Productivity Beliefs and Efficiency in Science

Fabio Bertolotti
Bank of Italy
& CEPR

Kyle R. Myers
Harvard University
& NBER

Wei Yang Tham
University of
Toronto

March 2026

[*Comments welcome*]

Many contend the allocation of resources across scientists is inefficient. But the extent of this inefficiency remains unclear because outputs and prices are challenging to measure. We develop a method to overcome these challenges and evaluate the market for science. Our model of researchers' labor supply shows how their willingness to pay for inputs reveals their productivity beliefs. We estimate the model's parameters using data from a nationally representative survey and find the distribution of productivity to be very skewed. The counterfactuals indicate that a more efficient allocation of the current budget would be equivalent, at the population level, to billions of dollars of research funding.

Fabio Bertolotti: fabiobertolotti@hotmail.it. Kyle Myers: kmyers@hbs.edu. Wei Yang 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 the authors. This project would not have been possible without the excellent work of Rachel Mural, Nina Cohodes, and Yilun Xu and input from Marie Thursby, Jerry Thursby, and Karim Lakhani. We received helpful comments from Stefano Baruffaldi, Nick Bloom, Judy Chevalier, Matt Clancy, Zoë Cullen, Shane Greenstein, Matt Grennan, Daniel Gross, Bronwyn Hall, Sabrina Howell, Bruce Kogut, Petra Moser, David Rivers, Jesse Shapiro, Tim Simcoe, Chad Syverson, Joel Waldfogel, as well as seminar participants at Columbia, JOMT, MIT, 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, IIOC, Washington University in St. Louis, the Conference on the Economics of Innovation in Memory of Zvi Griliches, and Catholic University of Milan.

1. Introduction

Science has continuously proven to be the engine of economic growth.[1] Academic science, in particular, is now a $100 billion engine for the United States (e.g., Azoulay et al. 2019; Fleming et al. 2019; Gibbons and NCSES 2024). However, there is growing evidence that science is getting harder (Jones 2009; Bloom et al. 2020), and there are a growing number of debates about the functioning of scientific institutions.[2] Understanding productivity and efficiency in science has become increasingly important.

A long line of work, which we review below, has documented allocative inefficiencies in science; researchers are denied resources for reasons unrelated to their productivity. While it is clear we can do better at allocating resources, it is unclear *how much* better we can do. How much more knowledge could be produced if we improved the allocation of resources to best support the most talented scientists? Answering this question first requires answering another question: what is the distribution of marginal returns across researchers? In this paper, we formalize and answer both of these questions.

Progress on these questions has been slow because estimating productivity in science poses some major challenges. First, the output of science is difficult to observe; how does one quantify a unit of knowledge? Bibliometric data is popular; however, as noted by Adams and Griliches (1998), "*what constitutes a scientific paper makes for an elastic yardstick of scientific achievement*." Second, scientific inputs are rarely allocated via price mechanisms and researchers face large adjustment costs (Myers 2020; Baruffaldi and Gaessler 2025). These facts violate key assumptions at the core of most conventional methods for productivity estimation.

To overcome these challenges, we develop a method for estimating productivity without data on output quantities or input prices. As in conventional methods, we leverage the fact that producers' input choices reveal information about their productivity. However, our approach involves a new step: the direct solicitation of producers' willingness to pay (WTP) for inputs.

We assume that each individual producer (i.e., a scientist) derives indirect utility from their output and other sources of consumption, and we allow their total factor productivities, technologies (i.e., factor mix), and preferences to be idiosyncratic. Producers also have idiosyncratic budgets (i.e., salaries), which they can use for consumption or use to

[1] For motivating theory, see: Jones (2005). For a recent review, see: Bryan and Williams (2021).

[2] For coverage in popular press, see: Rohn (2017); Ioannidis (2018); Piper (2019); Mills (2021).

purchase inputs from us, the experimenters. The key is that, conditional on individuals' budgets, technologies, and preferences, one producer will only have a higher WTP compared to another if they believe they are more productive. Thus, the variation in WTP for inputs can then be used to recover the distribution of individual-level productivity and, through counterfactuals, quantify the value from more efficiently allocating inputs.

In order to apply this method to the market for science, we first present a new model of researchers' labor supply and consumption decisions. Researchers have heterogeneous preferences and derive utility from three sources: their salary, their scientific output (i.e., the quantity of knowledge they produce), and their leisure time. An (exogenous) contract specifies their salary, administrative duties, and guaranteed funding. Given their contract, researchers choose how to allocate their time across fundraising (e.g., writing grants)[3], research, and leisure given their beliefs, budgets, and constraints. Each researcher has 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.

To identify researchers' WTP for inputs, we use the logic of compensating variation to solve for the salary change that would leave the researcher indifferent to alternative contracts, where these alternatives specify counterfactual amounts of guaranteed research funding or administrative duties. This variation in salary — our measure of researchers' WTP for inputs — depends on productivity because the expected change in scientific output, given the counterfactual funding or duties, depends on the researcher's productivity beliefs.

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).[4] 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. Throughout the paper, we define scientific markets as the

[3]We allow for heterogeneity in researchers' ability to fundraise.

[4]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, grant funding, and publication rates; some of this is reprinted in the Appendix.

interaction of fields of study and career-stage. We focus on variation and counterfactuals *within* these markets to avoid heterogeneity that is beyond the scope of our model.

Researchers' willingness-to-pay responses are of plausible magnitudes and behave as expected. 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 we find little evidence that the responses are driven by sample selection or systematic noise. Furthermore, researchers' willingness to pay for free time is strongly correlated with their implied hourly wages as expected.

Our estimates of productivity beliefs vary widely across researchers, even after accounting for outliers. Within the field×career-stage markets, the ratio of the 90^{th} and 10^{th} percentiles of TFP is approximately 30. 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 sign of face validity, our model-based estimates of scientific output are positively correlated with popular bibliometric proxies for output (e.g., citations) even after conditioning on input levels. Interestingly, our estimates indicate that approximately half of the variance in scientific output across individual researchers is due to variance in their productivity.[5] The high degree of dispersion suggests that it may be hard for the market to facilitate positive selection on productivity.

To evaluate allocative efficiency, we compare observed allocations to two alternatives. The first is a first-best, where a planner directly assigns funding and research time within each market to maximize social welfare. The second is a uniform-policy benchmark, where researchers within the same market face the same guaranteed funding and fundraising ability and then re-optimize their choices. While this allocation is likely suboptimal, it is useful to assess the role of heterogeneous funding opportunities for misallocation.

Compared to the observed allocation, the first-best raises average output by roughly 63% (in value terms), private utility by 0.5%, and social welfare by 2.9%. The uniform-policy benchmark captures a substantial share of these gains, delivering roughly 70% of the total

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

welfare improvement from the first-best. To provide a more tangible valuation of the gains from reallocation, we ask how much guaranteed funding would need to increase under the observed allocation structure to match the welfare gains from the counterfactuals. We find that doing so would require more than a doubling of research budgets. On the scale of our focal population, this amounts to tens of billions of research dollars per year.

Finally, we examine which observable characteristics predict researchers' input wedges. A large vector of observable features explains only a small share of the cross-sectional variation in these wedges. Still, we zoom in on gender disparity, and find female researchers are observed to receive less funding and research time than is optimal, and most of this gap can be accounted for by heterogeneity in guaranteed funding and fundraising ability.

Our estimates come with some important caveats. First, we face a set of challenges common to all existing productivity estimation techniques.[6] Second, we face some unique limitations: (*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. Throughout the paper we engage with these limitations by providing robustness tests of our assumptions and face validity tests of the data. Overall, the results from these tests give us confidence in our conclusions. Still, we interpret our results as plausible upper bounds on the productivity dispersion and gains from reallocation in science. Given the dearth of quantitative evidence in this area, we view our estimates and methodology as valuable inputs for future theoretical and empirical studies in the economics of science and innovation.

After a brief review of our connection to the literature, the rest of the paper proceeds as follows: Section 2 provides the general framework and key assumptions underpinning our methodology in general, and our specific model of researchers' production and consumption; Section 3 details the survey data; Section 4 describes and reports the results of the survey experiments; Section 5 provides the model estimates; Section 6 contains our counterfactual allocation exercises; and Section 7 concludes with a discussion of our results and the usefulness of our methodology more generally.

Related Literature

Our paper sits primarily at the intersection of two bodies of literature: empirical studies of science and markets for innovation (i.e., Merton 1973; Stephan 1996; Bryan and Williams

[6]Namely, (*i*) noise due to mismeasurement or misspecification; and (*ii*) the presence of unmodeled heterogeneity in output prices (or preferences over payoffs).

2021); and economic studies of producers' productivity and factor misallocation (i.e., Syverson 2011; Restuccia and Rogerson 2017; De Loecker and Syverson 2021). More generally, we are focused on a classic challenge in labor economics of estimating worker-level productivity metrics (i.e., Hoffman and Stanton 2024).

The meta-science literature has long been interested in misallocation. Following a long line of sociological work (Merton 1973; Zuckerman 1988; Shapin 1995) economists have quantified some status-based frictions (Azoulay, Fons-Rosen, and Graff Zivin 2019) and have also studied potential frictions such as: political lobbying (Hegde and Sampat 2015), information asymmetries regarding researchers' output (Hager, Schwarz, and Waldinger 2024), and competitive pressures from priority-based credit mechanisms (Hill and Stein 2025). However, the scientific production function is rarely formalized in a way that yields the structural parameters necessary for quantifying productivity dispersion or allocative efficiency. Rosenbloom et al. (2015) finds the marginal returns to scientific funding (per bibliometrics) are relatively similar across US universities, which is suggestive of allocative efficiency. Taking a more dynamic view, Qiu (2023) estimates the discount rate that rationalizes funding allocations at the US National Institutes of Health and finds a relatively high discount rate, which suggests young scientists may be under-funded.

Looking towards the productivity literature, our methodology is centered on understanding producers' factor demand, which follows the logic of identification based on input cost shares (e.g., Hsieh and Klenow 2009) or the inversion of input demand into control functions (e.g., Olley and Pakes 1996; Levinsohn and Petrin 2003; Ackerberg, Caves, and Frazer 2015). Much work has been done to extend these methods to account for features such as unobservable prices (De Loecker et al. 2016), adjustment costs (Petrin and Sivadasan 2013; Asker, Collard-Wexler, and De Loecker 2014), as well as measurement error and heterogeneity (Kim, Petrin, and Song 2016; Gollin and Udry 2021). To our knowledge, we are the first to formalize an approach for estimating production functions without data on output quantities and input prices.

We also follow a growing body of work using surveys to study productivity in settings where inputs and outputs are subjective or difficult to measure (e.g., Bloom, Sadun, and Van Reenen 2012; Atkin, Khandelwal, and Osman 2019). We are not the first to solicit producers' WTP for inputs (c.f., Cole et al. 2013; Wossen et al. 2024); however, prior work typically has an inherent interest in producers' demand for a specific input. Our work also runs parallel to the development of methods that use (quasi-)experimental variation to estimate misallocation (e.g., Sraer and Thesmar 2023; Carrillo et al. 2023).

2. Methodological Framework and Model of Science

In this section, we show how a scientist's expectations about their productivity can be estimated by using information about their WTP for a fixed amount of inputs. We discuss the method in general terms in Section 2.1 and then describe our specific model of researchers' labor supply, including functional form assumptions, in Section 2.2. Additional details regarding the model and estimation are contained in Appendix C.

To illustrate our approach to estimating productivity, the following simple example shows how soliciting scientists' WTP for an input can be used to estimate their productivity:

> Utility maximizing scientists are indexed by i and are endowed with expected income, M (e.g., a guaranteed salary). Scientists produce research output, Q, using a single variable input, X (e.g., research funding). Scientists obtain the input by incurring heterogeneous, convex costs given by: $X_i^{c_i}$, where $c_i > 1$. Such convex costs summarize potential frictions identified in the meta-science literature. Scientists produce research output according to a linear research production function: $Q_i = \alpha_i \cdot X_i$, where α_i is the focal productivity parameter. Scientists' utility function is linear in income and research output less costs. They maximize utility by choosing X_i:
>
> $$\max_{X_i \geq 0} \; M_i + b \cdot Q_i - X_i^{c_i} \quad \text{s.t.} \quad Q_i = \alpha_i \cdot X_i,$$
>
> where b is the marginal utility of research output relative to income, the numeraire. We cannot observe the productivity or cost parameters (α_i, c_i), and we cannot observe output (Q_i). Our goal is to estimate productivity (α_i).[7]
>
> Our solution is to solicit scientists' willingness to pay (WTP_i) for a fixed quantity of input using their income (M_i). Consider an experiment where scientists are asked to report their WTP to obtain 1 unit of X without incurring the convex costs. By definition, they will choose WTP_i that makes them indifferent between purchasing and not purchasing the additional unit of X. Formally, WTP_i solves:
>
> $$\max_{X_i} \left[M_i + b \cdot \alpha_i \cdot X_i - X_i^{c_i} \right] = \max_{\widetilde{X}_i} \left[(M_i - WTP_i) + b \cdot \alpha_i \cdot (\widetilde{X}_i + 1) - X_i^{c_i} \right].$$

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

This yields a simple expression for identifying scientists' productivity expectations that does not require output quantities or input costs: $\alpha_i = WTP_i/b$. Clearly, this requires estimating, observing or making assumptions about b, a point we highlight below.

2.1. Framework, Assumptions, and Estimation

2.1.1. General Framework

Our goal is to estimate individual producers' rational expectations about their productivity in a single future period of production. Despite revolving around forecasts, we present the framework as static; there is no dynamic optimization.[8]

There are N individuals indexed by i. For simplicity, we focus on the case where individuals use a single, variable input X to produce some quantity Q of output; extensions to multiple inputs and stocks are possible. Individuals have heterogeneous production functions with a Hicks-neutral total factor productivity (TFP) term: $Q_i = \alpha_i \cdot f_i(X_i)$, which is monotonically increasing in X.[9] Estimating the TFP parameter α_i is our primary goal.

