Decoding team and individual impact in science and invention

Mohammad Ahmadpoor[a,b] **and Benjamin F. Jones**[a,b,c,1]

[a]Strategy Department, Kellogg School of Management, Northwestern University, Evanston, IL 60208; [b]Northwestern University Institute on Complex Systems and Data Science, Evanston, IL 60208; and [c]National Bureau of Economic Research, Cambridge, MA 02138

Edited by Jose A. Scheinkman, Columbia University, New York, NY, and approved May 28, 2019 (received for review July 20, 2018)

Scientists and inventors increasingly work in teams, raising fundamental questions about the nature of team production and making individual assessment increasingly difficult. Here we present a method for describing individual and team citation impact that both is computationally feasible and can be applied in standard, wide-scale databases. We track individuals across collaboration networks to define an individual citation index and examine outcomes when each individual works alone or in teams. Studying 24 million research articles and 3.9 million US patents, we find a substantial impact advantage of teamwork over solo work. However, this advantage declines as differences between the team members' individual citation indices grow. Team impact is predicted more by the lower-citation rather than the higher-citation team members, typically centering near the harmonic average of the individual citation indices. Consistent with this finding, teams tend to assemble among individuals with similar citation impact in all fields of science and patenting. In assessing individuals, our index, which accounts for each coauthor, is shown to have substantial advantages over existing measures. First, it more accurately predicts out-of-sample paper and patent outcomes. Second, it more accurately characterizes which scholars are elected to the National Academy of Sciences. Overall, the methodology uncovers universal regularities that inform team organization while also providing a tool for individual evaluation in the team production era.

team science | collaboration | prediction | team organization

Teams are increasingly prevalent across virtually all fields of science and patenting (1–4), raising fundamental questions about the nature of team-based creativity and team assembly and creating fundamental challenges for individual assessment (5–11). For example, while Heisenberg developed his uncertainty principle without building a team and received credit in a straightforward manner as the solo author, more recent breakthroughs, such as Milstein and Kohler's monoclonal antibodies and Faggin, Hoff, and Mazor's microprocessor, often come from collaborations that both combine and obscure individual contributions (2, 4, 5). Here we investigate two intertwined questions. First, how do individuals combine to predict team output? Second, how can individual impact be inferred when people work in teams?

Concretely, consider a paper written by two individuals. At one extreme, the team outcome could be a max process, $y = max\{a_{low}, a_{high}\}$, where y is the success of the joint outcome, a_i is an index characterizing each individual team member, and $a_{high} \geq a_{low}$. In this max specification, the joint output is determined by the higher-index individual; for example, perhaps this individual, by shaping the research question and methods, drives the ultimate success of the project. By contrast, at the other extreme, team outcomes could be a min process, $y = min\{a_{low}, a_{high}\}$, where the joint result is determined by the lower-index individual. For example, perhaps this team member creates bottlenecks at certain tasks and determines the ultimate outcome. Alternatively, the outcome may lie between these max and min extremes, perhaps as the arithmetic, geometric, or other mean of the individual indices.

These alternative views have fundamentally different—indeed, opposite—implications for science. Organizationally, in a max specification, a team could expect a successful outcome so long as one person has a high index, and an organization might sprinkle around its best people to great effect (12–14). However, in a min specification, the opposite is true. Here the person with the lowest index on a team would determine the outcome, and the collective output of science would be greatest not by sprinkling the top people around but rather through positive assortative matching, where individuals of similar index measures work together (14–16). Credit considerations in collaboration (5, 10, 17, 18) are also germane; in a max specification, audiences would reward the top author, akin to some versions of the Matthew effect (5), but in a min specification the joint outcome is informative for the lowest-index member of the team (17). Of course, the true relationship may lie between these max and min extremes.

This paper introduces a transparent and computationally feasible method for informing the relationship between individual and team outcomes. This descriptive approach is applied both to reveal central facts about science and invention and to predict individual and team results. We leverage the generalized mean (or Hölder mean) to write

$$y = \beta_n \left[\frac{1}{n} \sum_{i=1}^{n} a_i^{\rho} \right]^{\frac{1}{\rho}}, \qquad [1]$$

where y is the outcome and n is the team size. The parameters a_i track individuals across their works to estimate a fixed effect for

Significance

Scientists and inventors increasingly work in teams. We track millions of individuals across their collaboration networks to help inform fundamental features of team science and invention and help solve the challenge of assessing individuals in the team production era. We find that in all fields of science and patenting, team impact is weighted toward the lower-impact rather than higher-impact team members, with implications for the output of specific teams and team assembly. In assessing individuals, our index substantially outperforms existing measures, including the h index, when predicting paper and patent outcomes or when characterizing eminent careers. The findings provide guidance to research institutions, science funders, and scientists themselves in predicting team output, forming teams, and evaluating individual impact.

Author contributions: M.A. and B.F.J. designed research, performed research, analyzed data, and wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

Data deposition: The patent data sets and NAS publication data have been deposited in Figshare (https://doi.org/10.6084/m9.figshare.8242571).

[1]To whom correspondence may be addressed. Email: bjones@kellogg.northwestern.edu.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1812341116/-/DCSupplemental.

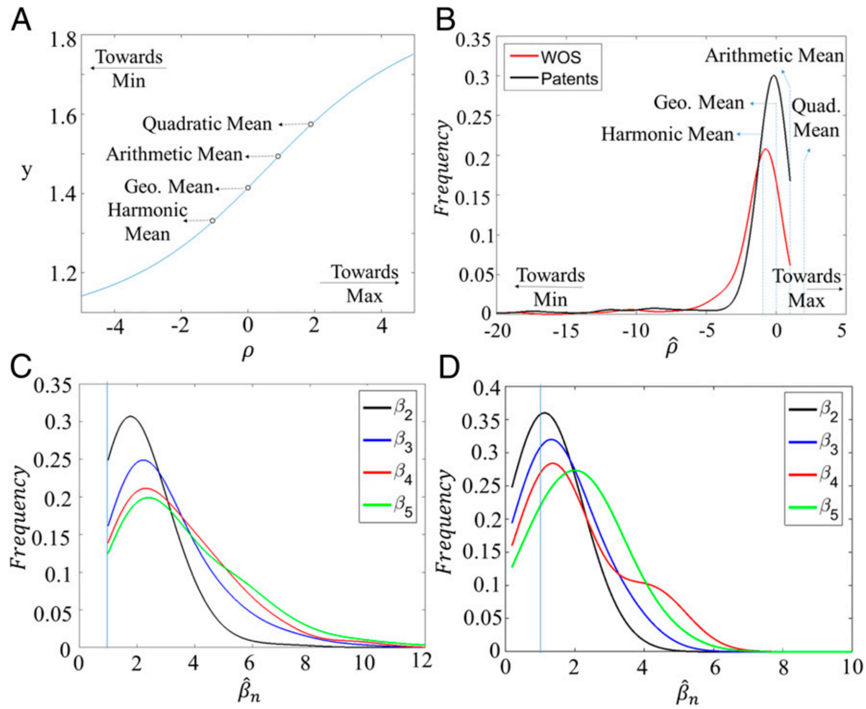


Fig. 1. The generalized mean. (*A*) An example of the generalized mean function for two individuals. (*B*) The distribution of the generalized mean parameter $\hat{\rho}$ across Web of Science fields (red) and patenting fields (black). (*C*) The distributions of the team impact parameters ($\hat{\beta}_2, \ldots, \hat{\beta}_5$) across Web of Science fields. (*D*) The distributions of the team impact parameters ($\hat{\beta}_2, \ldots, \hat{\beta}_5$) across patenting fields.

individual i on a per-paper (or per-patent) basis. The key team parameter is ρ, which defines how the individual parameters a_i combine. At the extremes, the Hölder mean allows for the max ($\rho \to \infty$) and min ($\rho \to -\infty$) functions while also incorporating other means, including the arithmetic mean ($\rho = 1$), geometric mean ($\rho = 0$), and harmonic mean ($\rho = -1$) as special cases (Fig. 1*A*). An important intuition is that the person with the lowest (highest) a_i becomes more influential for the joint output as ρ declines (increases). The arithmetic mean provides the boundary where each individual is equally important.

In addition, the parameter β_n captures impact benefits associated with teamwork (specifically, for a team of size n), including advantages of aggregating effort, skill, or marketing, as well as disadvantages through coordination costs in teams (1, 2, 4). We normalize the model by setting $\beta_1 = 1$ for solo-authored work. This normalization implies that $y = a_i$ for solo-authored work. Thus, the individual index (the estimated a_i) is interpreted as the expected outcome when that person works alone. Further, taking a team of size n, the magnitude of $\hat{\beta}_n$ is interpreted as the outcome advantage of teamwork over solo-work when the individual team members share a common value of a_i.

We estimate this function, by field, in two large datasets. First, for research articles, we examine all 182 different fields of science, engineering, social sciences, and arts and humanities in the WOS that have at least 500 papers in the field. Second, for patents, we examine all 384 different primary technology classes of the US Patent and Trademark Office (USPTO) that have at least 500 patents in the class. The estimates further deploy name disambiguation to identify a given individual across a body of their work. For the WOS, we use Thomson Reuters' name-disambiguated author dataset (19–21). For the USPTO data, we use Li et al.'s (22, 23) name-disambiguated inventor dataset. We further restrict the data to the 97% of papers and 99% of patents with team sizes of eight or fewer members (24). The team outcome measure in our main analyses is the number of citations received by the paper or patent in the first 8 y after publication (1). We consider robustness to alternative outcome measures in the *SI Appendix*, which also provides further details about these datasets. Our final estimation samples include 24 million research articles written by 13 million individuals (WOS, 1945–2005 period) and 3.9 million patents produced by 2.6 million individuals (USPTO, 1975–2006 period).

Results

Fig. 1*B* presents the distribution of the estimated $\hat{\rho}$ across fields. We see substantial similarity in the science and patenting domains. First, in all fields of science and patenting, we find $\hat{\rho} < 1$. This finding indicates that while everyone on the team has influence, team output is weighted toward the lower-index rather than the higher-index members of the team. This finding is robust to various computational checks (*SI Appendix*) and consistent with raw data analysis as we will show below. The generality of this finding—appearing across diverse fields of sciences, engineering, social sciences, and disparate technology areas of invention, many of which feature different norms and institutions—indicates a profound regularity to team-based research outcomes. Second, we see that the modal field in both the science and patenting domains centers below the geometric average, with median values near the harmonic average ($\hat{\rho}_{median} = -1.49$ for paper fields and $\hat{\rho}_{median} = -0.95$ for patent fields). Third, the distribution is asymmetric toward lower $\hat{\rho}$, with a substantial mass of fields below the harmonic average and a long left tail stretching toward the min specification.

