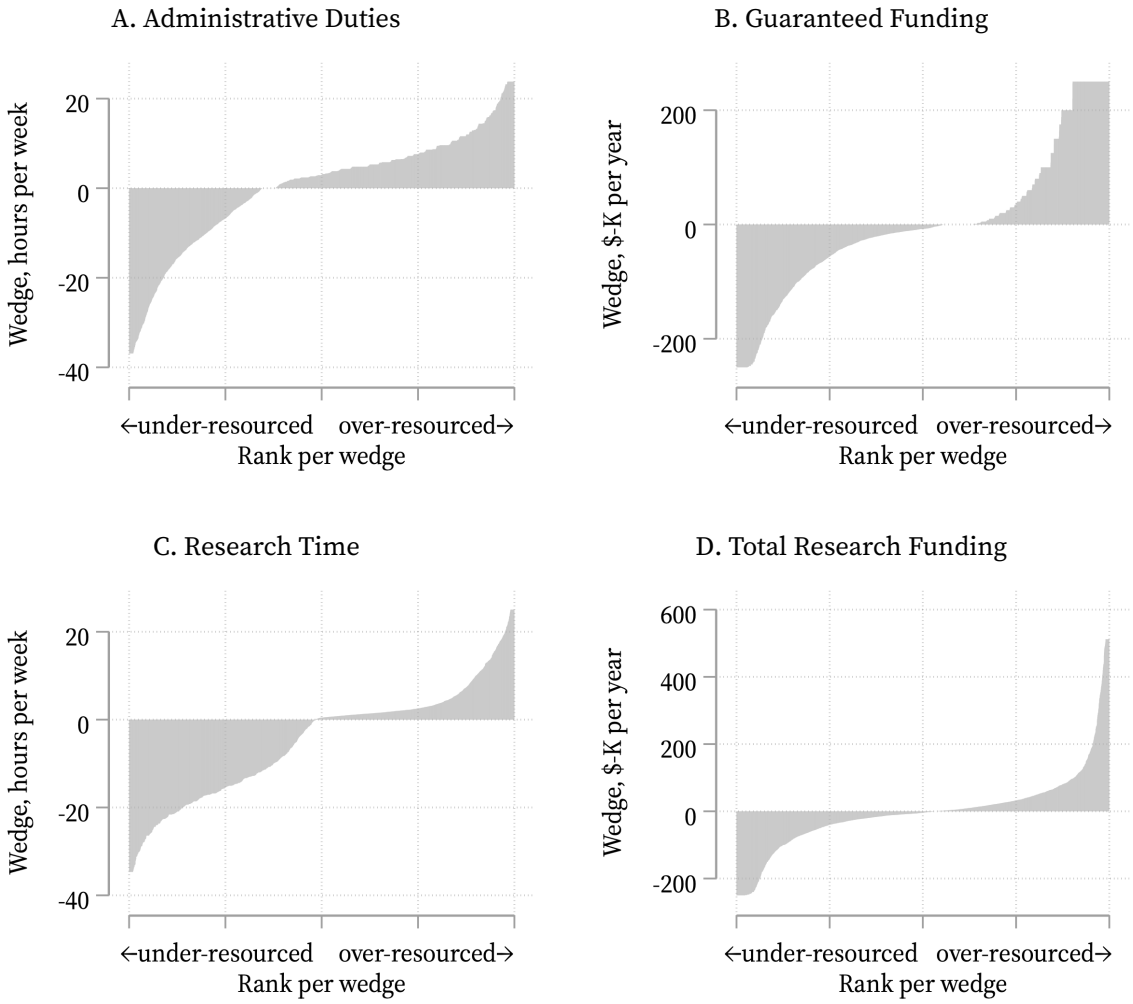
Note: Reports summary statistics for inputs and outcomes under actual allocations (Col. 1) and the utility-maximizing allocation (Cols. 2–5). Column (2) reports the main results from our preferred specification. Columns (3–5) report results under alternative specifications of the model and its estimation. *Absolute input cap* refers to the scenario where researchers' behavioral responses are limited with a global absolute maximum (as opposed to a relative maximum in the preferred scenario). *Homogeneous param.* refers to the scenario where all non-production-function parameters are homogeneous. *Across field B* allows for the reallocation of inputs across the five major fields of study (which is not allowed in the preferred scenario).