Individuals' benefits are governed by idiosyncratic functions $b_i(M_i, Q_i)$, which depend on output levels (Q_i) and an endowment of money that is guaranteed regardless of output (M_i). For individual scientists, M_i would reflect their guaranteed salary, and the benefit function would describe their utility from their salary as well as any additional benefits that they expect to receive from producing more research output (e.g., prestige, expectations of a promotion).[10] Total costs depend on an input cost function $c(X_i, \mu_i)$ and there may also be constraints on input levels denoted by $g(X_i, \mu_i) = 0$.

Thus, individuals choose input levels that maximize their payoff per:

$$\begin{aligned} \max_{X_i} \quad & b_i\big(M_i, \alpha_i \cdot f_i(X_i)\big) - c(X_i) \\ \text{subject to} \quad & g_i(X_i) = 0\,, \end{aligned} \tag{1}$$

where we define X_i^* as the argument that maximizes Equation (1). Solving the first order conditions of the producer's problem (Eq. 1), ignoring constraints in what follows for

[8]In so doing, we implicitly load dynamic considerations related to career progression onto other parameters in the model. We also omit the expectation operator despite variables being forecasts.

[9]Thus, α_i is a TFP-Quantity (TFPQ) parameter.

[10]For simplicity, we do not allow for financial markets, but they could be incorporated.

simplicity, yields:

$$\alpha_i \cdot \frac{\partial}{\partial X_i} b_i\big(M_i, \alpha_i \cdot f_i(X_i)\big) \cdot \frac{\partial}{\partial X_i} f_i(X_i) - \frac{\partial}{\partial X_i} c_i(X_i) = 0 \,, \tag{2}$$

which shows that productivity (α_i) is an implicit function of observables and the unknown parameters that govern the functions $b_i(...)$,$f_i(...)$, and $c_i(...)$. At this point, productivity is not separately identified from these unknown parameters.

In order to identify the parameters in Eq. 2, we experimentally solicit individuals' WTP for some amount of the input using their own money M_i (e.g., directly from us, the experimenters). For example, we offer the producer an opportunity to purchase Δ units of X and solicit their WTP (e.g., via stated preferences or incentive-compatible revealed preferences), which would be deducted from their M_i.

To see the value of this experiment, first consider the individual's optimization problem in the scenario where they choose to pay their WTP to purchase Δ inputs from us:

$$\max_{\widetilde{X}_i} \quad b_i\big(M_i - WTP_i, \alpha_i \cdot f_i(\widetilde{X}_i + \Delta)\big) - c_i(\widetilde{X}_i) \,, \tag{3}$$

where we define $\widetilde{X}_i^*$ as the argument that maximizes Equation (3). In this scenario, the individual's WTP_i is subtracted from their M_i since they are both in monetary units and treated as perfect substitutes in the benefit function. The inputs purchased, Δ, are added to the production function, but the producer's cost function, $c_i(...)$, still only depends on the quantity of inputs they choose to obtain after the experiment.

Therefore, the producer's WTP for Δ units of the input equates the following:

$$\underbrace{b_i\big(M_i, \alpha_i \cdot f_i(X_i^*)\big) - c_i(X_i^*)}_{\substack{\text{expected utility in} \\ \text{no-purchase scenario}}} = \underbrace{b_i\big(M_i - WTP_i, \alpha_i \cdot f_i(\widetilde{X}_i^* + \Delta)\big) - c_i(\widetilde{X}_i^*)}_{\substack{\text{expected utility in} \\ \text{scenario paying } WTP \text{ for } \Delta}} \,, \tag{4}$$

That is, their WTP make them indifferent between (*i*) the scenario where they don't purchase the inputs from us (left-hand side of Eq. 4) and (*ii*) the scenario where they do purchase from us (right-hand side of Eq. 4). Next, we walk through the key assumptions that facilitate identification of α_i through the inversion of Equation (4).

2.1.2. Key Assumptions

Assumption 1 — Convex Input Costs. The direct costs of inputs are convex: $\partial c_i(\ldots)/\partial X_i > 0$ *and* $\partial^2 c_i(\ldots)/\partial X_i^2 > 0$. These convexities can be due to any sort of friction or adjustment cost. If this assumption does not hold, then the producer's WTP in the experiment may not depend on their productivity, as we illustrate below.[11]

Assumption 2 — Monotonic Benefits. The benefit function $b_i(\ldots)$ *is strictly monotonically increasing in* Q_i *and* M_i. This implies that more money or more output always increases the individuals' payoff and, importantly, that $b_i(\ldots)$ is invertible.

Assumption 3 — Homogeneous or Observable Output Benefits. The benefits due to output, given by $\partial b_i/\partial Q_i$*, are either homogeneous or depend on observable covariates.* Unobservable heterogeneity in the returns to producing more output would not allow us to separately identify individual-specific returns to output from their productivity. Thus, these returns must either be observable (and assumed to capture all quality differences), a function of other observables (in which case they can be estimated), or unobservable and homogeneous (in which case individual productivity beliefs are identified up to a common level shifter). In the language of traditional producers, this assumption implies that, conditional on observables, the output is a commodity and all producers face common output prices.[12]

Assumption 4 — Observable Money and Plans. Individuals' monetary constraints, per M_i*, and optimal input plans* X_i^* *are observable*. There are some knife-edge cases where these variables need not be observed and productivity is identified by the experiment alone, such as the example above. However, those cases are likely rare in practice and current allocations must be observed for estimation.

[11]This is perhaps the least intuitive of our assumptions. Formally, this assumption ensures the necessary rank condition for estimation. Intuitively, the WTP experiment operates via an implicit linear cost schedule, so if the individual already faces a linear cost schedule when obtaining inputs, our WTP experiment may simply reflect those linear costs as a corner solution (i.e., the producer would never pay us, the experimenters, more for an input than what they pay in their own input market).

[12]Recall our simple, illustrative example at the beginning of this section — we showed that $\alpha_i = WTP_i/b$; however, if in fact the benefits of output are given by b_i, which is unobservable, then we cannot separately identify α_i from b_i.

2.1.3. Estimation

Let Δ be the amount of input offered to the scientist in the experiment and WTP_i^{obs} be scientists' reported willingness to pay for those Δ units. Furthermore, let μ_i be the vector of parameters that govern the production, benefit, and cost functions. Under Assumption 2 — the benefit function (b) is invertible — we can write a producer's WTP_i as a function:

$$WTP_i = \Pi(\Delta, M_i, X_i^* ; \mu_i) \equiv \widehat{WTP}(\Delta, M_i, X_i^* ; \mu_i) , \tag{5}$$

which leaves only the unknown parameters μ_i to be estimated. This approach makes use of the fact that productivity (α_i) and optimal allocations in the purchase scenario of the WTP experiment ($\widetilde{X}_i^*$) are implicit functions of observables and unknown parameters.[13] Now, we have a theoretical prediction of a producer's WTP_i for Δ.

In general, the GMM estimator for μ_i solves the minimization problem:

$$\min_{\mu_i} \quad \sum_{i=1}^{N} \Big(\widehat{WTP}(\Delta, M_i, X_i^* ; \mu_i) - WTP_i^{\text{obs}} \Big)^2 , \tag{6}$$

which identifies μ_i given standard GMM identification conditions being met. Particularly relevant for identification is the rank condition, which requires that the matrix of derivatives of the stacked moment conditions with respect to the vector of parameters has full rank.[14] It can be shown that Assumption 1 (Convex Input Costs) is a sufficient condition for the rank condition being met.

Note that instead of soliciting individuals' WTP directly, the experimenter could solicit individuals' current income, M_i, as well as their income that would make them indifferent to receiving δ, and this income-at-indifference could be used for identification. Also, if there are constraints on individuals' input levels, they could be offered to pay to relax those constraints. Furthermore, depending on the variation in the data, it may not be possible to allow for full heterogeneity in the parameters that govern the production, benefit, and cost functions. In this case, the experimenter could assume μ_i is given by a vector of common unobservables, say μ, and idiosyncratic observable variables, say Z_i, and estimate μ via the GMM estimator above.[15]

[13] Specifically, Assumptions 1 and 2 generally guarantee that WTP_i is a direct function of α_i, which, by Equation (C13) in the Appendix, depends on μ.

[14] This condition fails, for example, if the WTP does not directly depend on a specific parameter, or if a parameter is redundant.

[15] After recovering the parameters, we also shrink the dispersion in researchers' productivity within each

2.2. Researchers' Labor Supply and Functional Forms

Researchers supply their labor to 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).[16] Given their contract, researchers choose how much time to spend on research (R_i), fundraising (F_i), which together with duties (D_i) sum up to yield their total work hours (H_i).

The quantity of knowledge Q_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 $H_{\max}$.

Researchers' indirect utility is given by:

$$\mathcal{V}(\mathbf{S}_i, \theta_i, \mu_i) = \max_{R_i, F_i, H_i} b_{1i}(M_i) + b_{2i}(Q_i) - c_i(R_i, F_i, D_i) \tag{7a}$$

subject to

$$B_i = B_{\min} + G_i + \phi_i F_i \tag{7b}$$

$$R_i + F_i + D_i = H_i \tag{7c}$$

$$Q_i = \alpha_i B_i^{\gamma_i} R_i^{1-\gamma_i}, \tag{7d}$$

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 $b_{1i}(\ldots)$, $b_{2i}(\ldots)$, and $c_i(\ldots)$. The functional forms of these benefit and cost functions are assumed to be:

$$b_{1i}(\ldots) = \omega \cdot \frac{M_i^{1-\sigma_i}}{1-\sigma_i} \tag{8a}$$

$$b_{2i}(\ldots) = \frac{Q_i^{1-\eta_i}}{1-\eta_i} \tag{8b}$$

$$c_i(\ldots) = \psi \cdot \frac{(R_i + F_i + D_i^{\xi_i})^{1+\zeta_i}}{1+\zeta_i}. \tag{8c}$$

market before running the counterfactuals.

[16] 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. In the context of the methodological framework presented in Section 2, the function $b_{1i}(\ldots) + b_{2i}(\ldots)$ serves the role of the "benefit" function, while labor disutility $c_i(\ldots)$ represents the cost function, which is parameterized to be convex.

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 $\zeta_i > 0$.

Solving the researcher's maximization problem (Equation (7)) yields optimal choices of research time, fundraising time, and total working hours as a function of states, attributes, and parameters. We represent these optimal choices as policy functions $\mathcal{R}(\mathbf{S}_i, \theta_i, \mu_i)$, $\mathcal{F}(\ldots)$, and $\mathcal{H}(\ldots)$. The derivations of these policy functions are described in Appendix C.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 structural conditions to uniquely identify the production-related parameters $(\alpha_i, \gamma_i, \phi_i)$. Thus, we choose to specify each of the consumption-related parameters as a parametric function of researchers' observable features, whose common parameters we then estimate by GMM consistent with the framework of Section 2.

Fortunately, we have a large vector of observable features $(\mathbf{Z}_i)$ 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 B2). 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.

However, 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 $(\mathbf{Z}_i)$ into a one-dimensional index. We assume there are two clusters of researcher types ($k = 2$) and estimate each researcher's distance from these clusters with 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 C.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 a researcher's type. For instance, the parameter that governs the utility from private consumption is parameterized as: $\sigma_i = \exp(\delta_{\sigma,0} + T_i\delta_{\sigma,1})$. 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 $\mu_i(T_i, \mu)$ as a function of a researcher's type T_i and of the vector of common parameters to be estimated, denoted by $\mu = (\omega, \psi, \delta)$, where δ includes the deep parameters that govern the utility functions.

Allowing the μ_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.

3. Survey and Data Overview

3.1. Survey Design

We use the National Survey of Academic Researchers (Myers et al. 2023) and provide a brief overview of the survey methodology here. The population target is U.S. 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.[17] 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.

[17] 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).

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 follow-up emails. 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 B1 reports the results from a regression of an indicator for survey completion on the different incentive and reminder arms, showing 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 checks 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.[18] Appendix Figure B1 (replicated from Myers et al. (2023)) shows invite-respondent overlap on various measures of scientific inputs and outputs. Overall, there is little difference between the respondents and non-respondents both economically and statistically speaking.

Looking at the institutional level, Appendix Figure B2 (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 are some statistically significant differences; on average, respondents come from institutions with slightly less research funding (4–6%).

[18] 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.

3.2. Summary Statistics: Researchers and their Inputs

Table 1 reports the summary statistics of the key covariates in our analyses. See Myers et al. (2023) for a more detailed investigation of these summary statistics.

Markets. We define scientific markets according to two dimensions: field and career stage. Using the name of the professor's department and their self-reported scientific discipline, we assign them to a set of twenty fields of study.[19] We define career stage as a binary distinction based on whether a researcher is younger than the within-field median age ("early career"), or equal to or older than the within-field median age ("late career"). Thus, we obtain forty scientific markets. Importantly for our later analyses, all counterfactuals we estimate constrain the reallocation of inputs to only occur *within* these field×career-stage markets.

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.

[19] We also aggregate those fields into five broader "aggregate" fields for simplicity in reporting some results: Humanities and related; Engineering, Math, and related; Medicine and Health Sciences; Natural Sciences; and Social Sciences.

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 B3).

Position and Socio-demographics. Appendix Table B2 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).

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 focused on answering existing questions.

3.3. 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 focuses 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 B3). 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.

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 five-year period in exchange for a higher salary.[20] All of these hypotheticals are posed over a five-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 being filtered through the model. Appendix Figure C2 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 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 researchers' scientific productivity should plausibly be uncorrelated with their demand for high-speed internet access at home. In Appendix Table C1, we show that respondents' stated WTP for scientific inputs are rarely correlated with their stated WTP for the benchmark good, 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

[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.

the researcher's utility in the counterfactual scenario equal to their indirect utility $\mathcal{V}_i^*$ at current allocations. Formally, counterfactual guaranteed funding and duties in the four experiments are:

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

where G_i and D_i represent actual states.

As noted in Section 2, the values of $\widehat{M}_{ij}$ reported in these four experiments make the researcher indifferent between all possibilities, and therefore we can write these reported values to be a known function of observable data and the unknown parameters to be estimated.