Fig. 1*C* presents the distributions of $\hat{\beta}_2$ through $\hat{\beta}_5$ across fields for the Web of Science (WOS), and Fig. 1*D* presents these distributions for patents. Consistent with literature showing an impact advantage of teams over solo authors in raw data (1, 2, 25), we find that these team-impact parameters are large on average. Focusing on two-person teams, we see that $\hat{\beta}_2 > 1$ for 99% of WOS fields and for 94% of patenting fields. The median

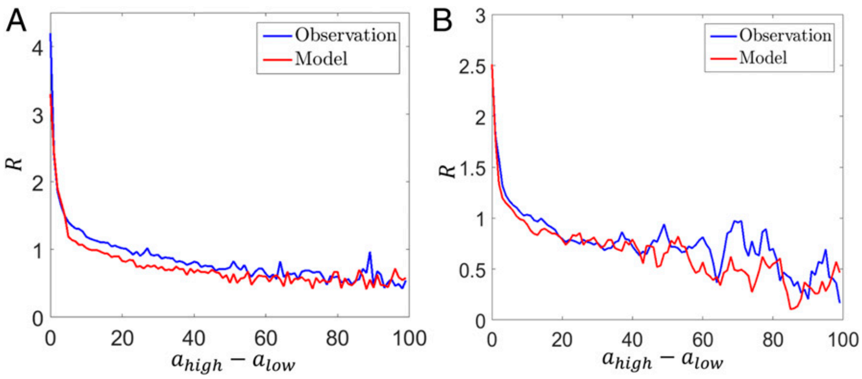


Fig. 2. Team impact. We examine different pairings of individuals in two-person teams. (*A*) The raw data (blue) and the model prediction (red) for the Web of Science. (*B*) The raw data (blue) and the model prediction (red) for US patents. The *x* axis is the difference in individual citation impact, $a_{high} - a_{low}$, between the two authors. The *y* axis is the normalized team outcome, measured as the ratio of the team citation outcome to the arithmetic mean of the team members' individual citation outcomes (see text). We see that the team impact advantage is large when the team members have similar individual impact measures but declines as the difference in individual impact widens within the team.

value is $\hat{\beta}_2 = 2.05$ for papers and $\hat{\beta}_2 = 1.44$ for patents, which rises further for larger teams, with some evidence that the teamwork advantage flattens for team sizes above 4. Notably, these findings indicate a team impact advantage, even when controlling for individual citation impact measures. Thus, the team advantage seen in prior literature (1, 2, 25) is not simply about higher-citation people tending to work in teams but rather appears conditional on the citation impact of the individual team members (10). *SI Appendix*, Tables S1 and S2, provides the estimated $\hat{\rho}$ and $\hat{\beta}_2$ through $\hat{\beta}_5$ for each field of science and patenting.

We thus see two offsetting features in team outcomes. There tends to be an impact advantage of teamwork over solo work ($\hat{\beta}_n > 1$), but this advantage declines as the gap between the team members' individual citation indices grows ($\hat{\rho} < 1$). On net, because the $\hat{\beta}_n$ values tend to be substantially greater than 1, teamwork tends to predict higher impact so long as the gap between the individuals is not itself substantial. Thus, individuals with different citation indices can still see higher impact when working together than working alone. We further find a negative relationship between a field's $\hat{\rho}$ and $\hat{\beta}_2$ (*SI Appendix*, Table S6 and Figs. S1 and S2). This relationship is consistent with a division of labor interpretation (4, 7, 25) where specialization may create substantial teamwork advantages (higher $\hat{\beta}_2$) but also accentuate bottlenecks in production (lower $\hat{\rho}$).

To develop further intuition for these findings and visually examine the fit of the model, we consider different pairings of individuals in two-person teams. We examine the ratio

$$R = \frac{y}{\frac{1}{2}\left(a_{low} + a_{high}\right)}, \quad [2]$$

where y is the team-based outcome for two individuals and a_{low} and a_{high} are their individual citation indices. Conceptually, $R = 1$ occurs when the team-based outcome is equivalent to the simple arithmetic average of the individual indices, while R will be greater (lower) than 1 if the team-based outcome outperforms (underperforms) the arithmetic average of the individual citation indices.

We first examine raw data, presenting a model-free analog of R. Here we measure y as the observed citation impact of the dual-authored paper and measure each a_i using each individual's solo-authored work and taking the arithmetic mean citation impact of that work. For the modeled version of R, we instead take

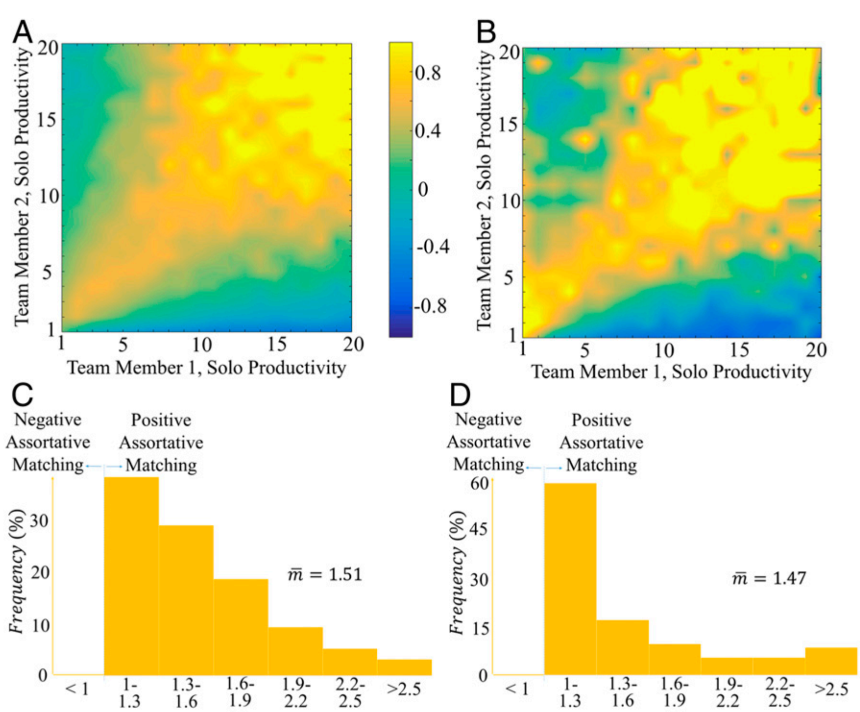


Fig. 3. Team assembly. The tendency for positive assortative matching on individual citation impact for (*A*) dual-authored papers and (*B*) dual-inventor patents. Matching tendencies between individuals are presented according to their solo outcomes, calculated based on each team member's solo works. For each given pairing of individuals, the plotted values are the amount by which the ratio of the observed matching frequency to the frequency expected by chance exceeds 1. The distribution of the mean trace (*m*) in the collaboration matrix when each field is analyzed separately for (*C*) papers and (*D*) patents. Consistent with $\hat{\rho} < 1$, we see a tendency toward positive assortative matching, which holds across all fields in both domains.

$y=\hat{\beta}_2\left[\frac{1}{2}(\hat{a}_{low}^{\hat{\rho}}+\hat{a}_{high}^{\hat{\rho}})\right]^{1/\hat{\rho}}$, where $\hat{\rho}$ and $\hat{\beta}_2$ are the model estimates for the relevant field and $\hat{a}_{low}$ and $\hat{a}_{high}$ are these individuals' model-estimated indices using all our data.

Fig. 2*A* shows raw data (blue line) and the model prediction (red line) for the WOS. Fig. 2*B* provides the same comparison for patents. In the figures, the vertical axis presents the moving average of R across all papers or patents with a given difference between the individual team members, $a_{high}-a_{low}$. We see that the model fits the raw data well. This visualization also reveals key intuition and implications. Namely, teams can have a large advantage over solo work, yet differences in individual impact indices within the team reduce this team advantage. Consider Fig. 2 *A* or *B* where the team members have the same index measure $(a_{high}=a_{low})$. Here the dual-authored output has a citation advantage substantially greater than what these individuals achieve alone. The raw data analog here corresponds directly to the model's estimate of β_2. However, as the gap between the individual impact indices widens, the impact advantage of dual-authored papers declines. This decline is consistent with $\rho<1$, so that the lower index team member dominates in determining the outcome. Had the team outcome been dominated by the higher index team member, then the raw data would slope upward in the figure (which would be consistent with $\rho>1$). Instead, as we see visually, heterogeneity in individual citation indices is impact-reducing. In fact, although the team advantage is sustained over fairly substantial differences in individual indices, once the differences in individual indices are large enough, teamwork is no longer more impactful as the organizational form. Overall, we see that the estimated team model (1) fits the shape of the raw data closely and that the impact advantages associated with teamwork are dissipated as the citation impact differences between team members grow.

Our next and related results consider team assembly. An organizational implication of $\rho<1$ is that heterogeneity of individual impact indices tends to reduce joint impact. From this perspective, research organizations would want to match people with similar indices (i.e., positive assortative matching) to maximize total research impact (15, 26, 27). Such sorting has implications for team assembly by individuals and institutions, with potentially wide implications across science and invention given the generality of $\rho<1$ (12, 28). Our next analyses therefore examine whether teams do indeed assemble to match on individual indices, consistent with our estimates of ρ.

Fig. 3 *A* and *B* focus on two-person teams. As for the raw data analysis in Fig. 2, we measure an individual's impact purely using their solo-authored work, producing an individual-level estimate that is independent of their coauthors. We then ask who works with whom. We present the ratio of (i) the observed frequency of two-person pairings to (ii) the frequency expected by chance,

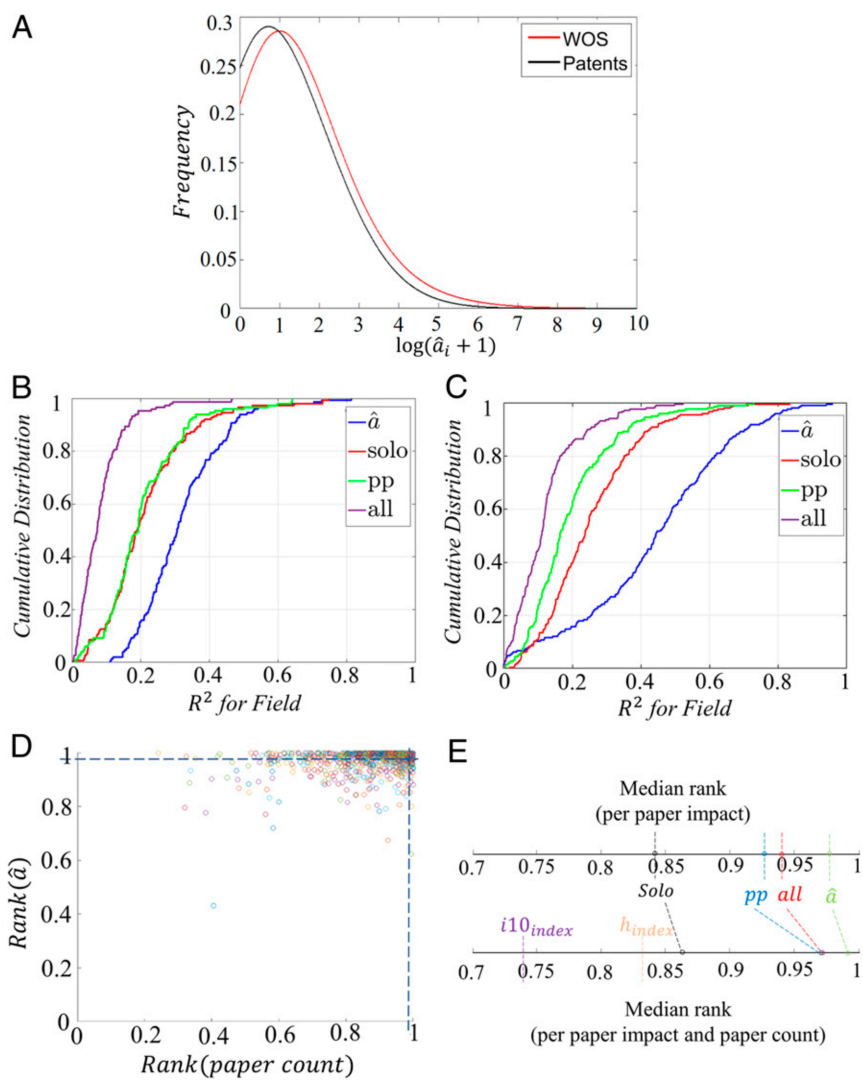


Fig. 4. Individual citation index. (*A*) The distributions of the individual citation index ($\hat{a}_i$) across Web of Science fields (red) and patenting fields (black). For paper or patent outcomes, the prediction of the citation impact for out-of-sample (*B*) solo-authored papers or (*C*) solo-invented patents. Predictive accuracy is measured in regressions, comparing the predictive capacity using $\hat{a}_i$ versus alternative measures (see text). The *x* axis presents the regression R^2 for a given field, and the *y* axis is the cumulative distribution across all fields. We see that $\hat{a}_i$ provides substantially more accurate predictions of out-of-sample citation outcomes compared with standard measures. For individual career outcomes, we rank each NAS member among that individual's corresponding cohort. (*D*) NAS members ranked by $\hat{a}_i$ (*y* axis) and publication count (*x* axis), with median ranks indicated by dashed lines. (*E*) Median rank for NAS members using alternative career metrics (see text). We see that $\hat{a}_i$ more accurately characterizes NAS members as high-rank individuals compared with standard career measures, including the h index.

drawing pairs of these individuals at random. We group individuals by mean citations to their solo work, rounded to the nearest integer. Fig. 3*A* shows a tendency toward assortative matching in the WOS, and Fig. 3*B* shows a similar tendency in patenting. Namely, collaborations are more frequent than expected by chance where $a_{high} = a_{low}$. Meanwhile, collaborations between individuals with different impact measures become increasingly unlikely as these differences become large.