FIGURE D7
Actual Versus Optimal Input Correlations—Policy Levers



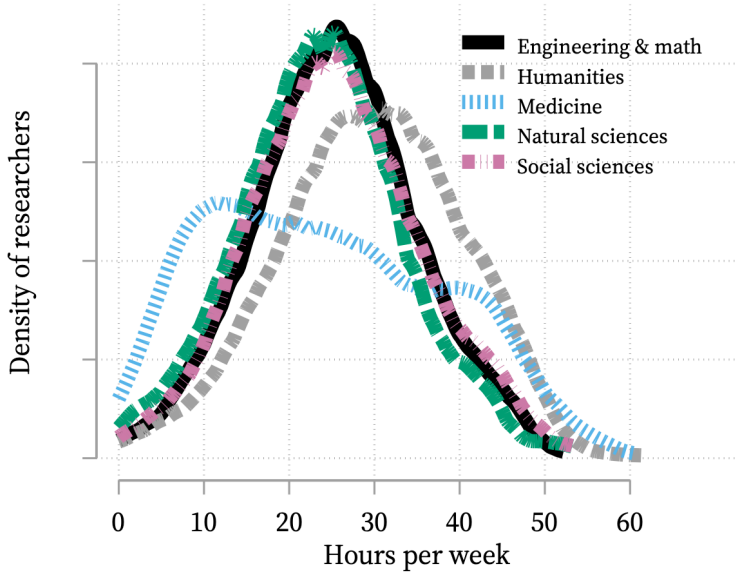
Note: Shows the binned scatterplot of actual and optimal input levels per an objective of maximizing output. Also reports the coeffificient estimate from a regression of actual on optimal input levels and the mean absolute difference between actual and optimal input levels. Based on the counterfactual described in Column (2) of Table 3.

FIGURE D8
Input Wedge Distributions



Note: Shows the distribution of the input wedge, which equals a researcher's actual input level minus their optimal input level; positive values indicate the researcher is over-resourced and negative values indicate the researcher is under-resourced relative to the optimum.

FIGURE D9
Teaching, Administration, Clinical and Other Duties by Major Field



Note: Shows the distribution of researchers' hours per week spent on teaching, administration, clinical, or other duties by major field.

TABLE D5
Input Wedges and Gender Differences

	(1)	(2)	(3)	(4)	(5)
Panel (a): Actual research time					
Optimal level	0.301***	0.294***	0.300***	0.352***	0.349***
	(0.00970)	(0.00989)	(0.00991)	(0.00985)	(0.0102)
Female				–1.333***	–0.885***
				(0.286)	(0.279)
R^2	0.18	0.22	0.26	0.24	0.32
X: position		✓			✓
X: output			✓		✓
X: socio–demog.				✓	✓
N obs.	4,003	4,003	4,003	4,003	4,003
Panel (b): Actual research funding					
Optimal level	0.987***	0.982***	0.928***	0.985***	0.932***
	(0.00856)	(0.00883)	(0.0102)	(0.00939)	(0.0109)
Female				–12111.4***	–7907.5**
				(3950.8)	(3872.1)
R^2	0.65	0.65	0.67	0.67	0.69
X: position		✓			✓
X: output			✓		✓
X: socio–demog.				✓	✓
N obs.	4,003	4,003	4,003	4,003	4,003

Note: Reports results from regressions of actual input levels on optimal input levels and as described in Equation 10. Robust standard errors reported; * $p < 0.10$,** $p < 0.05$,*** $p < 0.01$.

TABLE D6
Input Wedges per Common Productivity Proxies

	(1)	(2)	(3)	(4)
Panel (a): Actual research time				
Optimal level	0.508***	0.506***	0.509***	0.508***
	(0.0164)	(0.0166)	(0.0162)	(0.0163)
Recent research funding, own	0.130***			
	(0.0311)			
Recent research funding, institution		−0.0137		
		(0.0139)		
Recent publications			0.177***	
			(0.0197)	
Recent citations				0.148***
				(0.0223)
R^2	0.32	0.30	0.33	0.32
X: position, output, socio–demog.	✓	✓	✓	✓
N obs.	3,072	3,057	3,072	3,072
Panel (b): Actual research funding				
Optimal level	0.746***	0.763***	0.744***	0.753***
	(0.0110)	(0.00974)	(0.0106)	(0.0101)
Recent research funding, own	0.0767***			
	(0.0264)			
Recent research funding, institution		0.00202		
		(0.00958)		
Recent publications			0.0760***	
			(0.0153)	
Recent citations				0.0592***
				(0.0142)
R^2	0.70	0.69	0.69	0.69
X: position, output, socio–demog.	✓	✓	✓	✓
N obs.	3,072	3,057	3,072	3,072

Note: Reports results from regressions of actual input levels on optimal input levels and common proxies for researchers' producitivities. All variables are standardized. All proxies are based on data from one to two years prior to the survey. Robust standard errors reported; * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.