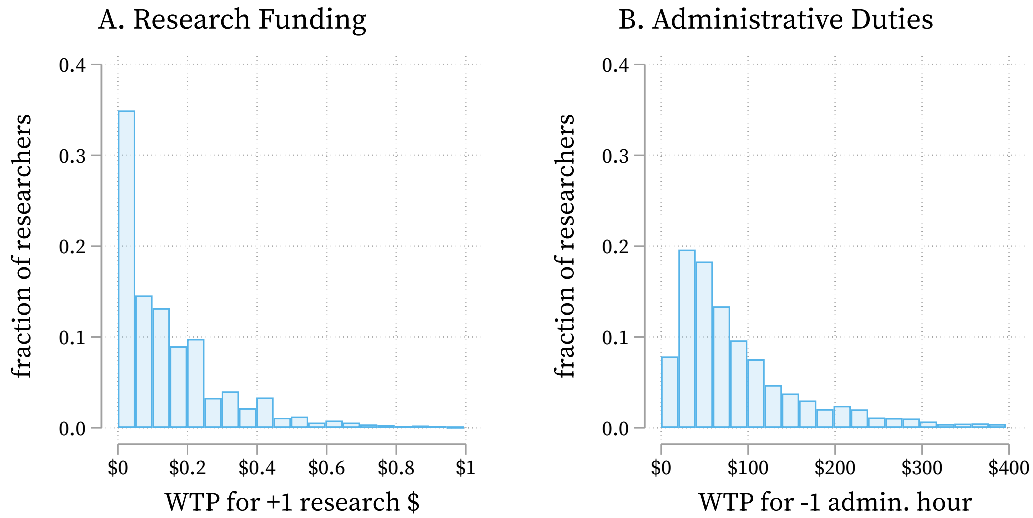
4.3. Survey Experiment Results

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 another portion 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, behavioral biases, or noise. Although these are practical trade-offs that professors face throughout their careers, systematic data have 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 C3A, we see that this is indeed the case. Furthermore, it should be

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.

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 C3B, we find this to be true.

As another test, Appendix Tables C1 and C2 describe analyses where we regress researchers' WTP responses on the variables in the model as well as (*i*) the large vector of observable features ($\mathbf{Z}_i$), (*ii*) an inverse Mills ratio constructed using the randomized survey participation incentives following Heckman (1979), and/or (*iii*) researchers' WTP for the benchmark good (high-speed internet) following Dizon-Ross and Jayachandran (2022). We find the variables in the model can explain 82–96% of the variation in WTP responses (see Appendix Table C1). This supports the model's ingredients as being key determinants of researchers' answers. When we include the large vector of observable features in addition as additional explanatory features, the R^2 statistics increase by only one to four percentage points. Residual variation in WTP responses (conditional on the variables of the model) are not well explained by these heterogeneous observables. This gives us confidence that we are not dramatically mis-specifying heterogeneity in the model.

We also find that the inverse Mills ratio is not a statistically significant predictor of responses (see Appendix Table C2). 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 (see Appendix Table C2). 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. However, the economic magnitude of this relationship is quite small. Furthermore, 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 allows us to control for this to some degree. Overall, despite not being incentivized experiments, researchers' responses behave as expected and are of plausible magnitudes.

5. Estimates for the Model of Science

5.1. A View of Researchers' Utility Functions

To illustrate our estimates of the utility function, Appendix Figure D1 shows how an average researcher's utility depends on the levels of the three state variables (salary, administrative duties, guaranteed research funding) and research output. The figure shows the percent change in a researcher's utility 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.

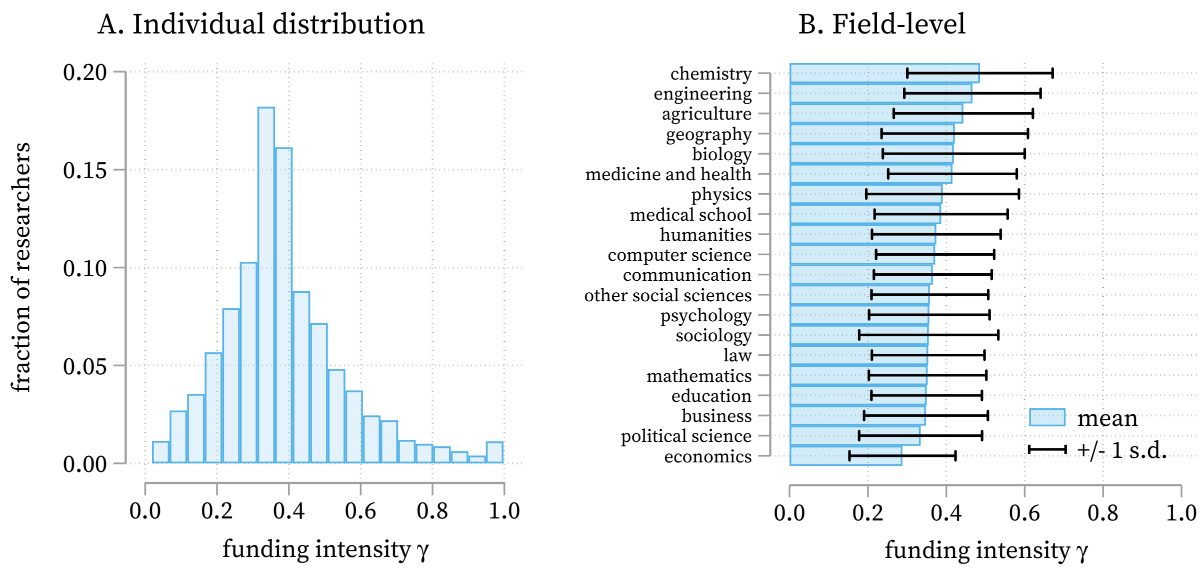
In terms of magnitude, salary and administrative duties are the most important for researchers' utility. Shifting the average researcher's salary from the 10^{th} percentile to the 90^{th} percentile increases their utility by roughly 60%. An equivalent relative increase in administrative duties reduces utility by about 20%. In contrast, similarly scaled increases in guaranteed funding or research output raise utility by only a few percentage points.[22]

5.2. Funding intensity

Figure 2 shows the unconditional distributions of the funding intensity parameter, which describes researchers' marginal returns to funding (per γ_i) relative to their own research

[22] 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. The numerical estimates of the common parameters and more detailed descriptions of the model's fit are available from the authors upon request.

FIGURE 2. Funding intensity



Note: Panel A shows the unconditional distribution of researcher-specific funding intensity (γ) estimates. Panel B shows the mean plus/minus one standard deviation of γ by field.

time (per 1-γ_i). Figure 2A shows the individual-level distribution and 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, Figure 2B plots the average funding intensity parameters (γ_i) for the twenty fields represented in the sample. We find that γ_i is highest in chemistry and engineering, where research is often capital intensive and involves the use of expensive lab equipment. In contrast, the social sciences (e.g., economics, political science) display the lowest funding intensity on average. Notably, the relatively large standard deviations reveal significant heterogeneity in funding intensity across researchers within each field of study.

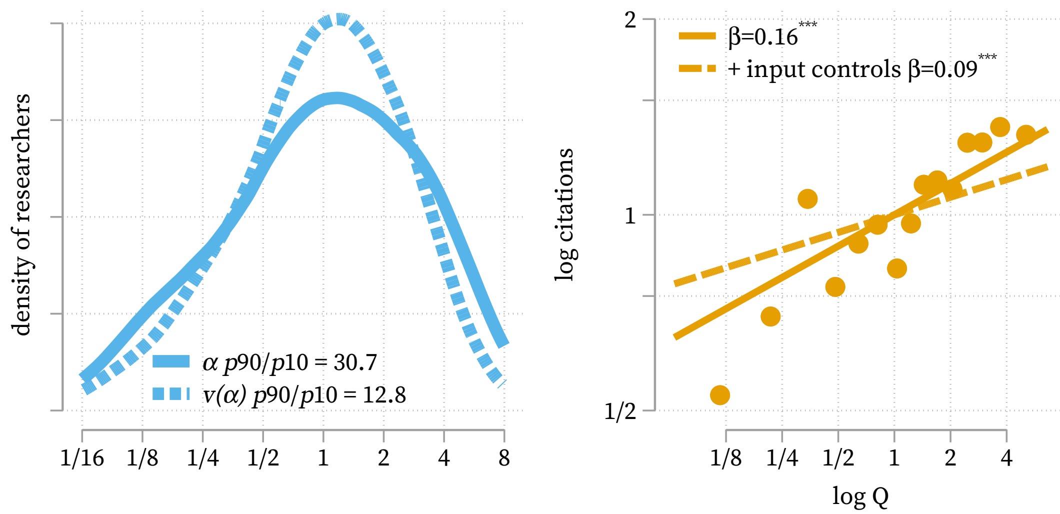
5.3. Total factor productivity and output

Figure 3A shows the distribution of the total factor productivity parameter (α), normalized to be median 1 within each market. Our estimates reveal a large skew in researchers' beliefs about their productivity. In addition to the parameter value, Figure 3A plots the distribution of mean value-equivalent TFP by filtering each researchers α_i parameter through the mean $\nu(\ldots)$ in their market.

Figure 3A also reports the ratio of the 90^{th} and 10^{th} percentiles, the "90–10 TFP ratio"

FIGURE 3. Productivity & output

A. Total factor productivity B. Output quantity (Q) & citations



Note: Panel A shows the distribution of TFP, α_i (solid line), and the distribution of mean-value-equivalent TFP, per $\nu(\alpha_i)$ (dashed line). Panel B shows binned scatterplots of citations (five-years pre-survey) against model-implied output (Q_i). The unconditional relationship (solid) and the relationship conditional on input levels B_i and R_i (dashed) are shown separately. In all cases, variables are normalized to be median 1 within market and shown on log scales.

measure of variance commonly reported in traditional productivity studies. The 90–10 TFP ratio is substantial: the 90^{th} percentile researcher believes they are roughly 30 times as productive as the 10^{th} percentile researcher. In mean value terms, per $\nu(\alpha)$, this translates to more than a 10-fold difference.

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 1.5× to 5× (Syverson 2011). To date, most 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.[23]

Another test of the validity of our model is whether our model-based output measure tracks observable proxies of research performance. To conduct this test, we use biblio-

[23] For some evidence of face validity, we note that the 90-10 ratio for field-normalized citations per year in our sample is 44, which is on the same scale of productivity dispersion we estimate.

metric data matched to approximately three-quarters of the sample and construct each researcher's cumulative citations over the five years preceding the survey. We lack post-survey bibliometric data; however, prior work documents that citation-based output is highly persistent at horizons of this length (e.g., Sinatra et al. 2016), so the timing mismatch between survey responses and bibliometric data is unlikely to attenuate the relationship substantially.

Figure 3B confirms that our model-based output measure and citation output are positively correlated within-markets. This is a reassuring test, but does not speak to the validity of our production function parameters per se: our model-based output depends on both productivity parameters and input levels, where the latter have long been known to correlate with citations. A sharper test conditions on input levels. If our productivity estimates are capturing genuine heterogeneity in α_i and γ_i, then the quantity-citation relationship should survive after absorbing inputs. Figure 3B shows that it does: the conditional correlation is attenuated, as expected, but remains statistically significant.[24]

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: $\mathrm{Var}(\log(\alpha_i)) + 2\mathrm{Cov}(\log(\alpha_i), \gamma_i \log(B_i)) + 2\mathrm{Cov}(\log(\alpha_i), (1-\gamma_i)\log(R_i))$. Using this relationship, we estimate that 54% 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 107%. 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 Section 6 below.

6. Social Welfare and Efficiency

The preceding sections show that researchers have significantly different production functions and input levels. We now seek to answer the allocative efficiency question: given our estimates, how much value could be gained if funding and research time were

[24]Note, we view this output-citation test as validating the model's output measure *on average*: researchers whom the model ranks as highly productive do, in fact, produce more cited work. But this test does not imply that citations are a sufficient statistic for research value at the individual level. The bibliometric correlation provides a check on the average quality of our output measure, not an endorsement of citations as the object of interest.

optimally allocated across researchers?

We proceed in four steps. First, we define social welfare and a first-best benchmark in which a planner can assign funding and research time directly (Section 6.1). Second, we report the changes induced by a shift from observed to first-best allocations, and convert the gains from this reallocation into budgetary equivalents (Section 6.2). Third, we construct a diagnostic counterfactual that allows us to quantify how much cross-sectional variation in funding inputs contributes to current misallocation (Section 6.3). Finally, we examine which observable characteristics predict who is over- and under-resourced (Section 6.4).

6.1. Welfare and the first-best allocation

Welfare definition. We define the social welfare generated by each individual researcher i in market j as:

$$\mathcal{W}_i = U_i + \kappa \cdot \nu_j(Q_i), \tag{9}$$

where $\kappa \geq 0$ determines the scale of the average knowledge externality, and ν_j is a market-level function that converts output Q into social value and determines how much each individual over- or under-appropriates the value of their output conditional on κ.[25] 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 20% ($\kappa = 4$) and as low as 2% ($\kappa = 49$) (Nordhaus 2004; Jena and Philipson 2008; Lakdawalla et al. 2010). In our empirical analyses, our conservative assumption is $\kappa = 4$, which implies that the average researcher in each market captures 20% of the social value their output creates.

First-best allocation. We define the first-best as the solution to a problem where the planner can directly assign funding B_i and research time R_i to each researcher within a market, holding the aggregate levels of B and R in each market at observed totals:

$$\max_{\{B_i, R_i\}} \sum_i \mathcal{W}_i \quad \text{s.t.} \quad \sum_i B_i = \sum_i B_i^{\text{obs}}, \;\; \sum_i R_i = \sum_i R_i^{\text{obs}}. \tag{10}$$

[25]We assume that $\nu_j(Q_i) = \frac{Q_i^{1-\bar{\eta}_j}}{1-\bar{\eta}_j}$, where $\bar{\eta}_{j(i)}$ is the market-level mean of η_i for $i \in j$.

Three features of this benchmark deserve emphasis. First, because the first-best planner assigns funding and research time directly, the problem is silent about the implementation of the optimal allocation in a decentralized equilibrium where the policy maker faces limited information or constraints on the available policy levers. In practice, the planner could act on guaranteed funds G_i or individual fundraising productivity ϕ_i, subject to constraints that reflect the fact that fundraising and guarantees may have value outside the model.[26] Second, we impose that all reallocation occurs within-market (i.e., field×career-stage cells); we only compare researchers to peers in their own discipline and cohort in order to highlight the role of heterogeneity in individual productivity for misallocation. Third, the aggregate constraints ensure that gains come purely from reallocation, not from expanding the total pool of funding or research time.