We further deploy this analysis for each field separately within each domain. As a summary statistic, we examine the mean ratio of observed to expected frequencies where $a_{high} = a_{low}$ (i.e., we take the mean of the diagonal terms in matching matrices like Fig. 3 *A* and *B* but now analyzed by field). Fig. 3 *C* and *D* presents the distribution across fields for papers and patenting. In all fields, we see this mean ratio is greater than 1, so that positive assortative matching is a universal tendency. This tendency is consistent with the organizational implications of $\hat{\rho} < 1$. At the same time, teams may assemble this way for many reasons; for example, individuals with similar citation indices may sort into the same organizations or narrow subfields, which in turn facilitate their collaboration.

Our second group of results focuses on the individual citation index. The distribution of the individual index is right-skewed (Fig. 4*A*). These distributions are close to lognormal (*SI Appendix*, Fig. S3), which is consistent with citation distributions (29). The median individual citation index measure is $a_i = 1.32$ (papers) and $a_i = 1.05$ (patents), while the 95th percentile individual shows $a_i = 23.07$ (papers) and $a_i = 19.81$ (patents). Interestingly, we see a similar distributional shape in both the paper and patenting domains.

Notably, each individual citation index estimate has been determined accounting for the citation behavior of an individual's coauthors (and, more distantly, the citation behavior of everyone else in an individual's broader collaboration network). Moreover, these individual estimates are determined in light of the team-production parameters. An important implication of $\rho < 1$ is that the lower-ranked author is relatively important to the team-based outcome. Team-based outcomes will thus tend to be more informative about, and credit will accrue toward, the lower-index members of the team. By contrast, current popular methodologies for evaluating individuals (1) typically either are team blind (e.g., counting an individual's citations with no adjustment for team size, as in Google Scholar) or take a fractional approach (e.g., dividing citations by the number of coauthors), and promotion committees and funding panels are known to utilize such methods in evaluating individuals (30, 31) despite evidence that these may be poor predictors (32).

To examine the accuracy of the individual index estimates, $\hat{a}_i$, we consider their capacity to predict outcomes for out-of-sample papers and patents. Recall that $\hat{a}_i$ tells us the citation impact we expect for a paper or patent when the individual is a solo author or inventor. We run our estimations again for 100 WOS fields and 100 USPTO technology classes but leaving out, at random, one output from each individual. We then predict the outcome, y, for the paper or patent that was dropped. Further, we compare the predictive capacity of $\hat{a}_i$ against alternative, commonly used individual metrics (33), including (*i*) mean citations to the individual's works ("all," with no adjustment for the number of collaborators), (*ii*) mean citations per collaborator to the individual's works ("pp," with citations to each work are divided by its number of collaborators), and (*iii*) mean citations for the individual's solo works only ("solo"). A wide range of additional measures are analyzed in the *SI Appendix*, Tables S7 and S8. To measure prediction success, we run regressions by field, where the dependent variable is the citation impact of the out-of-sample work and the regressor is the predictive measure we are testing. We take the R^2 of each regression to capture goodness of fit. The *SI Appendix* provides further detail on methods.

Fig. 4*B* examines predictive success for out-of-sample solo-authored papers. Because these are solo-authored papers, the model prediction is $y_i = \hat{a}_i$, thus providing a focused test of the individual parameters. The figure presents the cumulative distribution of R^2 (across fields) for $\hat{a}_i$ and the common approaches *i–iii*. We see that the $\hat{a}_i$ estimates tend to provide substantially higher R^2 than the other metrics do in predicting out-of-sample outcomes. Notably, the model-estimated individual indices do better even than a simple average of the individuals' solo-authored works. The advantage of $\hat{a}_i$ comes because it is estimated using all of the individual's papers, which, although many involve team-authorship, help pin-down the measure. Fig. 4*C* shows that the estimates $\hat{a}_i$ similarly outperform the commonly used metrics when examining the patenting sphere. The *SI Appendix*, Table S8, shows that $\hat{a}_i$ similarly outperforms alternative metrics collected in (33), including numerous variants based on author order.

SI Appendix, Fig. S4, further considers out-of-sample prediction for works with two or three collaborators. Here the model prediction is based on the $\hat{a}_i$ for individuals in the team and the relevant $\hat{\beta}_n$ and $\hat{\rho}$ parameters for the field (estimated in samples where we have left out the papers or patents in the prediction set). The model prediction is then compared with predictions based on the popular constructs *i–iii* above. See *SI Appendix* for further discussion of methods. We again find large advantages of the model estimates in predicting out-of-sample outcomes, compared with these other measures. Overall, these findings suggest that our methodology, which can be applied in standard databases, can better predict outcomes both when individuals work alone and when they work in teams.

Our final results consider career outcomes. Here we consider an entire body of an individuals' work. Standard career metrics, such as the h index (34), incorporate paper impact measures and paper counts. In our context, the estimated $\hat{a}_i$ provides a per-paper impact measure for an individual, and we further incorporate publication volume, v_i, counting the papers the individual has joined in producing. As an outcome, we consider election to the National Academy of Sciences (NAS). We examine how NAS members rank among all other scholars in their cohort, defined as all individuals who share the same initial publication year and field (see *SI Appendix*, Tables S9 and S10, for data detail). Fig. 4*D* presents the ranks of $\hat{a}_i$ (vertical axis) and v_i (horizontal axis) for individuals elected to the NAS. NAS members rank at the 97th percentile of the $\hat{a}_i$ distribution and the 98th percentile of the v_i distribution, comparing against other scientists in their cohort.

How do these measures compare with standard career metrics? Prominent career metrics include (*i*) the h index (34), (*ii*) total citations received, and (*iii*) the i10 index, which counts an individual's papers with at least 10 citations. While these measures (all featured by Google Scholar) are team blind, other measures attempt to adjust for teamwork, including adjustments for the number of authors or author position (33). To assess these different approaches, we again rank NAS members against the other scientists in their field and cohort but now using these alternative metrics. Fig. 4*E* presents the median rank of individuals elected to the NAS for prominent alternatives. Additional comparisons are presented in the *SI Appendix*, Table S11. Using purely the per-paper impact measure (Fig. 4 *E*, *Top*) we see that ranking individuals based on $\hat{a}_i$ more accurately characterizes NAS members than alternative measures. Additionally, incorporating publication counts (Fig. 4 *E*, *Bottom*) further improves ranks. The $\hat{a}_i$-based rank continues to outperform. Notably, it proves far more accurate in characterizing NAS members than the h index. By contrast, total citations ("all") and equal sharing of citations per team member ("pp") do quite well (if not as well as using $\hat{a}_i$). This finding is consistent with the positive assortative matching we see above, where the tendency to work with teammates of similar individual citation indices can make equal credit per author systems relatively useful in ranking individuals.

Conclusion

We have presented a computationally feasible method for analyzing team and individual outcomes and deployed this methodology across large repositories of papers and patents. The analysis reveals universal patterns about team science and invention while providing a tool for estimating individual impact and predicting outcomes. The descriptive regularities suggest that team-based science and patenting most typically centers near the harmonic average of the team members' individual citation indices. These findings imply that team output is predicted more by the lower-index rather than the higher-index members of the team. This remarkable generality is further consistent with an observed tendency for team assembly among individuals with similar citation indices, which appears across all fields. Meanwhile, the individual index developed here is shown to outperform other metrics in predicting out-of-sample paper or patent outcomes and in characterizing eminent careers.

Further work can extend and refine this methodology and assess mechanisms. While our method, based on an individual fixed effect, is computationally feasible and can be deployed in available, wide-scale databases, in the context of richer data, extended methods might explore specific team assembly and production processes (4, 7, 10). Assessing choice in team assembly, sorting of ideas across teams, credit concerns, and effort allocation in idea production and marketing are important areas for future work. Causal research designs, including field and laboratory experiments, may allow close observation and isolation of specific mechanisms to help unpack the descriptive and predictive regularities unveiled here. In science fields that use author order (9, 35), one could further refine the methodology to study hierarchical roles (14), although our methodology already appears to outperform assessments that use author order (*SI Appendix*, Table S8). More generally, institutional features, such as the rise of postdoctoral positions and shifting funding landscapes, may interface with these findings, suggesting additionally important and policy-relevant avenues for future work. One may also extend this methodology by using alternative measures, beyond citation measures, to characterize outcomes, and by investigating teams in additional contexts. From entrepreneurship to songwriting, from surgery to sports, team assembly, team outcomes, and individual assessment are first-order concerns for the institutions that support teams and for the individuals themselves (13, 14, 36, 37).

Methods

The estimation produces two sets of parameters. First, we compute field-specific team-outcome parameters, $\hat{\rho}$ and $\hat{\beta}_2, \ldots, \hat{\beta}_n$. Second, we produce the individual index, $\hat{a}_i$, for every individual in the field, which can be hundreds of thousands of people. Because our outcome measure is the citations received by a given work, the estimate $\hat{a}_i$ is interpreted as an individual citation index. It represents the expected citation outcome for an output this person produces when working alone. Intuitively, the estimation of the individual citation index is possible because a person may sometimes work alone, providing a direct signal of his/her outcomes in that case, and/or because the same individual moves between different teams, allowing one to see how outcomes vary when a specific person is involved. In practice, for patents, we estimate the individual citation index for everyone in the technology class. For papers, very large fields in the WOS make estimation slow. In the largest 25 WOS fields, we therefore take, at random, a coauthor network within the field that contains between 50,000 and 100,000 unique authors. *SI Appendix*, Tables S1 and S2, presents the number of individuals analyzed for each field. Our estimation method is nonlinear least squares and should be interpreted as producing descriptive regularities and a tool for out-of-sample prediction, rather than isolating causative mechanisms. See *SI Appendix* for detailed discussion of methods; *SI Appendix* further describes the computational insights that make such a large-scale analysis feasible, demonstrates the successful convergence of the algorithm for widely different starting values in the parameter space, and demonstrates run times for collaboration networks of different size (*SI Appendix*, Tables S3–S5).

ACKNOWLEDGMENTS. We gratefully acknowledge support from the Alfred P. Sloan Foundation under award G-2015-14014. We thank Brian Uzzi, Pierre Azoulay, Tom Hubbard, Dan Barron, Niko Matouschek, Heidi Williams, and Scott Stern for helpful discussions. The Web of Science data are available via Thomson Reuters.