Researchers' duties (e.g., administration, teaching) are held at observed levels. We treat these as producing social value outside the model. Furthermore, we impose researcher-specific budget bounds $B_{\min,i} \leq B_i \leq B_{\max,i}$ and hour constraints $R_i + D_i \leq H_{\max,i}$ to ensure that gains reflect feasible reallocations within the range of observed variation of the data and the survey experiments.

Relative to the observed allocation, the first-best addresses two categories of distortions in scientific markets:

- *Output-value distortions*, which arise because (*i*) researchers under-appropriate the value of their output to the extent that $\kappa > 0$, and (*ii*) conditional on κ, researchers' marginal utility from additional output is larger (if $\eta_i < \bar{\eta}_j$) or smaller (if $\eta_i > \bar{\eta}_j$) than the social value.
- *Input distortions*, which arise due to (*i*) the practical fact that scientific markets rely on a combination of guarantees and fundraising mechanisms for allocating funding instead of direct allocation (e.g., due to information asymmetries), and (*ii*) conditional on the use of these mechanisms, the (shadow) prices that researchers face when choosing their input levels may not reflect the socially optimal (shadow) prices necessary for an efficient allocation (e.g., due to biases in the peer review process or heterogeneous access to opportunities).

Separately identifying inefficient distortions from technology constraints here (e.g., as discussed in Bergquist, Lashkari, and Verhoogen (2026)) is challenging as it requires

[26] A minimum positive amount of guarantees may have independent value due to the uncertainty involved in research and because effort is not contractible. Similarly, a non-zero productivity of fundraising may be necessary insofar competitive fundraising is how non-market allocation works absent price signals.

taking a stance on what policies are feasible for a second-best planner facing limited information about individual-specific technologies and constraints related to the reality that funding guarantees and fundraising opportunities may have value outside the model. Therefore, in Section 6.2 we start by primarily focusing on comparing the observed versus the first-best allocations. Then, in Section 6.3 we examine a counterfactual scenario where we eliminate any heterogeneity in guaranteed funding and fundraising ability across researchers in the same market, a policy that would be certainly feasible to the policy maker. We interpret this exercise as a diagnostic tool to measure how much a (much-debated feature of scientific markets) — heterogeneous funding opportunities and costs — contributes to allocative inefficiencies.

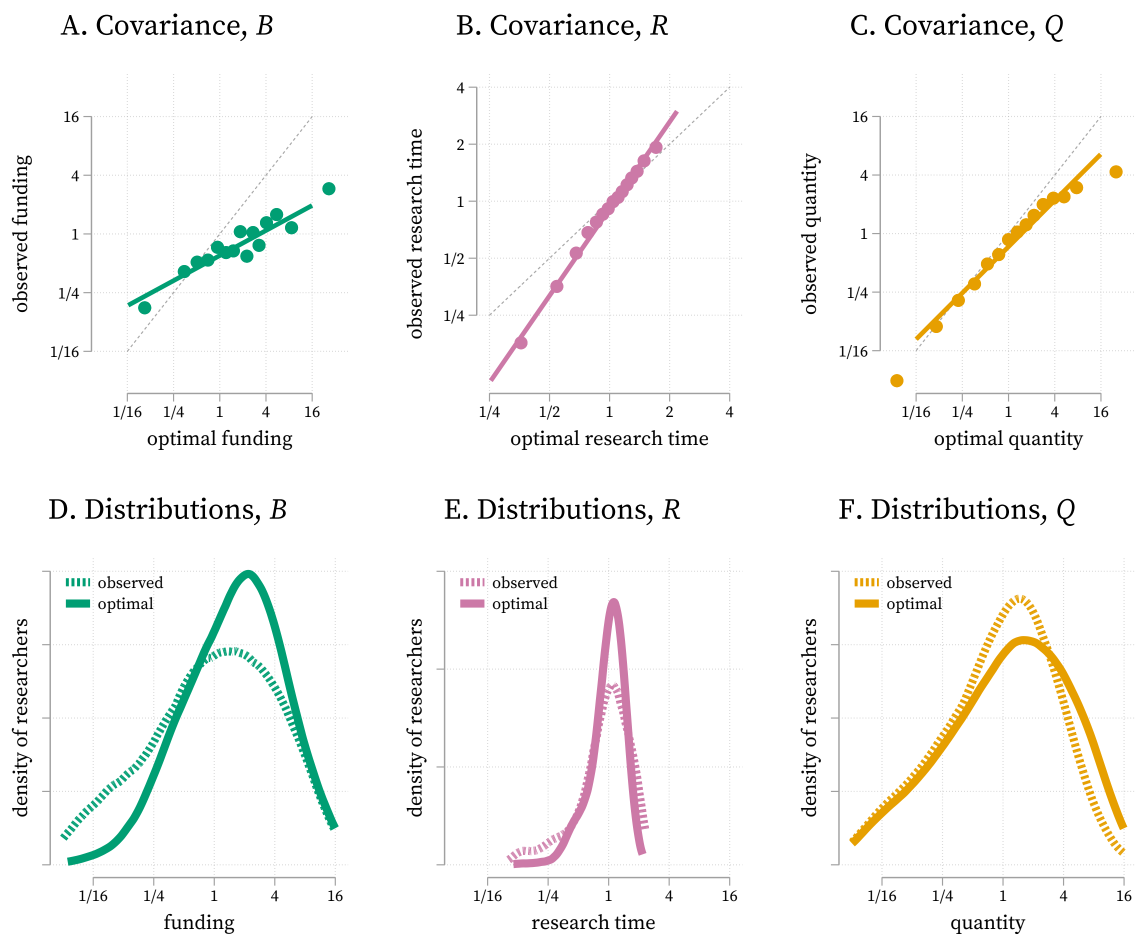
6.2. Reallocation to first-best

Because the planner in Equation (10) holds fixed the aggregate quantities of funding and research time within each market, the exercises in this subsection isolate gains from reallocation alone. There is no expansion of the resource envelope. At the same time, because the planner assigns B_i and R_i directly, the resulting magnitudes should be interpreted as a first-best upper bound relative to policies that must operate through decentralized funding mechanisms or other second-best instruments.

Inputs and outputs. Figure 4 compares observed and first-best input and output levels. Panels A–C show that observed and optimal allocations are positively correlated. This is important: the actual allocation is not random (or worse). Researchers who are more productive per our estimates do tend to receive more funding, devote more time to research, and in turn, produce more knowledge. In other words, scientific markets appear to do some positive selection on productivity *on average*. Panels D–F of Figure 4 show how the distributions vary between observed and optimal allocations. Because aggregate funding and research time are fixed within each market, any difference between the observed and optimal input distributions reflects a re-sorting of existing resources. Under the first-best, resources are shifted toward researchers with higher social marginal products, which results in more compressed input distributions, but a slightly more dispersed output distribution. This illustrates that some of the variance in input levels does not reflect productivity dispersion but rather distortions. Next, we estimate the aggregate costs of these distortions.

Aggregate changes in output, utility, and welfare. Table 3 reports the aggregate implications of these reallocations. Relative to the observed allocation, the first-best increases average

FIGURE 4. Observed and optimal input and output levels



Note: Panels A–C show binned scatterplots of first-best optimal and observed levels along with linear fits; 45^o lines of equality are shown. Panels D–F show the distribution of researcher-specific levels in optimal and observed allocations. All variables are normalized to be median 1 within market and shown on log scales.

output by roughly 63% (in value terms per $v(Q)$), private utility by 0.5%, and social welfare by 2.9%. Despite the positive correlations documented in Panels A–C of Figure 4, the noisy allocation of inputs leaves substantial aggregate value on the table.

To place the gains from reallocation in more familiar units, we translate them into a budget-equivalent increase under the observed allocation structure. Specifically, we ask how much guaranteed funding would need to rise, holding fixed the current allocation rules and allowing researchers to re-optimize their time allocations, in order to generate the same aggregate output as the first-best reallocation. We estimate that matching the first-best, in terms of its social welfare gains, would require researchers to receive an additional 218% of funding, which corresponds to approximately $400,000 per year. When scaled to the relevant population (using the inverse of our response rate assuming no sample selection), this is equivalent to nearly $50 billion in additional research funding. This budget-equivalent calculation is not itself a welfare object, but it provides a transparent way to benchmark the magnitude of the allocative inefficiencies implied by the model.

TABLE 3. Changes from alternative allocations

	(1)	(2)
funding (B_i) % mean	~0%	~0%
% s.d.	–22%	–32%
research time (R_i) % mean	~0%	+1.36%
% s.d.	–31%	–7.7%
fundraising time (F_i) % mean	–100.00%	+8.36%
% s.d.	–100.00%	+27.3%
Utility (U_i) % mean	+0.52%	+0.42%
% s.d.	+0.02%	+0.09%
Output ($v(Q_i)$) % mean	+63.1%	+43.2%
% s.d.	+119%	+52.5%
Welfare ($\mathcal{W}_i$) % mean	+2.87%	+2.03%
% s.d.	–0.21%	–0.35%
growth in budget through obs. allocations needed to match $\mathcal{W}$ gain from reallocation		
researcher-level (B_i) %	+218%	+132%
$/ yr	+$407,112	+$257,172
population-level ($\sum_i B_i \times \text{response rate}^{-1}$) $/ yr	+$49.4B	+$31.2B

Note: Reports changes in input, outputs, and value under the first-best optimal allocation and the uniform funding access policy.

Taken together, Figure 4 and Table 3 suggest a market in which actual allocations are directionally informative about productivity, but still far from efficient.

6.3. Decomposing the sources of misallocation

To better understand *where* misallocation may come from, we construct a second counterfactual based on a policy of uniform funding opportunities: within each market, every researcher receives same guaranteed funding $\bar{G}$ and the same fundraising ability $\bar{\varphi}$. The common guaranteed funding $\bar{G}$ is set so that total guaranteed funding matches its observed aggregate: $N_f \bar{G} = \sum_{i \in I_f} G_i^{\text{obs}}$. The common fundraising ability $\bar{\phi}$ is set so that, after researchers re-optimize their time allocations, aggregate fundraised dollars match the observed total: $\sum_{i \in I_f} \bar{\phi} F_i^{\text{cf}} = \sum_{i \in I_f} E_i^{\text{obs}}$. Researchers take $(\bar{G}, \bar{\phi})$ as given and maximize private utility U_i.

This uniform policy is not to be taken as an optimal allocation. Rather, we use it as a diagnostic benchmark that removes a specific source of heterogeneity (in G and ϕ) on which there is active debate in the scientific community. While $\bar{G}$ and $\bar{\phi}$ are almost surely socially suboptimal in a second-best world where the planner has to use these specific mechanisms – i.e., reallocate guaranteed funding and additional fundraising opportunities – this is still a useful scenario because it reflects a simple and much debated allocation that would certainly be feasible to the policy maker. The key property of this counterfactual is that it retains the fundraising mechanism and the externality structure intact. It removes *only* cross-sectional heterogeneity in the terms of the funding system. It asks: what would happen if every researcher in a cell faced the same funding terms, but everything else—the externality, the fundraising mechanism, each researcher's productivity, preferences, and duties—stayed the same?

Recall that researcher *i*'s observed funding budget is given by $B_i^{\text{obs}} = G_i + \phi_i F_i + B_{\min}$. Then, we can define deviations ε_{Gi} and $\varepsilon_{\phi i}$ such that:

$$B_i^{\text{obs}} = \bar{G} \cdot (1 + \varepsilon_{Gi}) + \bar{\phi} \cdot (1 + \varepsilon_{\phi,i}) \cdot F_i + B_{\min}, \tag{11}$$

where ε_{Gi} measures how much researcher *i*'s guaranteed funding deviates from the market average and $\varepsilon_{\phi i}$ measures how much *i*'s fundraising ability deviates from the market average. When $\varepsilon_{Gi} = \varepsilon_{\phi i} = 0$ for all i, we obtain the uniform-policy counterfactual. Note, the ε terms are descriptive objects: they measure heterogeneity around the market average, *not necessarily* distortions relative to an optimum.

Welfare-relevant marginal products and ratios. The welfare-relevant marginal product of funding for researcher i under allocation a for factor $X = \{B, R\}$ is defined by

$$\mathrm{MWPX}_i^a = \left.\frac{\partial \mathcal{W}_i}{\partial X_i}\right|_{X_i = X_i^a}.$$

In an efficient allocation, welfare-relevant marginal products are equated across researchers within a cell; dispersion in MWP_B indicates misallocation in funding. An analogous object measures misallocation in research time, replacing the gross marginal product with the net marginal product (gross minus marginal effort cost).