1. S. Wuchty, B. F. Jones, B. Uzzi, The increasing dominance of teams in production of knowledge. *Science* **316**, 1036–1039 (2007).
2. N. J. Cooke, M. L. Hilton, Eds.; National Academy of Sciences; National Research Council, *Enhancing the Effectiveness of Team Science* (National Academies Press, Washington, DC, 2015).
3. J. D. Adams, G. C. Black, J. R. Clemmons, P. E. Stephan, Scientific teams and institutional collaborations: Evidence from us universities, 1981–1999. *Res. Policy* **34**, 259–285 (2005).
4. B. F. Jones, The burden of knowledge and the 'death of the Renaissance man': Is innovation getting harder? *Rev. Econ. Stud.* **7**, 283–317 (2009).
5. R. K. Merton, The Matthew effect in science. *Science* **159**, 56–63 (1968).
6. National Academy of Sciences, National Academy of Engineering, and Institute of Medicine, *Facilitating Interdisciplinary Research* (The National Academies Press, Washington, DC, 2005).
7. S. M. Fiore, Interdisciplinarity as teamwork: How the science of teams can inform team science. *Small Group Res.* **39**, 251–277 (2008).
8. B. Uzzi, S. Mukherjee, M. Stringer, B. Jones, Atypical combinations and scientific impact. *Science* **342**, 468–472 (2013).
9. L. Allen, J. Scott, A. Brand, M. Hlava, M. Altman, Publishing: Credit where credit is due. *Nature* **508**, 312–313 (2014).
10. M. Bikard, F. Murray, J. Gans, Exploring trade-offs in the organization of scientific work: Collaboration and scientific reward. *Manage. Sci.* **61**, 1473–1495 (2015).
11. S. Fortunato *et al*., Science of science. *Science* **359**, eaao0185 (2018).
12. G. Grossman, G. Maggi, Diversity and trade. *Am. Econ. Rev.* **90**, 1255–1275 (2000).
13. B. Hamilton, J. Nickerson, H. Owan, Team incentives and worker heterogeneity: An empirical analysis of the impact of teams on productivity and participation. *J. Polit. Econ.* **111**, 465–497 (2003).
14. L. Garicano, T. Hubbard, Learning about the nature of production from equilibrium assignment patterns. *J. Econ. Behav. Organ.* **84**, 136–153 (2012).
15. M. Kremer, The o-ring theory of economic development. *Q. J. Econ.* **108**, 551–575 (1993).
16. P. Legros, A. F. Newman, Monotone matching in perfect and imperfect worlds. *Rev. Econ. Stud.* **69**, 925–942 (2002).
17. G. Z. Jin, B. Jones, S. F. Lu, B. Uzzi, The reverse Matthew effect: Consequences of retraction in scientific teams. *Rev. Econ. Stat.*, 10.1162/rest_a_00780 (2018).
18. J. Gans, F. Murray, "Credit history: The changing nature of scientific credit" in *The Changing Frontier: Rethinking Science and Innovation Policy*, A. Jaffe, B. Jones, Eds. (University of Chicago Press, 2015), pp. 107–131.
19. M. Levin, S. Krawczyk, S. Bethard, D. Jurafsky, Citation-based bootstrapping for large-scale author disambiguation. *J. Am. Soc. Inf. Sci. Technol.* **63**, 1030–1047 (2012).
20. L. Bai, Data unification and disambiguation. Presentation to "Web of Science as a Research Dataset" conference, Indian University, November 14–15 2016. https://cns.iu.edu//docs/research/workshops/161114/161114-Bai.pdf. 2016. Accessed 7 June 2019.
21. Clarivate Analytics. Web of Science product webpage. https://clarivate.com/products/web-of-science/. Accessed 7 June 2019.
22. G.-C. Li *et al*., Disambiguation and co-authorship networks of the U.S. patent inventor database (1975–2010). *Res. Policy* **43**, 941–955 (2014).
23. R. Lai, A. D'Amour, A. Yu, Y. Sun, L. Fleming, Disambiguation and Co-authorship Networks of the U.S. Patent Inventor Database (1975 - 2010). https://doi.org/10.7910/DVN/5F1RRI. Accessed 1 February 2017.
24. M. Ahmadpoor, B. F. Jones, Data sets for team and individual impact in science and invention. Figshare. https://doi.org/10.6084/m9.figshare.8242571. Deposited 7 June 2019.
25. B. Jones, *As Science Evolves, How Can Science Policy?* (NBER Innovation Policy and the Economy, 2010), vol. 11.
26. G. S. Becker, A theory of marriage: Part I. *J. Polit. Econ.* **81**, 813–846 (1973).
27. D. Topkis, *Supermodularity and Complementarity* (Princeton University Press, Princeton, NJ, 2011).
28. B. F. Jones, S. Wuchty, B. Uzzi, Multi-university research teams: Shifting impact, geography, and stratification in science. *Science* **322**, 1259–1262 (2008).
29. F. Radicchi, S. Fortunato, C. Castellano, Universality of citation distributions: Toward an objective measure of scientific impact. *Proc. Natl. Acad. Sci. U.S.A.* **105**, 17268–17272 (2008).
30. G. Ellison, How does the market use citation data? The Hirsch index in economics. *Am. Econ. J. Appl. Econ.* **5**, 63–90 (2013).
31. M. Perry, P. Reny, How to count citations if you must. *Am. Econ. Rev.* **106**, 2722–2741 (2016).
32. P. Jensen, J.-B. Rouquier, Y. Croissant, Testing bibliometric indicators by their prediction of scientists promotions. *Scientometrics* **78**, 467–479 (2009).
33. L. Waltman, A review of the literature on citation impact indicators. *J. Informetrics* **10**, 365–391 (2016).
34. J. E. Hirsch, An index to quantify an individual's scientific research output. *Proc. Natl. Acad. Sci. U.S.A.* **102**, 16569–16572 (2005).
35. T. Tscharntke, M. E. Hochberg, T. A. Rand, V. H. Resh, J. Krauss, Author sequence and credit for contributions in multiauthored publications. *PLoS Biol.* **5**, e18 (2007).
36. B. Uzzi, J. Spiro, Collaboration and creativity: The small world problem. *Am. J. Sociol.* **111**, 447–504 (2005).
37. J. Duch, J. S. Waitzman, L. A. Amaral, Quantifying the performance of individual players in a team activity. *PLoS One* **5**, e10937 (2010).

Supporting Information for

Decoding Team and Individual Impact in Science and Invention

Mohammad Ahmadpoor
Northwestern University and NICO

Benjamin F. Jones
Northwestern University, NICO, and NBER

April 29, 2019

1 Data

We study (1) journal article data from the Web of Science (WOS) and (2) patent data from the United States Patent and Trademark Office (USPTO), while also incorporating (3) data identifying members of the National Academy of Sciences.

1.1 Paper Data

The paper data contain 24 million publications, constituting all research articles indexed in the name disambiguated version of Thomson Reuters WOS database that were published over the 1945-2005 period. The WOS records paper titles, bibliographic information (journal, volume, issue, page), citations, author information (names, affiliations), and citation links to other papers in the database. Each document in our analysis is a research article as defined by WOS (as opposed to other WOS document categories such as letters, notes, editorial material, discussions, and meeting abstracts). The WOS data are available to researchers through Clarivate Analytics and described in detail at www.webofknowledge.com. The WOS database further provides name-disambiguated identifiers for individual authors using their Distinct Author Identification System (DAIS), which combines a machine learning approach that has high precision and recall (1) together with validated researcher identification sources like ORCID and ResearcherID, as well as user feedback (2).

We analyzed 240 fields of research as codified by the WOS. These fields include all those in sciences, engineering, social sciences, and arts and humanities where there at least 500 papers in the field. While our WOS dataset covers all research articles published up to 2013, we use citations received within the first eight years after publication (3) to measure the impact of a research article and hence study papers published up to 2005. See Table S1 for the number of papers and number of unique authors in each WOS field. Note that, for the 20 largest Web of Science fields, we used large subsamples rather than the entire field to assist with computational speed. Specifically, we drew, at random, one initial author in each of these fields and then built a coauthorship network outwards from that author until there were between 50,000 and 100,000 papers in the sample for that field.

1.2 Patent Data

The patent data contain all 3.9 million patents granted by USPTO with application dates between 1975 and 2006. These data integrate three different data sources: (i) the Patent Data Project of the National Bureau of Economic Research (https://sites.google.com/site/patentdataproject/Home); (ii) the updated patent data of (4) (https://iu.app.box.com/v/patents); and (iii) the name-disambiguated dataset (5), which provides identifiers for distinct inventors using a machine-learning approach. Together, these data record the patent number, application year, unique inventor id, number of citations that each patent received, and technological class of each patent.

We studied all 384 technological classes determined by the USPTO that have at least 500 patents in the class. We use citations received within the first eight years after patent application (3) to measure the impact of a patent and hence study patent applications up until 2010. See Table S2 for the number of patents and number of unique inventors in each technology class.

1.3 National Academy of Sciences Data

For the NAS members information, we extracted each NAS member's name, affiliation, and their field of research (primary and secondary field) from the NAS member search website (http://www.nasonline.org/member-directory/?page=form). The list consists of 2,757 names who are alive at the time of our study. For 10 NAS members there is no affiliation listed (therefore they are excluded from our study) and for 42 percent of members there is no secondary field listed.

1.4 Data Availability

The underlying journal article data, described in Section 1.1 above, are available from Clarivate Analytics, but restrictions apply to the availability of these data, according to institutional licenses, so are not publicly available. Data are however available from the authors upon reasonable request and with permission of Clarivate Analytics.

The underlying patent data sets are publicly available as described in Section 1.2. Integrated patent data that support the findings of this study are available from the authors upon request.

The list of NAS members and their publications are available from the authors by email or by download from the corresponding author's website.

2 Methods

2.1 The Generalized Mean Function

We consider an outcome metric, y, for team-produced output and model its expected value using the generalized mean function:

$$y = \beta_n(\frac{1}{n}\sum_{i=1}^{n} a_i^{\rho})^{\frac{1}{\rho}}, \tag{1}$$

where i indexes individual members of a team of size n and a_i is an individual index for person i that represents the outcome when this person works alone. The term ρ is the generalized mean parameter, which defines how the parameters a_i are averaged together, and the parameter β_n captures the advantage of teams of size n in producing high impact research. Recall that we use the normalization $\beta_1 = 1$, which implies that $y = a_i$ for solo-authored work and thus individual index is measured on the same scale as the outcome metric.

2.2 The Regression Model

Given a sample of team-produced outputs, including information regarding who worked with whom and the outcome of each collaboration, we can estimate the unknown parameters. We model a given outcome, y_k, as having the expected value given in (1) plus a stochastic error term. We can then formulate the optimization problem using non-linear least squares regression and solve:

$$\min_{\rho,\{\beta_n\},\{a\}} \sum_{k=1}^{K} \left(\beta_n(\frac{1}{n_k}\sum_{j=1}^{n_k} a_j^{\rho})^{\frac{1}{\rho}} - y_k \right)^2, \tag{2}$$

where k indexes specific team-produced outputs and there are K research outputs in the sample.

A regression sample is constituted by the patents in a given USPTO technological class or the journal articles in a given WOS field. We thus estimate field-specific values of ρ, $\{\beta_n\}_{n=2}^{T}$, and $\{a_i\}_{i=1}^{M}$, where M is the number of unique authors in that field and T is the maximum team size in the data.

The estimation operates through variation in authorship structure. In particular, a person may sometimes work alone, providing a direct signal of his/her outcome index, a_i, and may also move between different teams, allowing one to see how output varies when a specific person is involved. Tracing an individual across different settings has been used to study the role of CEOs (6), and it has been used to study paper outcomes for a sample of MIT faculty (7). A distinction with these other methods is that we consider the individual fixed effect in the context of the generalized mean function, allowing for a richer array of mappings between individual and group outcomes.

More broadly, while individual fixed effects are useful, the analysis is ultimately descriptive and predictive, rather than causative. For example, researcher team assembly, like the matching of CEOs to firms, reflect choices. These choices may in turn influence the outcome when a set of individuals work together. The descriptive regularities that emerge from the analysis may thus follow from various underlying team processes and actions. Experimental approaches, where team membership and other team features are varied exogenously, are important areas for future work that can help isolate underlying mechanisms.