We define three ratios, each comparing welfare-relevant marginal products at two allocations. For researcher i and inputs $X = \{B, R\}$:

$$\omega_i^{X^o/\hat{X}} = \frac{MWPX_i^{\text{observed}}}{MWPX_i^{\text{optimal}}} \quad \text{(all distortions)},$$

$$\omega_i^{X^o/\hat{X}} = \frac{MWPX_i^{\text{observed}}}{MWPX_i^{\text{uniform}}} \quad \text{(heterogeneous input costs)},$$

$$\omega_i^{X^o/\overline{X}} = \frac{MWPX_i^{\text{uniform}}}{MWPX_i^{\text{optimal}}} \quad \text{(distortions cond'l on het. input costs)}.$$

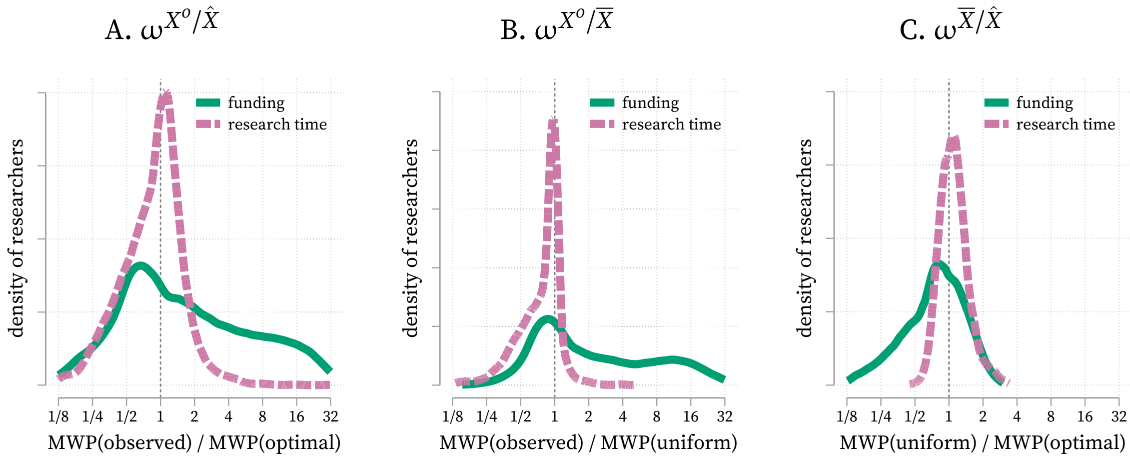
The first ratio captures the total gap between observed and first-best. The second isolates the contribution of heterogeneous funding terms: since the uniform-policy benchmark retains the fundraising mechanism and the externality structure, the gap between observed and uniform-policy WMPs is driven by cross-researcher differences in (G_i, ϕ_i). The third captures everything that remains after equalizing funding terms—the externality, the fundraising mechanism, and any residual heterogeneity in marginal products that persists (or that is introduced) under common $(\bar{G}, \bar{\phi})$.

The three ratios satisfy a multiplicative decomposition so that in logs, for each $X = \{B, R\}$:

$$\operatorname{var}\left(\log \omega_i^{X^o/\hat{X}}\right) = \operatorname{var}\left(\log \omega_i^{X^o/\overline{X}}\right) + \operatorname{var}\left(\log \omega_i^{\overline{X}/\hat{X}}\right) + 2 \cdot \operatorname{cov}\left(\log \omega_i^{X^o/\overline{X}}, \log \omega_i^{\overline{X}/\hat{X}}\right).$$

Figure 5 summarizes the results. Panel A shows that the observed-versus-first-best wedges are widely dispersed for both funding and research time, consistent with the large gains from reallocation documented above. Panel B then asks how much of that dispersion is removed when funding opportunities are equalized. Panel C shows the remaining gap between the uniform-policy benchmark and the first-best.

FIGURE 5. Marginal welfare product wedges



D. Variance decomposition

		$\text{var}(\log \omega^{X^o/\hat{X}})$	=	$\text{var}(\log \omega^{X^o/\overline{X}})$	+	$\text{var}(\log \omega^{\overline{X}/\hat{X}})$	+	$2\text{cov}(\ldots)$
$X =$	B	2.40		64%		20%		16%
	R	0.43		51%		24%		25%

Note: Panels A–C show the distribution of researcher-factor-specific wedges that compare marginal welfare products in the observed-v-optimal allocations (Panel A), the observed-v-uniform-policy allocations (Panel B), and the uniform-policy-v-optimal allocations (Panel C). Panel D shows how much of the variance in the observed-v-optimal wedges can be explained by the other two wedges and their covariance.

Whether the uniform policy compresses or expands wedge dispersion is, in principle, an empirical question. If the observed heterogeneity in funding terms is positively aligned with social marginal products, then equalizing those terms removes some useful variation. If it is negatively aligned, equalizing them reduces misallocation. The results of our counterfactuals show that $\text{var}(\log \omega_i^{\overline{X}/\hat{X}}) = 0.47$ for funding and 0.10 for research time and $\text{var}(\log \omega_i^{X^o/\overline{X}}) = 1.54$ for funding and 0.22 for research time . This implies that heterogeneous funding opportunities amplify total misallocation on net.

Panel D quantifies the decomposition. Setting aside covariances, for funding, the component associated with heterogeneous funding opportunities accounts for 64% of the total dispersion in wedges; for research time, the corresponding share is 51%. The remaining dispersion reflects the residual gap between the uniform-policy benchmark and the first-best, which includes the role of output-value distortions and any remaining input distortions not removed by equalizing (G_i, ϕ_i).

This interpretation is consistent with the aggregate results of the uniform policy counterfactual reported in Table 3. Compared to observed allocations, the uniform-policy benchmark yields 43% more output, 0.42% more private utility, and 2% more social welfare relative to the observed allocation. Relative to the first-best, this means the uniform policy yields roughly 70% of the total welfare gain from reallocation. Thus, heterogeneous funding opportunities can explain an economically meaningful, but not complete, share of the misallocation in scientific markets.

6.4. Who is over- and under-resourced?

The preceding analysis establishes that misallocation is large in the aggregate. We now estimate the extent to which the researchers who receive too much or too little of each input can be identified from standard observable characteristics.

Predictability of input wedges. To study this question, we estimate Poisson models of the following form

$$\mathbb{E}\left[\text{input ratio}_i\right] = \exp\left(\mu_{j(i)} + \mathbf{Z}_i \beta_Z\right), \tag{12}$$

where the dependent variable is the ratio of input levels under the following allocations: (*i*) observed / optimal per first-best; (*ii*) observed / uniform policy; or (*iii*) uniform / optimal. Ratios above one indicate that the researcher receives more of the input than the denominator allocation would assign, and vice versa. The term $\mu_{j(i)}$ denotes market fixed effects and $\mathbf{Z}_i$ is the large vector of observable characteristics that we use to generate our researcher type index (e.g., rank, tenure status, years since PhD, institution type, socio-demographics, etc.). Because the covariate set is high-dimensional, Table 4 reports results from an elastic net Poisson procedure with market fixed effects throughout. Furthermore, we refrain from reporting specific coefficients since this regression is not a test of causation, but rather predictive power. Thus, we report the pseudo-R^2 metrics from these regressions as our focal measure of predictability.

Table 4 reports that observable characteristics have limited predictive power for these input wedges. The resulting fit indicates that much of the variation in who is over- and under-resourced is not well captured by standard observable features of researchers' positions, backgrounds, or prior publication records. The pseudo-R^2 values indicate that, regardless of which allocation or input being compared, typically only 5–10% of the input wedge variation is predictably correlated with researchers' observable features.

This result should be interpreted narrowly. It does not imply that these observables are unrelated to input levels in general. Rather, they do not appear to be powerful predictors of which researchers receive more or fewer inputs than they should, as defined by the wedge of interest. It is also relevant for interpretation of the wedges themselves: if the wedges were primarily proxying for heterogeneous preferences that our model could not accommodate, then these observables would be expected to explain a larger share of their variation. That does not appear to be the case.

More broadly, these results suggest that much of the relevant heterogeneity for allocative efficiency in science is difficult to recover from conventional data that a policymaker might have access to.

TABLE 4. Observable features and input wedges

	funding			research time		
	(1)	(2)	(3)	(4)	(5)	(6)
pseudo-R^2	0.06	0.08	0.25	0.05	0.05	0.09
ratio	$\frac{obs}{opt}$	$\frac{obs}{unif}$	$\frac{unif}{opt}$	$\frac{obs}{opt}$	$\frac{obs}{unif}$	$\frac{unif}{opt}$

Note: Reports estimates from elastic net Poisson regressions of Equation 12. The dependent variable in each column is the researchers' input level ratio under the two allocations indicated by the "ratio" row. Only the pseudo-R^2 is reported.

Case study: Gender disparities. 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 (e.g., Witteman et al. 2019; Kim and Moser 2025. 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.

Table 5 reports the results of simple Poisson regressions with market fixed effects, which allows the gender coefficients to be read directly across allocations. Compared to the first-best optimal allocation, female researchers obtain 16% less funding (approximately $20,000 per year) and devote 5% less time to research (approximately 1 hour per week) than male researchers in observed allocations. Roughly the same magnitudes of under-resourcing are apparent when comparing the uniform policy to observed allocations.

However, both of these gaps become statistically insignificant and economically smaller when comparing the uniform policy allocation to the first-best benchmark. Thus, a substantial share of the observed gender gap in inputs appears to be associated with heterogeneity in guaranteed funding and fundraising ability. Moving from the uniform-policy benchmark to the first-best changes the gender gap by comparatively little, implying that the residual gap after equalizing funding terms is modest in the context of the full first-best reallocation.

This exercise is descriptive rather than causal. It indicates where, within the structure of the model, the gender gap in inputs appears to arise. It does not identify *why* researchers of different gender face different funding terms, nor does it imply that all observed differences are inefficient.

TABLE 5. Gender disparity in input wedges

	funding			research time		
	(1)	(2)	(3)	(4)	(5)	(6)
$\mathbb{1}\{\text{female}\}$	-0.177^{***}	-0.181^{***}	-0.041	-0.055^{***}	-0.060^{***}	-0.004
	(0.046)	(0.046)	(0.025)	(0.016)	(0.016)	(0.015)
% diff	−16.2%	−16.6%	−4.0%	−5.4%	−5.8%	−0.4%
level diff	−$19,294 / yr	−$19,828 / yr	−$4,757 / yr	−0.97 hrs/wk	−1.07 hrs/wk	−0.07 hrs/wk
ratio	$\frac{obs}{opt}$	$\frac{obs}{unif}$	$\frac{unif}{opt}$	$\frac{obs}{opt}$	$\frac{obs}{unif}$	$\frac{unif}{opt}$
$\mathbb{1}\{f\}$ partial R^2	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01

Note: Reports estimates from Poisson regressions of Equation 12 using only the female indicator as a covariate in addition to the market fixed effects. The dependent variable in each column is the researchers' input level ratio under the two allocations indicated by the "ratio" row. The sample includes only researchers who self-identify as female or male.

7. Discussion

By some measures, the market for science believes that the benefits from a more efficient allocation are low and declining (or, the costs are high and increasing). Consider the share of science committed to *meta-science* — efforts to improve the efficiency of science writ large: in recent years, approximately 0.1-1.0% of all publications relate to meta-science; furthermore, while US investments in science grew roughly 3-fold over the past fifty years, the number of meta-science studies per dollar declined by nearly 80% in that same period.[27] This could reflect a rational, fully-informed allocation of resources. Or, it

[27] See Appendix E for the methods used to produce these estimates.

may reflect significant uncertainty about the degree to which the aggregate productivity of the scientific enterprise could be improved. In this paper, we use a new approach to productivity estimation in order to shed new light on the benefits from improving allocative efficiency in science.

Our approach allows us to estimate researchers' productivity beliefs without observing output quantities or input prices. We find actual input allocations to be positively correlated with those that maximize a plausible social welfare function, but we also find very large gains to be had from more efficient allocations. In the counterfactuals we explore, total scientific output 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 relies on researchers' 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. Still, 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.

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 many potential biases affecting these beliefs. Still, science is inherently about forecasting uncertain outcomes; 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 (e.g., Becker, DeGroot, and Marschak 1964). However, the magnitude of costs in our setting is on the scale of tens to 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. 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 nonprofit 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 understand the distribution and determinants of productivity and efficiency.

References

Ackerberg, Daniel A, Kevin Caves, and Garth Frazer. 2015. "Identification properties of recent production function estimators." *Econometrica* 83 (6): 2411–2451.

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.

Asker, John, Allan Collard-Wexler, and Jan De Loecker. 2014. "Dynamic inputs and resource (mis) allocation." *Journal of Political Economy* 122 (5): 1013–1063.

Atkin, David, Amit K. Khandelwal, and Adam Osman. 2019. "Measuring Productivity: Lessons from Tailored Surveys and Productivity Benchmarking." *AEA Papers and Proceedings* 109: 444–49.

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.

Azoulay, Pierre, Joshua S Graff Zivin, Danielle Li, and Bhaven Sampat. 2019. "Public R&D investments and private-sector patenting: Evidence from NIH funding rules." *Review of Economic Studies* 86 (1): 117–152.

Baruffaldi, Stefano, and Fabian Gaessler. 2025. "The returns to physical capital in knowledge production: Evidence from lab disasters." *American Economic Journal: Applied Economics* Forthcoming.

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

Bergquist, Lauren F, Danial Lashkari, and Eric Verhoogen. 2026. "Wedges: A Microeconomic Perspective on Misallocation." *Mimeo*.

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.

Bloom, Nicholas, Raffaella Sadun, and John Van Reenen. 2012. "The organization of firms across countries." *The Quarterly Journal of Economics* 127 (4): 1663–1705.

Bryan, Kevin A, and Heidi L Williams. 2021. "Innovation: Market failures and public policies." In *Handbook of Industrial Organization*, vol. 5, 281–388: Elsevier.

Carrillo, Paul, Dave Donaldson, Dina Pomeranz, and Monica Singhal. 2023. "Misallocation in firm production: A nonparametric analysis using procurement lotteries." *Mimeo*.

Cole, Shawn, Xavier Giné, Jeremy Tobacman, Petia Topalova, Robert Townsend, and James Vickery. 2013. "Barriers to household risk management: Evidence from India." *American Economic Journal: Applied Economics* 5 (1): 104–135.

De Loecker, Jan, Pinelopi K Goldberg, Amit K Khandelwal, and Nina Pavcnik. 2016. "Prices, markups, and trade reform." *Econometrica* 84 (2): 445–510.

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.

Fleming, Lee, Hillary Greene, G Li, Matt Marx, and Dennis Yao. 2019. “Government-funded research increasingly fuels innovation.” *Science* 364 (6446): 1139–1141.

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.

Gollin, Douglas, and Christopher Udry. 2021. “Heterogeneity, measurement error, and misallocation: Evidence from African agriculture.” *Journal of Political Economy* 129 (1): 1–80.

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.”

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

Hegde, Deepak, and Bhaven Sampat. 2015. “Can private money buy public science? Disease group lobbying and federal funding for biomedical research.” *Management Science* 61 (10): 2281–2298.

Hill, Ryan, and Carolyn Stein. 2025. “Race to the bottom: Competition and quality in science.” *The Quarterly Journal of Economics* 140 (2): 1111–1185.

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.

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.

Ioannidis, John P.A. 2018. “Science Funding is Broken.”