Complementary to our approach, modeling and estimating team assembly choices can also be revealing. The study (7) considers a framework where an individual scientist weighs the potential gains from collaboration against the limited credit the individual may receive when the output is jointly produced. In a sample of approximately 650 MIT scientists, (7) finds that team-authored outputs have higher impact than solo-authored outputs and estimate a credit-sharing rule in their sample that can make team assembly choices rational. They also find evidence, although it is more ambiguous, that collaborations between junior and senior scientists may result in lower impact, which might be interpreted as consistent with $\rho < 1$. However, by contrast with our approach, (7) considers output in an additive regression framework (as with $\rho = 1$), leading to a question of how their findings would look when an index of solo-authored outcomes combines in a non-linear fashion and at the much lower values of ρ estimated in this paper. More generally, modeling approaches like those in (7) allow one to leverage formal considerations of choices to reveal structural parameters of interest, providing additional avenues forward in understanding team-based outcomes.

2.3 Computational Algorithm

To solve the optimization problem formulated in Eq. (2), we use the gradient descent method. To do so, we need to calculate the derivative of the objective function, (2), with respect to all its relevant parameters: ρ, $\{\beta_n\}_{n=2}^{T}$, and

$\{a_i\}_{i=1}^{M}$ in a given field.

To improve computational efficiency, and since large team sizes are rare, we consider outputs with 8 or less collaborators only, which account for 97% of papers and 99% of patents. Further, we collect rare, larger teams into a single β parameter, estimating $\{\beta_2, \beta_3, \beta_4\}$ for teams of size 2, 3, 4, and letting β_5 account for teams of size 5 through 8.

2.3.1 Gradients

The algorithm considers the first derivatives of $F_k = \left(\beta_{n_k}(\frac{1}{n_k}\sum_{i=1}^{n_k} a_i^{\rho})^{\frac{1}{\rho}} - y_k\right)^2$ for any given output k and for each relevant parameter. Using gradient descent, we search for the parameters that minimize the sum of F_k across all outputs in the sample.

For an individual who is part of the team that produced output k, the derivative with respect to that individual's a_i is:

$$\frac{\partial F_k}{\partial a_i} = \frac{2}{n_k}\beta_{n_k} {a_i}^{\rho-1}\left(\frac{1}{n_k}\sum_{j=1}^{n_k} a_j^{\rho}\right)^{\frac{1}{\rho}-1}\left(\beta_{n_k}\left(\frac{1}{n_k}\sum_{j=1}^{n_k} a_j^{\rho}\right)^{\frac{1}{\rho}} - y_k\right) \tag{3}$$

The first derivative with respect to β_{n_k} is:

$$\frac{\partial F_k}{\partial \beta_{n_k}} = 2\left(\beta_{n_k}(\frac{1}{n_k}\sum_{j=1}^{n_k} a_j^{\rho})^{\frac{1}{\rho}} - y_k\right)(\frac{1}{n_k}\sum_{j=1}^{n_k} a_j^{\rho})^{\frac{1}{\rho}}). \tag{4}$$

And the first derivative with respect to ρ is:

$$\frac{\partial F_k}{\partial \rho} = 2\left(\beta_{n_k}(\frac{1}{n_k}\sum_{j=1}^{n_k} a_j^{\rho})^{\frac{1}{\rho}} - y_k\right)\times$$
$$\frac{\beta_{n_k}}{\rho}(\frac{1}{n_k}\sum_{j=1}^{n_k} a_j^{\rho})^{\frac{1}{\rho}}\cdot\left[\frac{\sum_{j=1}^{n_k} a_j^{\rho} ln(a_j)}{\sum_{j=1}^{n_k} a_j^{\rho}} - \frac{1}{\rho}(\frac{1}{n_k}\sum_{j=1}^{n_k} a_j^{\rho})\right] \tag{5}$$

For the optimization problem, (2), the relevant gradient for updating each parameter is then the sum of that parameter's derivatives across the set of outputs $k = 1, ..., K$.

2.3.2 Algorithm for Individual Citation Index

A primary computational challenge is the large number of individuals, M, in a field, where each individual has his/her own impact index value, a_i, and collaborates with others across a complex network structure and with varying team sizes. While some WOS fields have a relatively small number of individuals, the largest fields have hundreds of thousands of different authors. The scale of these networks can thus require estimation of a very large number of individual parameters in each iteration, and requires keeping track of large but sparse matrix of specific collaborations.

Key computational insights regards how one organizes the collaboration matrices and how one updates these individual index parameters. In particular, rather than confronting $M \times M$ collaboration matrices and looping over each individual author, which can require problematic numbers of separate executions of the code for each round of the gradient descent, we instead simultaneously update the vector of author index values. This vector approach, which requires storing information in a particular way, speeds up the algorithm by many orders of magnitudes for large fields.

In particular, we proceed by building collaboration matrices with a dense information structure. Taking a given field, we first sort the ensemble of authors by assigning an integer from 1 to M to each individual and then building a matrix in which the first column is the a_i of each author. Then we build each row (from the second column to the last column) to list the individual identifiers, a number 1 to M, that indicates a specific coauthors of the person in the first column.

For example, for the dual-authored papers, the collaboration matrix has the following structure:

$$A_2 = \begin{bmatrix} a_1 & p_1^1 & p_2^1 & \cdots & \cdots & \cdots & \cdots \\ a_2 & p_1^2 & p_2^2 & \cdots & \cdots & \cdots & \cdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ a_M & p_1^M & p_2^M & \cdots & \cdots & \cdots & \cdots \end{bmatrix} \tag{6}$$

where, e.g., the p_j^1 in the first row are the individual identifiers (a number 1...M) for author 1's coauthor in author 1's j^{th} dual-authored paper. Define the largest number of dual-authored papers from any given author as Q. This matrix has the nice feature that it is comparatively small size: Rather than using an $M \times M$ matrix to define collaborators, here the matrix A_2 is an $M \times (Q+1)$ matrix, where $Q << M$. Algorithmically, this set-up allows us to calculate (3) by looping across the relatively small number of columns, rather than down the potentially very large number of rows.

To calculate the gradient for updating each author's index parameter, via (3), we also need another matrix containing the outcome, y, for each dual-authored paper. Specifically, we build a similar matrix to (6) in which the first column is removed and the i^{th} row lists each y of the dual-authored papers written by author i, producing an $M \times Q$ matrix.

$$Y_2 = \begin{bmatrix} y_1^1 & y_2^1 & \cdots & \cdots & \cdots & \cdots & \\ y_1^2 & y_2^2 & \cdots & \cdots & \cdots & \cdots & \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ y_1^M & y_2^M & \cdots & \cdots & \cdots & \cdots & \end{bmatrix} \tag{7}$$

Now, note that $\vec{a} = A_2(:,1)$ is the current vector of individual parameter estimates. For the jth dual-authored outputs, the vector of outcomes is $\vec{y_j} = Y_2(:,j)$ and the coauthor parameters for these outputs are $\vec{a_j} = A_2(A_2(:,j+1),1)$. Thus we can calculate Eq. (3) in vector form. We then iterate across the columns of these matrices, i.e., summing across all the dual-authored outputs in which an individual is involved.

2.3.3 Extending the Approach to General Team Sizes

In order to build a similar matrix for team-authored papers with larger numbers of coauthors, the first column of that matrix is again the a_i's of individuals (as in the matrix A_2). Then each row contains the individual identifier for each coauthor of the person in the first column, organized by the specific paper or patent. The coauthor identities, p, are again defined by the integer $(1...M)$ that gives the co-author position in the first column.

$$
A_n = \begin{bmatrix}
a_1 & p_{11}^1 & p_{12}^1 & \cdots & p_{1n}^1 & p_{21}^1 \cdots & \cdots & \cdots \\
a_2 & p_{11}^2 & p_{12}^2 & \cdots & p_{1n}^2 & p_{21}^2 \cdots & \cdots & \cdots \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \\
a_M & p_{11}^M & p_{12}^M & \cdots & p_{1n}^M & p_{21}^M \cdots & \cdots & \cdots
\end{bmatrix} \tag{8}
$$

Following the same procedure for dual-authored papers, the derivative of the objective function with respect to the individual index parameters for n-authored papers can be calculated by looping over the columns of matrix A_n.

2.3.4 Computational Efficiency

We used two computational resources for this analyses: (1) the Kellogg Data Center which contains 120 CPUs, 2TB of RAM, and (2) the Kellogg Linux Cluster which is a set of five Linux servers, each having 28 CPU cores and 1.5TB RAM. Table S3 lists computational run-times of the algorithm for several collaboration networks from the Web of Science, along with their corresponding average team size.

2.3.5 Examples of Parameter Convergence

To examine the convergence of the algorithm, we first turned to an ensemble of 1,178 mathematics papers from a community of 657 authors. We started with the following initial conditions: $\rho = 1$, $\{a_i\}_{i=1}^{657} = 1$, $\beta_2 = 1$, $\beta_3 = 1$, $\beta_4 = 1$, and $\beta_5 = 1$. To demonstrate the ability of the algorithm to reach convergence regardless of the initial conditions, we ran the algorithm for the aforementioned example of mathematicians from alternative sets of initial conditions. We chose initial conditions that are distant from the ultimate estimated values: (1) $\rho = 3$, $\beta_2 = \beta_3 = \beta_4 = \beta_5 = .1$, and $\{a_i\} = 1$ (2) $\rho = -3$, $\beta_2 = .1$, $\beta_3 = .2$, $\beta_4 = .3$, $\beta_5 = .4$ and $\{a_i\} = 1$. Using of these initial conditions, we can examine the convergence history for the parameters and see that they converge to extremely similar estimates. Visual representations of this convergence is available from the authors upon request.

More generally, we considered 20 additional fields at random (see Table S4). For each field, we ran the algorithm staring from four different sets of initial conditions, as listed in Table S4A. Tables S4B-C show that the final estimates for ρ and β_N are similar regardless of the parameter starting points.

2.4 Alternative Outcome Measures

For the output measure, the main text follows (3) and emphasizes the count of citations within the first eight years after publication (papers) and application (patents). Our methodology can also be applied to alternative output measures, and robustness to other outcome measures may be useful to refine interpretations. For example, different fields can have different citation distributions, and differences between solo and team-authored citation impacts may in part reflect field differences across author configurations.

In our analysis, we isolate hundreds of subfields of the Web of Science and similarly isolate hundreds of technology classes of the USPTO, but one can also go further using alternative outcome measures. Here we consider two other metrics for y. First, we consider the log citation count, specifically taking the natural logarithm of the eight-year citation count (and adding 1 to the citations so that the logarithm is well defined for works that receive no citations). This logarithmic measures acts to reduce the role of upper-tail citation outliers in influencing the results. Second, we consider a binary measure, where a paper is consider high impact if it is among the top 20 percent of citations received, with the upper 20th percentile being defined by field and year. This approach forces each sub-field to have exactly the same outcome distribution. The data sample is, as above, 20 fields.

As can be seen from Table S5, the estimated ρ with these alternative outcome measures is below 1, indicating relative weighting toward the lower-index members of the team. The estimated βs are greater than 1, also indicating the impact advantage when researchers are working in a team. The findings regarding team outcomes and organization thus appear broadly robust to alternative outcome measures. At the same time, it is possible that solo and team-authored work reflect variations across higher-resolution subfields within the field categories we analyze, where different subfields have different citation patterns.

2.5 Individual Index Estimation

Figure 4A presents the distribution of a_i for all paper authors and, separately, for all inventors. We plot $log(a_i + 1)$ on the x-axis given the right-skewed distributions of a_i. The figure suggests the lognormal nature of these distributions, as further shown in Fig. S3A-B, which is consistent with citation distributions (8). The median individual impact measure is a_i=1.32 (papers) and a_i=1.05

(patents), while the 95th percentile individual shows a_i=23.07 (papers) and a_i=19.81 (patents). Interestingly, we see a similar distributional shape in both the paper and patenting domains.

2.6 Visual Fit of Model

Figure 2 considers the visual fit of the model compared to raw data using the construct R (see main text). The raw data version calculates a_i using the mean citations to an individual's solo authored work and takes the citations to the dual-authored paper as the outcome. By necessity, this analysis requires each individual to have at least 1 solo-authored paper in addition to the dual-authored paper. The modeled version of R uses the same sample as the raw data but now takes the model estimated values $\hat{a}_i$ and for the outcome calculates the generalized mean (1) given $\hat{a}_{high}$ and $\hat{a}_{low}$ and the appropriate $\hat{\rho}$ and $\hat{\beta}$ of the field.

In Fig. 2, the x-axis presents $x = a_{high} - a_{low}$ using the two individuals' solo-work averages, bucketed by integer values of the difference. The y-axis presents R (for the raw data and model versions separately). The figure presents the moving average of R at each x value, with uniform weighting of observations over the $[x-2, x+2]$ interval.

2.7 Matching Analyses

The analysis of matching in Figs. 3A-D considers the actual frequency of collaboration compared to what is expected by chance, given authors' index measures. As in Fig. 2, we first calculate a_i using the mean citations to an individual's solo authored work. We then calculate the discrete frequency distribution of individuals according to the nearest integer value of a_i. In a sample with Z individuals, this frequency distribution is

$$f(a) = \frac{1}{Z}\sum_{i=1}^{Z} 1(a_i = a) \tag{9}$$

Now let there be J observed pairings, where a given pairing has individuals with index measures (a_{j1}, a_{j2}). The observed frequency of individual index pairings is then counted as

$$g_{obs}(a_1, a_2) = \frac{1}{J}\sum_{j=1}^{J} 1(a_{j1} = a_1, a_{j2} = a_2) \tag{10}$$

The null model is then developed as follows. Given the distribution $f(a)$, the probability under random matching that a dual-authored work occurs between two people with index measures a_1 and a_2 is

$$g_{null}(a_1, a_2) = f(a_1)f(a_2) \tag{11}$$

We can then compare the observed versus expected pairing outcomes through the ratio

$$V(a_1, a_2) = \frac{g_{obs}(a_1, a_2)}{g_{null}(a_1, a_2)} \tag{12}$$

Figs. 3A-B plot $V - 1$ for all $a_1, a_2 \in \{1, ..., 25\}$. In these figures, $V - 1 > 0$ indicates that the observed pairing happens more often than expected by chance and $V - 1 < 0$ indicates that the observed pairing happens less often than expected by chance. Note that we do not triangularize these matrices.[1]

For Fig. 3C-D, we separately calculate the observed and expected distributions and $V(a_1, a_2)$ for each field. As a summary statistic for assortative matching, we then examine the trace of $V(a_1, a_2)$. Specifically, for each field, we take $\bar{V}$ as the arithmetic mean of $V(a, a)$. Consistent with Figs. 3A-B, we consider $a \in \{1, ..., 25\}$. This produces one observation per field, where a mean greater than 1 indicates a tendency toward positive assortative matching. Fig. 3C-D present the distribution of $\bar{V}$ across fields.

2.8 Out-of-Sample Prediction Regressions

To examine the accuracy of the individual index estimates, a_i, we consider their capacity to predict outcomes for out-of-sample papers and patents. Recall that a_i tells us the citation impact we expect for a paper or patent when the individual is a solo author or inventor. We run our estimations again for 100 WOS fields and 100 USPTO technology classes but leaving out, at random, one output from each individual. We then predict the outcome, y, for the paper or patent that was dropped.

[1] That is, since $g_{null}(a_1, a_2)$ is symmetric, we could alternatively enforce the order $a_1 < a_2$ and collect the observed off-diagonal terms, with the null model now being $g_{null}(a_1, a_2) = 2f(a_1)f(a_2)$ where $a_1 \neq a_2$.

Using the test sample, we run ordinary-least-squares regressions of the form:

$$y_i = \alpha_0 + \alpha_1 P_i + \epsilon_i \tag{13}$$

for each field, where y_i is a left-out output, P_i is a prediction metric, and ϵ_i is an i.i.d. error term. Note that, by construction, there is one observation in this regression for each individual in the field.

We consider four variants of the prediction metric. Recall that these are all calculated in the estimation sample (i.e., they do not use the outcome in the test sample, which we are trying to predict).

- Solo. Here we define P_i as the arithmetic mean citation impact of the individual's solo-authored work.
- All. Here we define P_i as the arithmetic mean citation impact of all outputs associated with that individual, regardless of team size and with no adjustment for team size.
- PP. Here we define P_i as a per-person (PP) average, taking the arithmetic mean across all outputs associated with that individual but now dividing citations for a given output by n, the team size.
- $\hat{a}_i$. Using the estimation samples, we re-run our computations for each field, producing new estimates of $\{a_i\}_{i=1}^{M}$, as well as ρ and $\{\beta_n\}_{n=1}^{T}$, for that field. The estimated parameters are then taken to calculate P_i using the generalized mean function.

The above measures are featured in the main text. However, we further consider a broad arrange of additional metrics for individuals, as reviewed in (9). These measures, which include numerous metrics based on author order, are defined in Table S7.

2.8.1 Data

We take 100 medium-sized fields in the WOS and 100 medium-sized technology classes of the USPTO. For each field, we then randomly take a single paper (patent) for each author (inventor) to create (i) a test sample, which is constituted purely from these left-out outputs, and (ii) an estimation sample, which contains all the other outputs in that field (i.e., all the papers or patents except those in the test sample). In practice, for each WOS field or technology class, we create three versions of these test and estimation samples depending on the team size we are studying:

- T1. The test sample leaves out 1 solo-authored work for each individual.
- T2. The test sample leaves out 1 dual-authored work for each individual.
- T3. The test sample leaves out 1 three-authored work for each individual.

Naturally, an individual must have at least two works to be considered, so that one work can be left out and we can still construct a prediction metric from the estimation sample. Moreover, because one of our alternative prediction metrics depends purely on an individual's solo-authored work, we restrict each field dataset to individuals who have at least two solo-authored works.

Note that for solo-authored work (the T1 test sample), the generalized mean function simply gives $P_i = \hat{a}_i$. For team-based work (the T2 and T3 test samples), the generalized mean function predictor further incorporates the appropriate estimates $\hat{\rho}$ and $\hat{\beta}_n$ for the field.

For the other predictors (Solo, All, and PP as defined above and featured in the main text, as well as the additional measures defined in Table S7) we take the individual-level measure for the solo-authored prediction regressions (the T1 test sample). For team-based work (the T2 and T3 test samples), we take an arithmetic mean of the individual measures in the specific team to calculate P_i.

2.8.2 Out-of-Sample Prediction Results

The regression (13) is then run separately for each field, for each prediction metric, and for each test sample (T1, T2, T3). For each regression, we record the R^2 to capture goodness-of-fit. Figs. 4B-C in the main text focus on predicting solo-authored outcomes (T1 sample). Fig. S4B presents the cumulative frequency distribution of these R^2 for paper fields, comparing the predictive success for each of the four prediction metrics ($\hat{a}_i$, Solo, All, PP). Fig. S4C considers the same but for patenting fields.

Following a similar format, Figs. S4A-B examine the T2 samples (dual-authored works) for both papers and patents. Figs. S4C-D examine the T3 samples (three-authored works) for both papers and patents.

For the many additional metrics defined in Table S7, we further summarize their predictive success in Table S8. We provide the mean and median R^2 for paper fields, using each of the metrics. For comparisons, we also present the mean and the mean and median R^2 using $\hat{a}_i$ as well the metrics features in the main text.

2.9 National Academies of Sciences Analysis

2.9.1 Matching NAS Members to WOS Data

To find the corresponding WOS Author ID for each NAS member we matched their last name, first initial, affiliation, and their field of research with our WOS database. Since primary and secondary fields listed in the NAS membership are broader/coarser than WOS fields, we created a crosswalk from NAS primary and secondary fields to multiple WOS fields. Table S9 summarizes this crosswalk. For each NAS member we followed the following matching procedure. First, we looked for the set of WOS Author IDs with the same last name and first initial. Second, we trimmed this set by looking for authors who have more than one third of their publications in the primary and secondary fields listed for that individual in the NAS membership database and have at least 10 publications. Finally, we take the Author ID(s) where one of the WOS affiliations matches the NAS member's affiliation in the NAS database. Our analysis considers those NAS members for whom we find a unique WOS Author ID and for whom the corresponding $\hat{a}$ is also available in our estimations, which represents 45% of all NAS members. Table S10 summarizes the results of this matching algorithm.

2.9.2 Cohort Comparisons

After finding the corresponding WOS Author ID for NAS members, we can compare them to broader cohorts of scientists based on different indexes (e.g., h-index, i10-index, pp-avg, all-avg, solo-avg, and $\hat{a}$). To define a relevant cohort for each individual NAS member, we take all WOS authors who (i) published their majority of work in the same WOS field as the NAS member and (ii) share the same first publication year with the NAS member. The mean cohort size for an NAS member includes 1,967 individuals, and the median cohort size 1,430 individuals. We rank each NAS member within their corresponding cohort set.

In Fig. 4E, we present two different rankings for NAS members among their cohort. First we ranked them based only on their per-paper impact measures (e.g., pp-avg, all-avg, solo-avg, and $\hat{a}$). Fig.4E (top line) shows the results of this ranking. Second, We further incorporated the number of papers published by each NAS members in ranking NAS members. To this end, for each NAS member, we take the rank of the NAS member and members of his/her corresponding cohort based on per-paper impact measures (q_i^r), we also rank them based on their number of papers in the same cohort (v_i^r). Finally, we rank the NAS member based on $q_i^r \times v_i^r$. Fig. 4E (bottom line) shows the results of this per paper impact and paper count ranking for q_i^r defined by our main measures (pp-avg, all-avg, solo-avg, and $\hat{a}$).

We further show in Fig. 4E (bottom line) rankings based on well-known career indices, h_{index} and $i10_{index}$. The $i10_{index}$ is the number of the individual's publications with at least 10 citations each. The h_{index} is the largest integer h such that the individual has published at least h papers each of which has been cited in other papers at least h times.

Finally, we further ranked NAS members based on a broad range of additional indexes defined on a per-paper basis (See Table S7 for a list of these indexes). For each NAS member, we calculated their rank in their corresponding cohort based on these per-paper impact measures, as well as on the per-paper impact measure and paper count. We followed the same ranking procedure as described above. Table S11 summarizes the results of these rankings.

2.10 Code Availability

The computational algorithm and other analyses were implemented in Matlab, drawing on a SQL database. The one exception is the out-of-sample prediction regressions, which were performed in Stata. All code is available from the authors upon request.

References

1. Levin, Michael, Stefan Krawczyk, Steven Bethard, and Dan Jurafsky. 2012. Citation-based bootstrapping for large-scale author disambiguation. Journal of the American Society for Information Science and Technology 63(5):1030–1047.

2. Bai, Linge, 2016. Data unification and disambiguation. Presentation to "Web of Science as a Research Dataset" conference, Indian University, November 14-15 2016. http://cns.iu.edu/docs/research/workshops/161114/161114-Bai.pdf

3. Wuchty, S., B.F. Jones, and B. Uzzi. 2007. The increasing dominance of teams in production of knowledge. Science 316, no. 5827: 1036–1039.

4. Kogan, Leonid, Dimitris Papanikolaou, Amit Seru, and Noah Stoffman. 2017. Technological innovation, resource allocation, and growth," The Quarterly Journal of Economics 132(2): 665-712.

5. Li, Guan-Cheng, R. Lai, A. D'Amour, D. Doolin, Y. Sun, V. Torvik, A. Yu, and L. Fleming. 2014. Disambiguation and co-authorship networks of the u.s. patent inventor database (1975 - 2010). Research Policy 43(6), 941-955.

6. Bertrand, M. and A. Schoar. 2003. Managing with style: the effect of managers on firm policies. Quarterly Journal of Economics, 143 (4): 1169-1208.

7. Bikard, M., F. Murray, and J. Gans 2015. Exploring trade-offs in the organization of scientific work: collaboration and scientific reward. Management Science 61 (7): 1473-1495.