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, Kyoo, Amil Petrin, and Suyong Song. 2016. “Estimating production functions with control functions when capital is measured with error.” *Journal of Econometrics* 190 (2): 267–279.

Kim, Scott Daewon, and Petra Moser. 2025. “Women in science. Lessons from the Baby Boom.” *Econometrica*.

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.

Levinsohn, James, and Amil Petrin. 2003. “Estimating production functions using inputs to control for unobservables.” *Review of Economic Studies* 70 (2): 317–341.

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

The University of Chicago Press.
Mills, M. Anthony. 2021. “Fix Science, Don’t Just Fund it.”
Myers, Kyle R. 2020. “The Elasticity of Science.” *American Economic Journal: Applied Economics* 12 (4): 103–134.
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.
Nordhaus, William D. 2004. “Schumpeterian profits in the American economy: Theory and measurement.”
Olley, G Steven, and Ariel Pakes. 1996. “The dynamics of productivity in the telecommunications equipment industry.” *Econometrica* 64 (6): 1263.
Petrin, Amil, and Jagadeesh Sivadasan. 2013. “Estimating lost output from allocative inefficiency, with an application to Chile and firing costs.” *Review of Economics and Statistics* 95 (1): 286–301.
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.
Piper, Kelsey. 2019. “Science funding is a mess. Could grant lotteries make it better?”.
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. 2017. “The causes and costs of misallocation.” *Journal of Economic Perspectives* 31 (3): 151–174.
Rohn, Jenny. 2017. “The way research is funded is harmful to science — it’s time for a change.”
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.
Sinatra, Roberta, Dashun Wang, Pierre Deville, Chaoming Song, and Albert-László Barabási. 2016. “Quantifying the evolution of individual scientific impact.” *Science* 354 (6312): aaf5239.
Sraer, David, and David Thesmar. 2023. “How to use natural experiments to estimate misallocation.” *American Economic Review* 113 (4): 906–938.
Stephan, Paula E. 1996. “The economics of science.” *Journal of Economic Literature* 34 (3): 1199–1235.
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.

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.

Wossen, Tesfamicheal, David J Spielman, Arega D Alene, and Tahirou Abdoulaye. 2024. “Estimating seed demand in the presence of market frictions: Evidence from an auction experiment in Nigeria.” *Journal of Development Economics* 167: 103242.

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

Supplemental Appendix

A. Example Applications of Methodology

A.1. Application 1: Constant Returns to Scale and Convex Costs

Consider first the case of a firm that operates a linear production function in a single input, e.g., labor, and faces convex adjustment costs because, e.g., there are hiring and firing frictions. The optimization problem is:

$$\max_{l_i} \quad \alpha_i l_i - w l_i - c l_i^{\psi} \tag{A1}$$

with $b(\cdot) = \alpha_i l_i + m_i$ and $m_i = 0$, $c(\cdot) = w l_i + c l_i^{\psi}$ ($c > 0$ and $\psi > 1$), and $\boldsymbol{\mu}_i = \boldsymbol{\mu} = (w, c, \psi)$. The optimality condition of the problem is:

$$\alpha_i - w - c\psi l_i^{\psi-1} = 0 \tag{A2}$$

In the first step of the estimation procedure, we evaluate the optimality condition at observed allocations to characterize α_i as a function of parameters and observed allocations:

$$\alpha_i(\hat{l}_i, w, c, \psi) = w + c\psi \hat{l}_i^{\psi-1} \tag{A3}$$

where $\hat{l}_i$ denotes the observed input allocation. Moreover, conditional on attributes, parameters, and prices, the solution is determined by:

$$l_i^* = \left(\frac{\alpha_i - w}{\psi c}\right)^{\frac{1}{\psi-1}} \tag{A4}$$

In the second step, we implement the thought experiment and we offer to the firm $\Delta > 0$ extra units of labor, which the firm can hire by just paying WTP_i dollars and without incurring the convex adjustment cost. Of course, the firm can re-optimize the quantity of labor that hires at market conditions. Therefore, the problem is:

$$\max_{\tilde{l}_i} \quad \alpha_i(\tilde{l}_i + \Delta) - w\tilde{l}_i - c\tilde{l}_i^{\psi} - WTP_i \tag{A5}$$

and the solution:

$$\tilde{l}_i^* = \left(\frac{\alpha_i - w}{\psi c}\right)^{\frac{1}{\psi-1}} = l_i^* \tag{A6}$$

Replacing the solution in the profit function and equating profits at current allocations and in the thought experiment, we obtain:

$$WTP_i = \alpha_i(\hat{l}_i, w, c, \psi)(\tilde{l}_i^* + \Delta) - w\tilde{l}_i^* - c\left(\tilde{l}_i^*\right)^{\psi} - \alpha_i(\hat{l}_i, w, c, \psi)l^* + wl^* + c\,(l^*)^{\psi} \tag{A7}$$

Using the fact that $\tilde{l}_i^* = l_i^*$, we obtain:

$$WTP_i = \alpha_i(\hat{l}_i, w, c, \psi)\Delta = \left(w + c\psi\hat{l}_i^{\psi-1}\right)\Delta \tag{A8}$$

which allows the identification of w, c, ψ and thus $\alpha_i(\hat{l}_i, w, c, \psi)$, given $\hat{l}_i$ and Δ. The same logic applies to the case of decreasing returns to scale, namely:

$$\max_{l_i} \;\; \alpha_i l_i^{\beta} - wl_i - cl_i^{\psi}$$

with $v(\cdot) = \alpha_i l_i^{\beta} + m_i$ for $m_i = 0$ and $\beta \in (0,1)$, $c(\cdot) = wl_i + cl_i^{\psi}$ for $c > 0$ and $\psi > 1$, and $\mu_i = \mu = (\beta, w, c, \psi)$. In the general case, the optimal input choice does not admit a closed-form expression, but we can show that $\frac{\partial \tilde{l}_i^*}{\partial \Delta} \neq -1$, i.e., the additional input does not perfectly crowd out the quantity of input sourced in the market. This is a sufficient condition such that WTP_i does depend on α_i and thus on all estimands μ.

A.2. Application 2: Decreasing Returns to Scale and Linear Costs

Here, we show an example of how a linear cost schedule renders our approach unable to recover productivity. The setting is:

$$\max_{l_i} \;\; \alpha_i l_i^{\beta} - wl_i$$

i.e., the problem coincides with the previous application for $\psi = 0$. The first order condition is $\beta\alpha_i l_i^{\beta-1} = w$, which identifies:

$$\alpha_i(\hat{l}_i, \beta, w) = \left(\frac{w}{\beta}\right)\hat{l}_i^{1-\beta} \tag{A9}$$

and determines the optimal allocation as:

$$l_i^* = \left(\frac{\beta\alpha_i}{w}\right)^{\frac{1}{1-\beta}} \tag{A10}$$

Under the thought experiment, the optimization problem becomes:

$$\max_{\tilde{l}_i} \quad \alpha_i(\tilde{l}_i + \Delta)^{\beta} - w\tilde{l}_i$$

which implies:

$$\tilde{l}_i^* = \left(\frac{\beta\alpha_i}{w}\right)^{\frac{1}{1-\beta}} - \Delta = l_i^* - \Delta \tag{A11}$$

Replacing in the profit function and deriving WTP_i, we obtain:

$$\begin{aligned} WTP_i =& \alpha_i(\tilde{l}_i^* + \Delta)^{\beta} - w\tilde{l}_i^* - \alpha_i l_i^* + wl_i^* \\ =& \alpha_i(l_i^* - \Delta + \Delta)^{\beta} - w(l_i^* - \Delta) - \alpha_i l_i^* + wl_i^* \\ =& w\Delta \end{aligned}$$

where from the first to the second line we replace $\tilde{l}_i^* = l_i^* - \Delta$. Because the willingness to pay just depends on w, β cannot be identified, which implies that also α_i is not pinned down.

A.3. Application 3: Individuals as Producers with Utility Function

Consider the problem of an agent that gets utility from income m_i and output $\alpha_i d_i^{\beta} l_i^{1-\beta}$—produced using a fixed input d_i and labor l_i—and gets disutility from working. The problem is:

$$\max_{l_i} \quad \ln m_i + \left(\alpha_i d_i^{\beta} l_i^{1-\beta}\right)^{\eta} - \phi l_i^{\psi} \tag{A12}$$

with $b(\cdot) = \ln mi_i + \left(\alpha_i d_i^{\beta} l_i^{1-\beta}\right)^{\eta}$ for $\beta \in (0,1)$ and $\eta \in (0,1)$, $c(\cdot) = \phi l_i^{\psi}$ for $\phi > 0, \psi > 1$, and $\mu_i = \mu = (\beta, \eta, \phi, \psi)$. The optimality condition is:

$$\eta(1-\beta)\alpha_i^{\eta} d_i^{\eta\beta} l_i^{(1-\beta)\eta-1} = \phi\psi l_i^{\psi-1} \tag{A13}$$

Therefore:

$$\alpha_i(\hat{l}_i, d_i, \beta, \eta, \phi, \psi) = \left(\frac{\phi\psi\hat{l}_i^{\psi-1}}{\eta(1-\beta)d_i^{\eta}\hat{l}_i^{\eta(1-\beta)-1}}\right)^{\frac{1}{\eta}} \tag{A14}$$

and

$$l_i^* = \left(\frac{\eta(1-\beta)\alpha_i^{\eta} d_i^{\eta\beta}}{\phi\psi}\right)^{\frac{1}{\psi-\eta(1-\beta)}} \tag{A15}$$

In the setting of a thought experiment where Δ additional units of the fixed input are offered to the agent, the problem becomes:

$$\max_{\tilde{l}_i} \ \ln(m_i - WTP_i) + \left(\alpha_i(d_i+\Delta)^{\beta}\tilde{l}_i^{1-\beta}\right)^{\eta} - \phi\tilde{l}_i^{\psi} \tag{A16}$$

with

$$\tilde{l}_i^{*} = \left(\frac{\eta(1-\beta)\alpha_i^{\eta}(d_i+\Delta)^{\eta\beta}}{\phi\psi}\right)^{\frac{1}{\psi-\eta(1-\beta)}}$$

Therefore, WTP_i solves:

$$\ln(m_i - WTP_i) + \alpha_i^{\eta}\left(\frac{\eta(1-\beta)}{\phi\psi}\right)^{\frac{\eta(1-\beta)}{\psi-\eta(1-\beta)}}(d_i+\Delta)^{\frac{\beta\psi\eta}{\psi-\eta(1-\beta)}} - \phi\left(\frac{\eta(1-\beta)\alpha_i^{\eta}(d_i+\Delta)^{\eta\beta}}{\phi\psi}\right)^{\frac{\psi}{\psi-\eta(1-\beta)}} +$$
$$-\ln m_i - \alpha_i^{\eta}\left(\frac{\eta(1-\beta)}{\phi\psi}\right)^{\frac{\eta(1-\beta)}{\psi-\eta(1-\beta)}} d_i^{\frac{\beta\psi\eta}{\psi-\eta(1-\beta)}} - \phi\left(\frac{\eta(1-\beta)\alpha_i^{\eta}d_i^{\eta\beta}}{\phi\psi}\right)^{\frac{\psi}{\psi-\eta(1-\beta)}} = 0 \tag{A17}$$

and depends on all parameters.

B. Additional Survey Statistics and Comparisons

Table B1 shows the effects of the randomized participation incentives and reminders on survey completion. Figures B1–B2 illustrate sample representativeness and Figure B3 documents the alignment between self- and publicly-reported salaries (see Myers et al. (2023) for more on the sample). Table B2 describes the summary statistics for the sample on numerous dimensions. Table B3 shows the pairwise correlations of the subjective output measures.

TABLE B1. 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

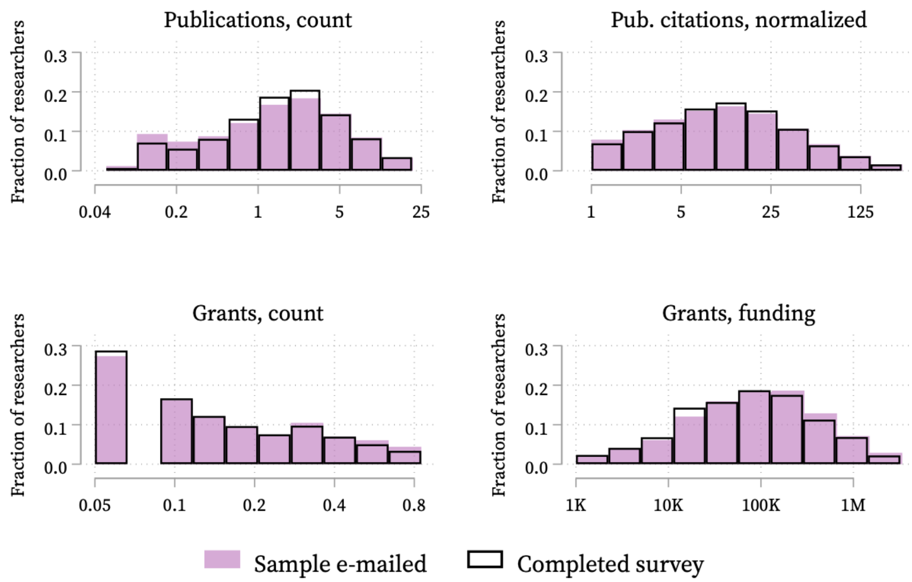
Note: Reports estimates from a regression of a binary indicator of survey completion on binary indicators for the randomized incentives and reminders including observations for all researchers emailed. Robust standard errors reported; * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

TABLE B2. 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
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. *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.

FIGURE B1. Sample Representativeness per Publication and Grant Funding Measures



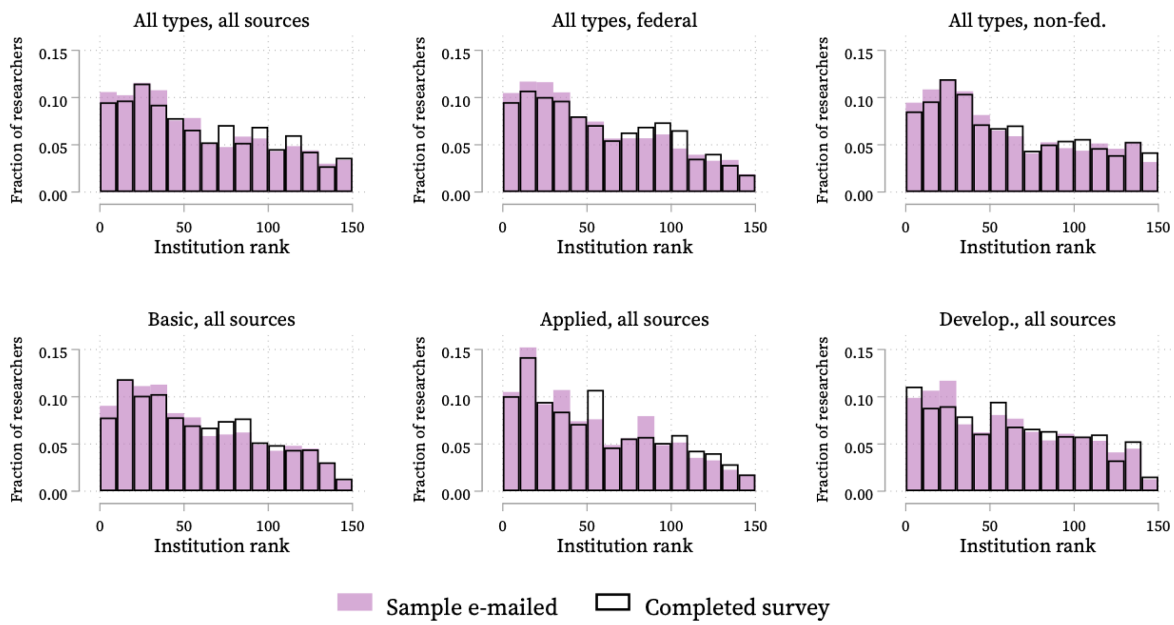
Note: Shows the distribution of publication and grant outcomes split by whether the researcher was emailed (i.e., the full sample) versus those who completed the survey; note the log x axes. See Myers et al. (2023) for regression-based estimates of the differences.

TABLE B3. 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}	

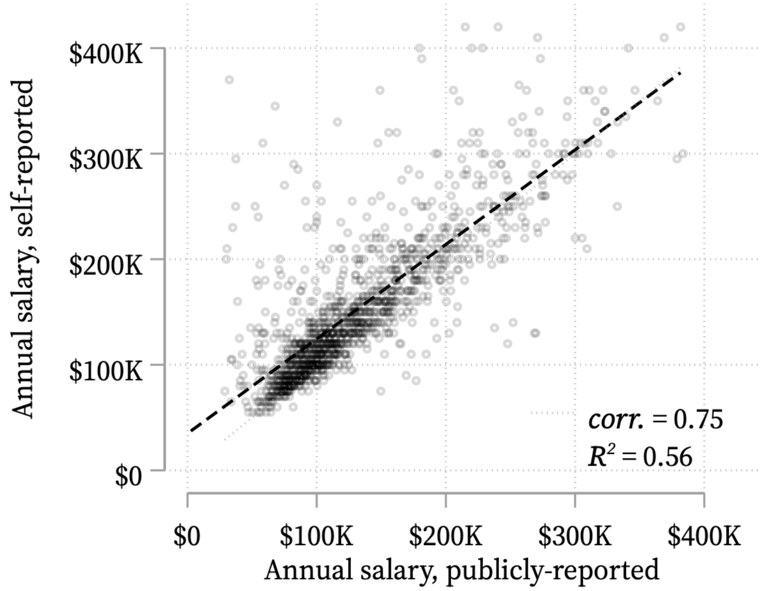
Note: Reports pairwise correlations for the four subjective measures of researchers' intended output types (Journal articles; Books; Materials or Methods; Products) and the four subjective measures of researchers' intended audiences (Academic peers; Policymakers; Businesses and organizations; General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

FIGURE B2. Sample Representativeness per Institutional Funding Measures



Note: Shows the distribution of institution-level funding split by whether the researcher was emailed (i.e., the full sample) versus those who completed the survey. See Myers et al. (2023) for regression-based estimates of the differences.

FIGURE B3. 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.

C. Additional Model and Experiment Details

C.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 (7) for each individual scientist i as a function of states $\mathbf{S}_i$, attributes θ_i, and parameters μ_i. They determine indirect utility $\mathcal{V}_i^* = \mathcal{V}(\mathbf{S}_i, \theta_i, \mu_i)$ at current allocations. Substituting the constraints, the utility maximization problem is:

$$\mathcal{V}(\cdot) = \max_{F_i, H_i} u_{1i}(M_i) + u_{2i}(\alpha_i (B_{\min} + G_i + \phi_i F_i)^{\gamma_i} (H_i - F_i - D_i)^{1-\gamma_i}) - u_{3i}(H_i, D_i) , \tag{C1}$$

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

$$\frac{\partial u_{2,i}(Y_i)}{\partial Y_i}\frac{\partial Y_i}{\partial F_i} + \lambda_{F,i} = 0 \tag{C2}$$

$$\frac{\partial u_{2,i}(Y_i)}{\partial Y_i}\frac{\partial Y_i}{\partial H_i} - \frac{\partial u_{3,i}(H_i, D_i)}{\partial H_i} - \lambda_{H,i} = 0 , \tag{C3}$$

and $\mathcal{R}(\cdot,\cdot,\cdot)$ being residually determined based on the time-constraint (7c). To derive the policy functions, we start from model's optimality conditions (C2) and (C3) and evaluate them using the functional forms described in the main text. We obtain the conditions:

$$Y_i^{1-\eta_i}\left[\gamma_i\phi_i(B_{\min} + G_i + \phi_i F_i)^{-1} - (1-\gamma_i)(H_i - D_i - F_i)^{-1}\right] + \lambda_{F,i} = 0 \tag{C4}$$

$$Y_i^{1-\eta_i}(H_i - D_i - F_i)^{-1}(1-\gamma_i) - \psi(H_i - D_i + D_i^{\xi_i})^{\zeta_i} - \lambda_{H,i} = 0 . \tag{C5}$$

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 fundraising time allocation is a corner solution ($F_i = 0$) or not ($F_i > 0$). We identify the threshold $H_{i,F>0}$ below which optimal fundraising is zero by solving (C5) for F_i as a function of H_i assuming that F_i is strictly positive, i.e., $\lambda_{F,i} = 0$:

$$F_i = \gamma_i(H_i - D_i) - \frac{1-\gamma_i}{\phi_i}(B_{\min} + G_i) . \tag{C6}$$

Therefore, (C6) implies that fundraising time is strictly positive as long as:

$$H_i > D_i + \frac{1-\gamma_i}{\gamma_i\phi_i}(B_{\min} + G_i) , \tag{C7}$$

which identifies the threshold $H_{i,F>0}$ that 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 (C5), which takes the piece-wise functional form:

$$\begin{cases} (1-\gamma_i)[\alpha_i(B_{\min}+G_i)^{\gamma_i}]^{1-\eta_i}(H_i-D_i)^{(1-\gamma_i)(1-\eta_i)-1} - \psi(H_i - D_i + D_i^{\xi_i})^{\zeta_i} - \lambda_{H,i} = 0 & \text{if } H_i \leq H_{i,F>0} \\ \left[\alpha_i(\phi_i\gamma_i)^{\gamma_i}(1-\gamma_i)^{1-\gamma_i}\right]^{1-\eta_i}\left(H_i - D_i\frac{B_{\min}+G_i}{\phi_i}\right)^{-\eta_i} - \psi(H_i - D_i + D_i^{\xi_i})^{\zeta_i} - \lambda_{H,i} = 0 & \text{if } H_i > H_{i,F>0} \end{cases} \tag{C8}$$

and defines an implicit solution $\mathcal{H}_i$ for H_i as a function of all parameters and state variables. (C8) is continuous but not differentiable at $H_{i,F>0}$. Moreover, the Lagrange multiplier $\lambda_{H,i}$ equals zero for $H_i < H_{\max}$ and H_i is always strictly larger than D_i because $\lim_{H_i \to D_i}$ (C8) $= +\infty$. 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 are:

$$\mathcal{F}_i = \begin{cases} 0 & \text{if } \mathcal{H}_i \leq H_{i,F>0} \\ \gamma_i(\mathcal{H}_i - D_i) - \frac{1-\gamma_i}{\phi_i}(B_{\min}+G_i) & \text{if } \mathcal{H}_i > H_{i,F>0} \end{cases} \tag{C9}$$

and

$$\mathcal{R}_i = \mathcal{H}_i - D_i - \mathcal{F}_i \,. \tag{C10}$$

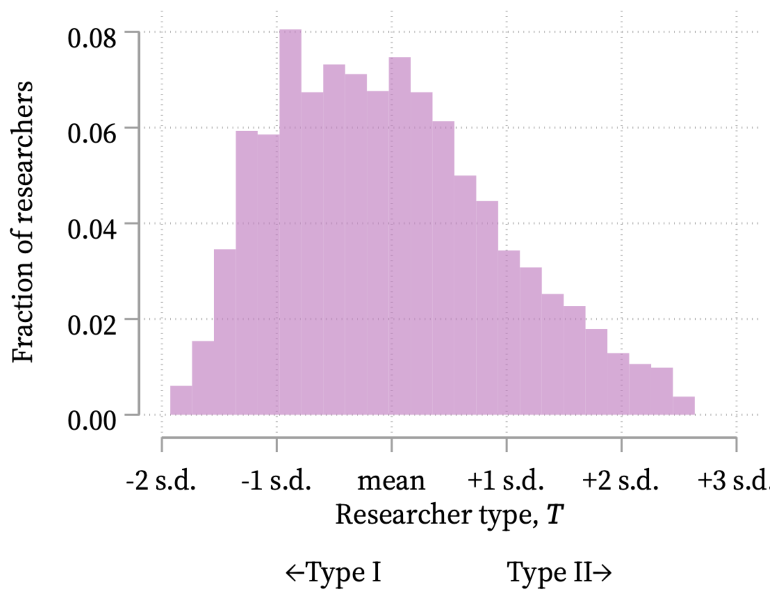
C.2. Reducing Dimensions of Heterogeneity: *k*-means Clustering Results

Ideally, our model would allow for rich heterogeneity in researchers' preferences (i.e., the $u_{1,i}(\cdot)$, $u_{2,i}(\cdot)$, and $u_{3,i}(\cdot)$ functions). This would help ensure that (true) variation in preferences would not mistakenly be attributed as (estimated) variation in productivity. We lack enough data to estimate researcher-specific preference functions, but we do have the large vector of observables from the survey $\mathbf{Z}_i$.[28] This motivates the following approach. We use *k*-means clustering to collapse the multi-dimensional heterogeneity from the observables $\mathbf{Z}_i$ into a single-dimensional index, T_i. Specifically, we assume there are two clusters of researcher types (*k*=2), and we use *k*-means clustering to estimate each researcher's euclidean distance from one of those groups; we use that distance as T_i, which provides a smooth, continuous measure of heterogeneity. Thus, for the preference function parameter σ_i, we assume that $\sigma_i = \exp(\delta_{\sigma,0} + T_i\delta_\sigma)$. Figure C1 shows the distri-

[28] We could allow the parameters that govern the preference functions to depend on deep parameters and these observables. For example, for the preference function parameter σ_i, we could assume that $\sigma_i = \exp(\delta_{\sigma,0} + \sum_z Z_i\delta_{\sigma,Z})$. However, this presents some practical estimation challenges (i.e., estimating approximately (35×3=) 105 additional parameters; the $\delta_{\sigma,z}$ parameters in the previous example).

bution of the one-dimensional euclidean similarity scores. Researchers with larger values of T_i tend to be, among other things, older, white, males that are full professors in the humanities and natural sciences. A table reporting the mean differences between the two types of researchers is available upon request.

FIGURE C1. 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.

C.3. Estimation Outline

We infer individual attributes and parameters 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. We use survey information on 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.

Conditional on some common parameters, we can infer individual-specific attributes θ_i from researchers' optimality conditions and model's structure. One important note is that we have separate estimation routines 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, $\forall i = 1, ..., N_F$:

$$\widehat{\phi}_i = \frac{EG_i}{F_i} . \tag{C11}$$

Expression (C11) is well-defined if and only if observed $F_i > 0$, which is why the inference strategy must differ for researchers with $F_i = 0$ at observed allocations.[29] 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 C.1. For the chosen functional forms, this takes the expression:

$$\widehat{\gamma}_i = \frac{B_i}{B_i + \phi_i R_i} = \frac{B_{\min} + G_i + \widehat{\phi}_i F_i}{B_{\min} + G_i + \widehat{\phi}_i F_i + \widehat{\phi}_i R_i} . \tag{C12}$$

Equation (C12) 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 $\widehat{\gamma}_i$:

$$(1 - \widehat{\gamma}_i)\widehat{\alpha}_i^{1-\eta_i}(B_{\min} + G_i + \widehat{\phi}_i F_i)^{(1-\eta_i)\widehat{\gamma}_i} R_i^{(1-\eta_i)(1-\widehat{\gamma}_i)-1} = \psi(H_i - D_i + D_i^{\xi_i})^{\zeta_i} , \tag{C13}$$

This equation holds exactly if $H_i < H_{\max}$. Therefore, Equations (C11), (C12), and (C13) 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 (C11), nor can we compute $\widehat{\gamma}_i$.[30] 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 $F_i > 0$.[31]

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

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

[31] 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.

Finally, given the estimates attributes $\widehat{\phi}_i$ and $\widehat{\gamma}_i$, the state variables, and the vector of common parameters $\widehat{\mu}$, we infer productivity beliefs $\widehat{\alpha}_i$ through Equation (C13). Given estimates of individual attributes $\widehat{\theta}_i = (\widehat{\alpha}_i, \widehat{\gamma}_i, \widehat{\phi}_i)$ as functions of parameters $\mu = (\omega, \psi, \delta)$, calibrated values $(B_{\min}, H_{\max})$, and observed allocations, we estimate the vector μ by generalized method of moments as:

$$\widehat{\mu} = \arg\min_{\mu} \mathcal{L}(\mu) = \arg\min_{\mu} \sum_{i=1}^{N} \sum_{j=1}^{4} \left(M_{ij}^{\text{obs}} - \widehat{M}_{ij}(\mathbf{S}_{ij}, \widehat{\theta}_i(\mu_i(\mu, T_i)), \mu_i(\mu, T_i)) \right)^2 , \tag{C14}$$

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 μ which determines individual attributes $\widehat{\theta}_i$ and individual-specific parameters μ_i.[32] Parameter estimates $\widehat{\mu}$ solve (C14) conditional on parameter restrictions specified in Section 2.2. Appendix C.4 provides additional details on the search algorithm.