8. Radicchi, Filippo, Santo Fortunato and Claudio Castellano. 2008. Universality of citation distributions: Toward an objective measure of scientific impact. PNAS 105 (45) 17268-17272.

9. Waltman, Ludo. 2016. A review of the literature on citation impact indicators. Journal of Informetrics 10, 365–391.

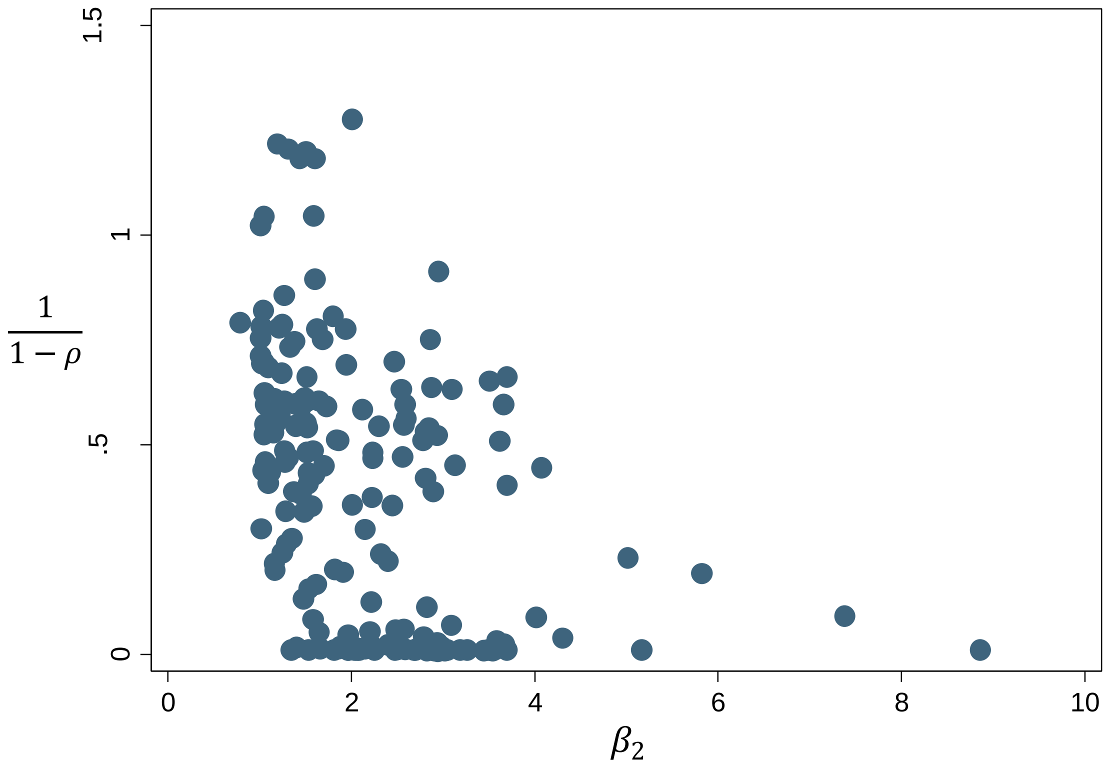


Fig. S1. Relationship between team production parameters for papers. There is evidence that $\hat{\beta}_2$ and $\hat{\rho}$ are negatively correlated, but with substantial residual variation. Table S6 provides regression results exploring the correlation.

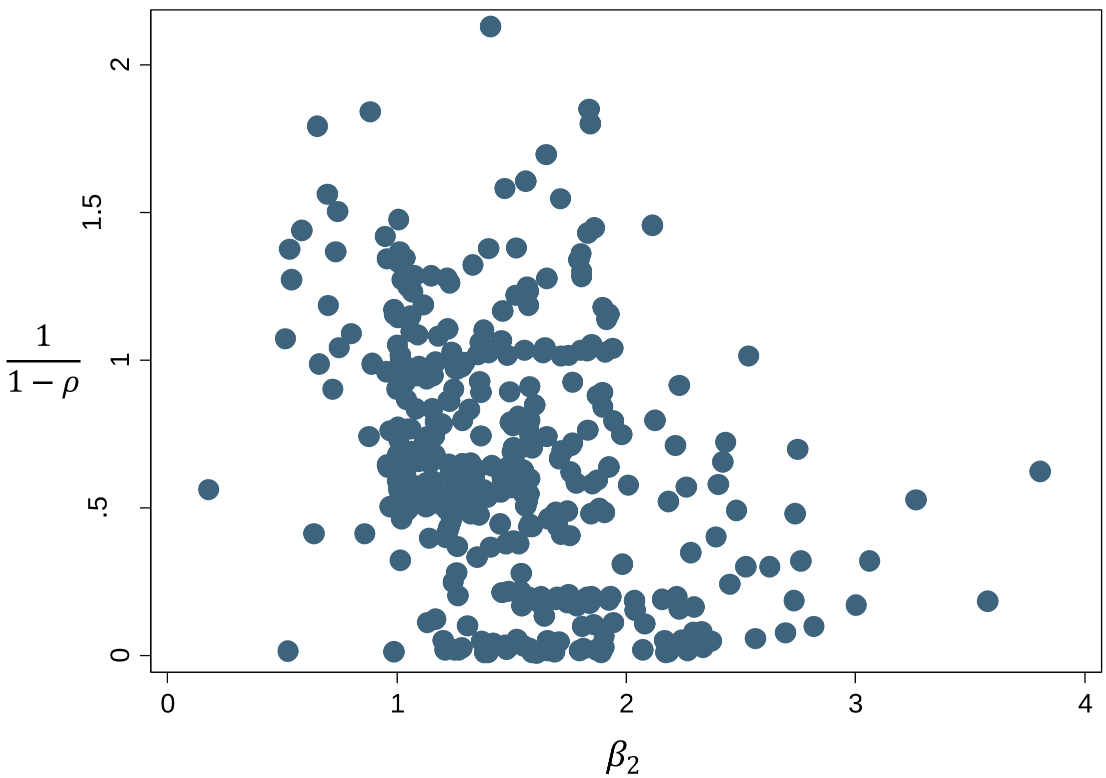
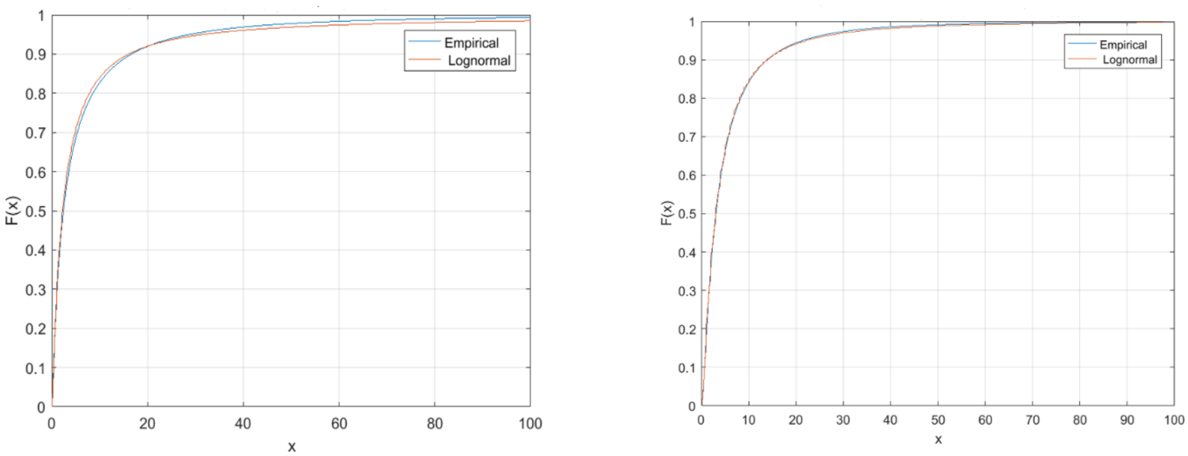


Fig. S2. Relationship between team production parameters for patents. There is evidence that $\hat{\beta}_2$ and $\hat{\rho}$ are negatively correlated, but with substantial residual variation. Table S6 provides regression results exploring the correlation.



(A) (B)

Fig. S3. Individual productivity. The distributions of the individual productivity parameters ($\hat{a}_i$) across Web of Science fields (**A**) and patenting fields (**B**). The empirical cumulative density function is shown for $\hat{a}_i > 1$ (blue) together with fitted log-normal cumulative density function (red) in each panel.

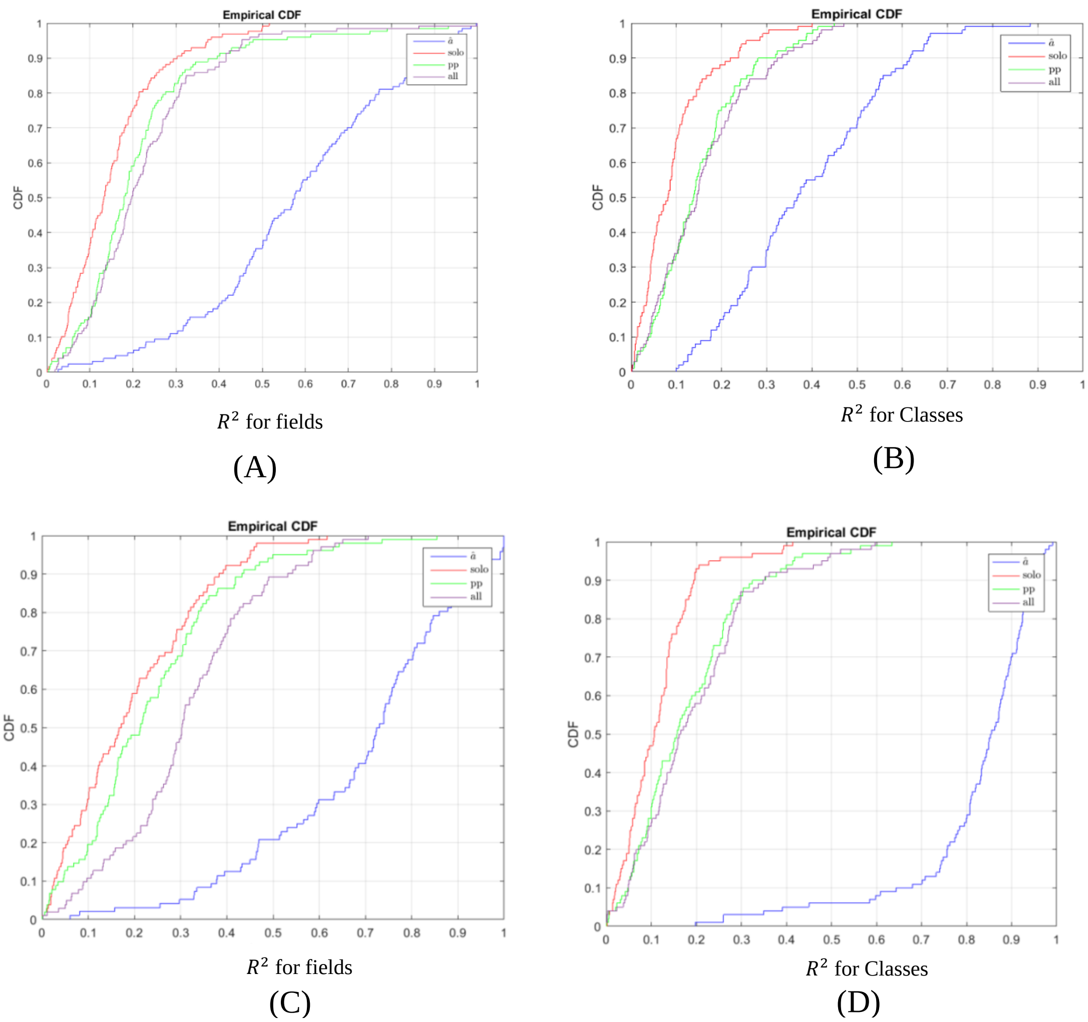


Fig. S4. Out-of-sample predictions. Predictions for two-author papers (**A**), two-inventor patents (**B**), three-author papers (**C**), and three-inventor patents (**D**). We see large advantages of the model estimates in predicting out-of-sample outcomes compared to the other measures. Model predictions are based on the $\hat{a}_i$ for individuals in the team and the relevant $\hat{\beta}_n$ and $\hat{\rho}$ parameters for the field (estimated in samples where the we have left out the papers or patents in the prediction set). See SI text for detailed methods.