C.4. Estimation Algorithm

Here, we describe the estimation of the common parameters $\mu = (\omega, \psi, \delta_{\sigma,0}, \delta_{\sigma,1}, \delta_{\eta,0}, \delta_{\eta,1}, \delta_{\xi,0}, \delta_{\xi,1}, \delta_{\zeta,0}, \delta_{\zeta,1})$. First, we define a grid of parameter values at which we perform a preliminary evaluation of the loss function in (C14). The grid is defined by the Cartesian product of the following:

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

Second, we select the grid-points where the value of the loss function is within 0.5% of the minimum, and we use them as initial points for the numerical solution. We perform a preliminary "Nelder-Mead simplex direct search" 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 minimum, and we use them

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

as initial points 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.

FIGURE C2. 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:

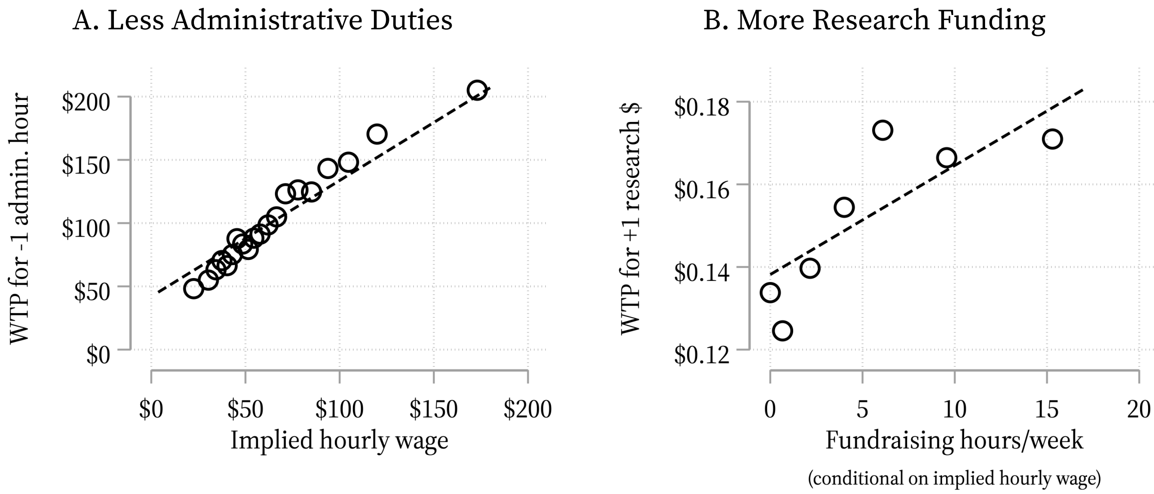
	Guaranteed research funding	*Annual salary*
Currently:	$400,000 total	$160,000
New offer: *more funding smaller salary*	$1,400,000 total	$______?

Q. What is the <u>smallest</u> 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 C3. 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.	✓	✓	✓	✓	✓	✓	✓	✓
Z		✓		✓		✓		✓
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; **Z** includes the full vector of variables that comprise the type index.

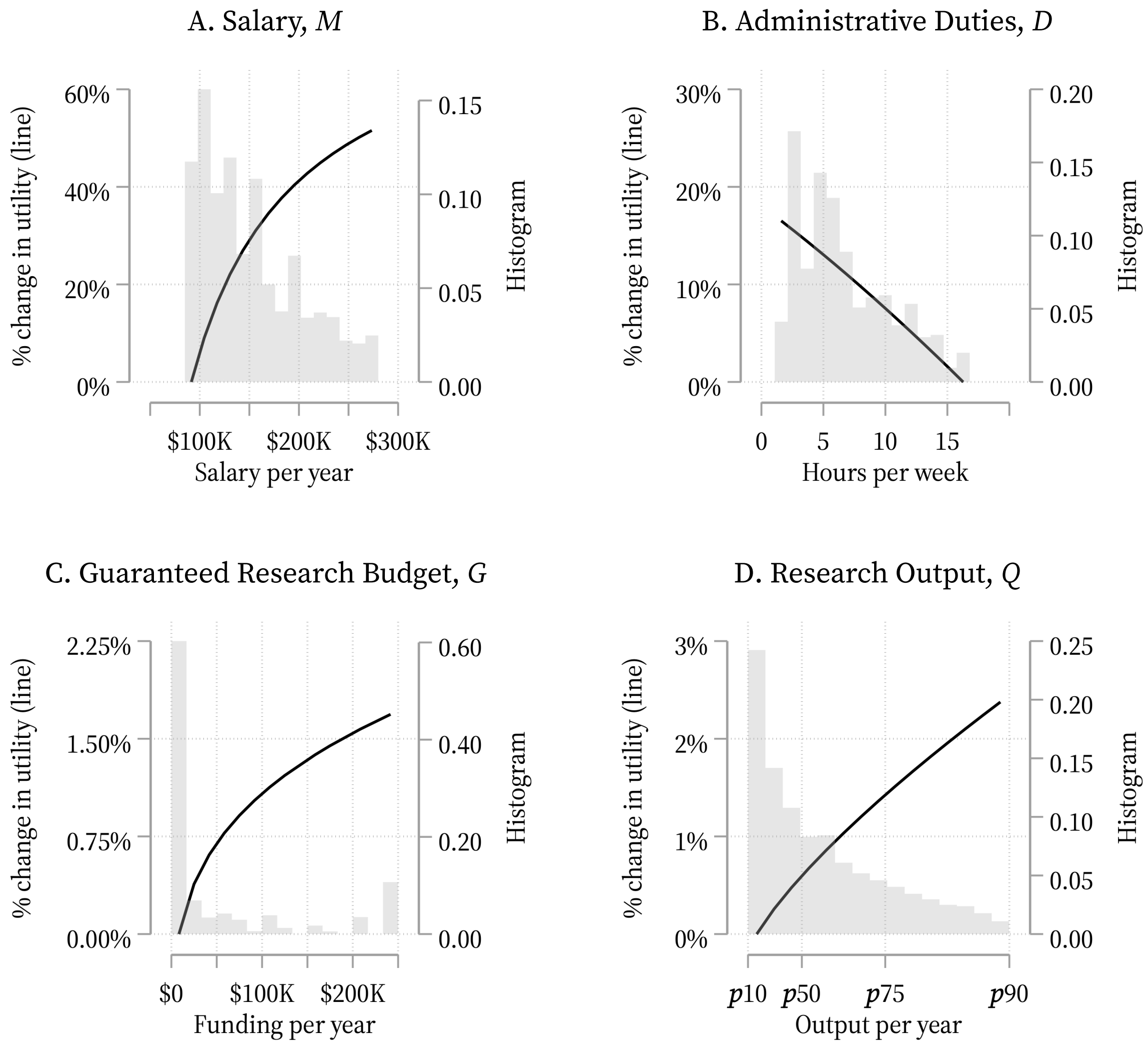
TABLE C2. Potential Role of Non-response Bias and Stated Preference WTP Bias

	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; $^{*}\ p < 0.10$, $^{**}\ p < 0.05$, $^{***}\ p < 0.01$.

D. The Average Researcher's Utility Function

FIGURE D1. 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. For simplicity, excludes behavioral responses.

E. Trends in Science and Metascience

Data on total US R&D spending is sourced from the National Center for Science and Engineering Statistics (NCSES).[33] Data on meta-science studies is sourced from (*i*) the *OpenAlex* publication database, which covers virtually all fields of science,[34] and (*ii*) the *PubMed* publication database, which focuses primarily on the life and biomedical sciences.[35]

OpenAlex is valuable in that its coverage spans a tremendous range of scientific fields; however, the categorization scheme used to index publications by topics is relatively coarse and lacks a clear set of topic codes for meta-science. In order to identify meta-science publications in OpenAlex, we restrict our focus to all journal-article publications flagged with the following topic codes: *scientometrics and bibliometrics research; scientific research and philosophical inquiry; philosophy and history of science; academic writing and publishing; science and science education; international science and diplomacy; innovation, technology, and society; research, science, and academia*.

PubMed is clearly a strict subset of scientific fields; however, publications are indexed with relatively fine-grain topic codes that facilitate the identification of meta-scientific studies. In order to identify meta-science publications in PubMed, we restrict our focus to all journal-article publications flagged with the following topic codes: *efficiency; research personnel; research support as topic; grants peer review; knowledge discovery; scholarly communication; allocative efficiency; grants; funding, capital*.

Clearly, in both cases, the set of topics we use to flag meta-science will include some false positives, and we surely have a non-trivial number of false negatives. However, we have no reason to suspect these error rates should be evolving significantly over time.

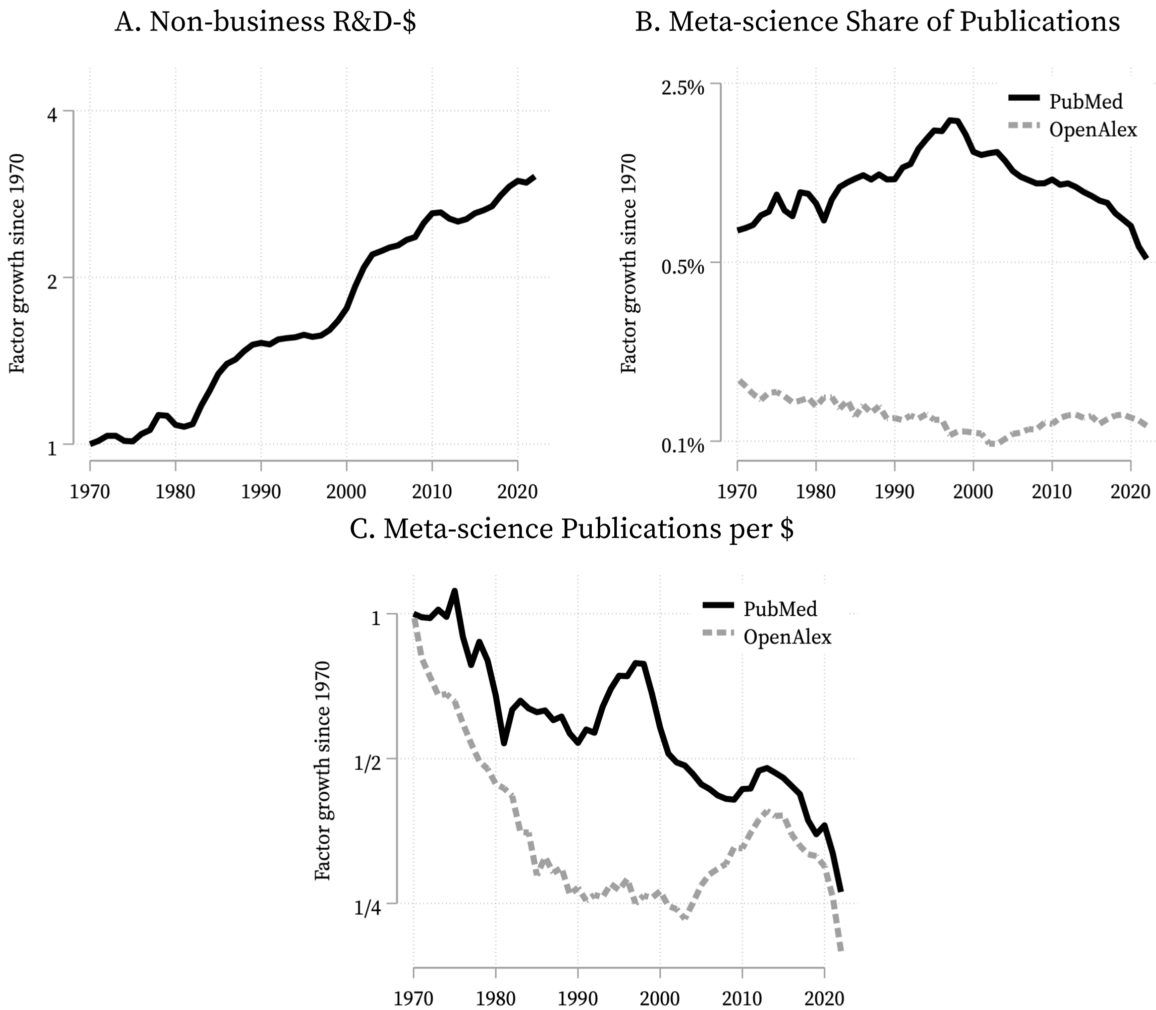
Figure E1 reports the trends for the past 50 years for total non-business R&D in real-$ (see Panel A), the share of all publications focused on meta-science (see Panel B), and the count of meta-science studies per $ shown in Panel A (see Panel C).

[33] See: National Center for Science and Engineering Statistics (NCSES). 2025. *U.S. R&D Totaled $892 Billion in 2022; Estimate for 2023 Indicates Further Increase to $940 Billion*. NSF 25-327. Alexandria, VA: U.S. National Science Foundation. Available at https://ncses.nsf.gov/pubs/nsf25327/.

[34] See: Priem, J., Piwowar, H., & Orr, R. (2022). OpenAlex: A fully-open index of scholarly works, authors, venues, institutions, and concepts. ArXiv. https://arxiv.org/abs/2205.01833.

[35] See: PubMed [Internet]. Bethesda (MD): National Library of Medicine (US), National Center for Biotechnology Information; 2004 – 2025. Available from: https://pubmed.ncbi.nlm.nih.gov.

FIGURE E1. Relative Growth in Science and Meta-science since 1970



Note: Shows the relative growth in real, non-business R&D spending in the US since 1970 (Panel A), the share of publications focused on meta-science (Panel B), and the count of meta-science publications per non-business R&D dollar (Panel C), using either the OpenAlex or PubMed publication databases.