Table S1. Parameter Estimates by WOS Field.

Field Name	Paper Count	Author Count	ρ	β_2	β_3	β_4	β_5
ACOUSTICS	46219	41171	-0.63791	1.493372	1.938176	1.497213	1.369275
AUTOMATION & CONTROL SYSTEMS	15081	18528	-89.8321	3.574877	6.221267	9.777709	7.935242
AGRICULTURE, DAIRY & ANIMAL SCIENCE	81552	89360	-0.60771	1.0521	1.013926	1.081317	1.208377
AGRICULTURAL ENGINEERING	5218	8762	-1.62641	1.445881	2.753613	2.654927	3.764589
AGRICULTURAL ECONOMICS & POLICY	1763	2610	-0.28957	1.935274	2.15423	1.125864	3.295714
AGRICULTURE, MULTIDISCIPLINARY	17914	38464	-0.97065	3.621679	4.377908	3.225043	2.999245
AEROSPACE ENGINEERING & TECHNOLOGY	69166	71139	-52.1045	2.467049	3.746375	5.169489	6.733012
AGRICULTURAL EXPERIMENT STATION REPORTS	10716	13212	0.153888	1.600808	1.296755	1.816645	2.489018
ALLERGY	10959	19088	0.169956	1.312328	1.385236	1.568199	1.517174
ANATOMY & MORPHOLOGY	32743	42617	-52.3464	2.225173	3.147256	4.379505	3.88886
ANDROLOGY	7949	16433	-5.44096	1.537023	3.00265	4.929398	5.783273
ANESTHESIOLOGY	66753	104173	-111.667	3.450145	9.775834	9.85756	13.8958
BIODIVERSITY CONSERVATION	582	1259	-0.43321	2.469624	2.818322	5.70108	4.951838
ASTRONOMY & ASTROPHYSICS	27711	25879	-1.31421	1.533542	1.996623	2.421541	2.535465
PSYCHOLOGY, BIOLOGICAL	238109	110464	-1.58831	2.893714	4.615971	4.060856	4.183848
BEHAVIORAL SCIENCES	5574	11448	-1.4576	1.093925	1.572939	1.979235	3.624471
BIOCHEMICAL RESEARCH METHODS	7649	26122	-1.30355	1.1121	1.229243	1.308001	2.712256
BIOCHEMISTRY & MOLECULAR BIOLOGY (*)	506903	639419	-0.09576	2.954008	4.336389	4.047762	3.094111
BIOLOGY	106129	155476	-0.75062	1.220331	1.285951	1.493752	1.630004
BIOLOGY, MISCELLANEOUS	18800	25119	0.153376	1.439536	1.625051	1.483013	1.104654
BIOPHYSICS (*)	251171	384708	-1.06294	1.276597	1.652232	1.843318	2.047217
BIOTECHNOLOGY & APPLIED MICROBIOLOGY	73059	149556	-0.60464	1.057775	1.12324	1.231146	1.373899
PLANT SCIENCES	317483	296630	-26.1647	4.310178	7.59658	7.396153	5.013696
ONCOLOGY (*)	268205	480904	-0.46099	1.091445	1.715886	1.872022	1.667831
CARDIAC & CARDIOVASCULAR SYSTEMS	158025	266453	-101.652	3.555779	5.339207	5.476935	5.431346
CELL BIOLOGY (*)	259648	463733	-0.83648	2.29605	2.594003	3.524602	2.851882
CRITICAL CARE MEDICINE	11201	32676	-0.67656	3.661254	2.049003	2.402336	5.950759
THERMODYNAMICS	224	560	-16.2543	2.490921	3.3441	3.347841	3.677349
CHEMISTRY, APPLIED	34724	53593	0.023201	1.010594	1.0824	1.020761	1.053366
CHEMISTRY, CLINICAL & MEDICINAL	28082	88517	-0.71112	1.133065	1.125599	1.299041	1.583159
CHEMISTRY (*)	570413	627204	-0.83187	1.391222	2.071487	2.059216	1.83452
CHEMISTRY, ANALYTICAL (*)	256807	346298	-1.21661	3.129685	3.695538	4.326129	5.801274
CHEMISTRY, INORGANIC & NUCLEAR	160816	180081	-0.96343	2.781692	1.675107	2.66423	3.387871
CHEMISTRY, ORGANIC (*)	430069	454216	-13.5873	3.095284	8.007852	4.290727	3.019204
CHEMISTRY, PHYSICAL (*)	375838	415571	-0.69174	1.724373	2.52813	2.506055	2.603641
COMPUTER SCIENCE, ARTIFICIAL INTELLIGENCE	49595	69091	-101.678	8.865485	14.06572	14.02042	17.25533
COMPUTER SCIENCE, CYBERNETICS	7771	10092	-0.65629	1.648285	1.482982	2.521744	1.844606
COMPUTER SCIENCE, HARDWARE & ARCHITECTURE	9933	16828	-0.81334	1.511928	2.6482	2.096534	2.813544
COMPUTER SCIENCE, INFORMATION SYSTEMS	25071	37120	-1.12232	2.558623	4.016282	3.989967	3.143734

COMMUNICATION	22335	20944	-86.6834	2.851679	3.89169	5.222586	5.313739
COMPUTER SCIENCE, INTERDISCIPLINARY APPLICATIONS	16514	31578	-0.34264	1.378029	1.095223	1.257015	1.120185
COMPUTER SCIENCE, SOFTWARE ENGINEERING	65011	84046	-0.96226	1.863387	2.105735	2.256298	3.253631
COMPUTER SCIENCE, THEORY & METHODS	141745	174833	-0.67198	1.394943	1.4221	1.359803	1.451823
COMPUTER APPLICATIONS & CYBERNETICS	20606	20196	-0.27871	1.012799	2.3451	3.2196	4.432331
CONSTRUCTION & BUILDING TECHNOLOGY	6582	8419	0.21645	2.010935	1.235803	1.168498	1.837289
CRITICAL CARE	33024	70286	-0.28468	1.214319	1.310406	1.51784	1.2619
CRYSTALLOGRAPHY	144016	203754	-104.473	2.68195	5.134793	8.397342	9.479953
CYTOLOGY & HISTOLOGY	16840	28268	-2.80887	1.297725	2.913943	3.653931	4.861115
DENTISTRY, ORAL SURGERY & MEDICINE	107474	122993	-0.82971	2.572759	2.63091	2.380465	1.140758
DERMATOLOGY	105494	147986	-0.44127	1.02342	1.05427	1.061446	1.35338
GEOCHEMISTRY & GEOPHYSICS	106161	111075	-103.679	5.169603	8.023401	13.80391	13.76922
SUBSTANCE ABUSE	23998	41758	-61.708	2.044579	3.193409	4.293499	5.264149
HEALTH CARE SCIENCES & SERVICES	10549	26735	-0.24005	1.806062	2.011922	1.863885	1.938528
ELECTROCHEMISTRY	65005	80753	-101.883	1.962651	4.280817	6.942339	9.586182
EVOLUTIONARY BIOLOGY	13614	26991	-97.7749	1.962728	3.22855	4.120825	5.877534
DEVELOPMENTAL BIOLOGY	46813	81476	-1.80035	2.015423	2.84344	3.74355	4.19703
ENDOCRINOLOGY & METABOLISM (*)	223030	353840	-7.07481	2.217918	3.06772	3.5957	3.4588
ENERGY & FUELS	38843	65493	-0.81549	1.173603	1.784844	1.981919	2.166043
ENGINEERING	61652	74782	-21.3232	1.966776	3.235285	4.490401	4.121911
ENGINEERING, BIOMEDICAL	36790	78513	-1.13048	1.313425	1.680647	2.189769	2.264813
ENGINEERING, ENVIRONMENTAL	5792	10254	-1.93539	1.29025	2.367601	2.861718	3.86809
ENGINEERING, CHEMICAL	251266	293895	-6.59734	1.476822	2.562346	3.337268	4.221702
ENGINEERING, INDUSTRIAL	13174	16616	-1.48192	3.702246	4.293428	4.674577	5.213447
ENGINEERING, MANUFACTURING	14746	21376	-0.33336	1.688578	2.169566	2.53509	2.321862
ENGINEERING, MARINE	4276	4540	-1.24933	4.076491	5.192444	3.665302	3.159417
ENGINEERING, CIVIL	60048	70186	-16.0894	2.569574	3.191825	3.7461	3.479811
ENGINEERING, OCEAN	1970	3457	-4.13915	1.91679	4.846438	5.199722	4.29098
ENGINEERING, PETROLEUM	46860	66609	-0.85134	2.8542	4.410236	3.671214	4.55771
ENGINEERING, ELECTRICAL & ELECTRONIC (*)	313697	360623	-17.7462	2.199189	3.169605	3.609759	3.969994
ENGINEERING, MECHANICAL	134602	167237	-0.83592	1.40146	1.806103	2.126207	1.977048
ENGINEERING, GEOLOGICAL	1916	3873	-40.7468	3.675097	6.877459	5.551427	6.814168
ENTOMOLOGY	128414	106423	-0.58442	3.102023	3.437352	3.33601	4.882095
ENVIRONMENTAL SCIENCES	159381	260559	-0.32896	1.010906	1.023951	1.022337	1.010309
ENVIRONMENTAL STUDIES	18116	21716	-0.1694	1.269239	1.101157	1.056135	1.675431
ERGONOMICS	8718	12834	-1.38791	2.809661	3.29163	4.682956	5.078241
FOOD SCIENCE & TECHNOLOGY	244371	292114	-36.6396	2.937385	4.560781	9.575752	3.079918
GASTROENTEROLOGY & HEPATOLOGY	136039	265328	-0.57295	2.878564	1.920291	2.841251	2.654128
GENETICS & HEREDITY (*)	245084	420347	-9.94017	7.388191	5.09104	3.874768	3.270564
GEOGRAPHY	39520	31985	-79.9383	2.506544	3.414743	4.968912	5.678845
GEOGRAPHY, PHYSICAL	3300	7468	-7.92628	2.820216	6.075744	6.145648	8.135737
GEOLOGY	35343	47643	-0.36769	1.329444	1.630075	1.716371	1.641418
GEOSCIENCES, INTERDISCIPLINARY	178142	203008	-106.546	3.702687	6.102979	9.819542	4.440829

GERIATRICS & GERONTOLOGY	38400	66684	-0.53777	3.504515	2.83629	1.631092	2.001465
GERONTOLOGY	9379	22116	-0.85629	1.524469	2.167452	2.75779	3.7378
HEALTH POLICY & SERVICES	26291	36760	-0.26411	0.787621	1.606379	1.517944	1.367428
HEMATOLOGY	118171	244064	-1.27818	1.034736	1.317651	2.060012	2.711946
MATHEMATICAL & COMPUTATIONAL BIOLOGY	4879	13379	-0.69625	1.44601	2.303042	1.428128	2.782249
HOSPITALITY, LEISURE, SPORT & TOURISM	319	612	-10.3155	4.014139	5.206053	3.476568	11.33113
PUBLIC HEALTH	165301	286232	-56.9635	1.973961	3.095042	4.368136	5.14929
IMMUNOLOGY	208111	322397	-81.7039	2.861622	5.250574	6.950381	9.750015
INFECTIOUS DISEASES	89560	203454	-0.67828	1.064515	1.381112	1.182873	1.474199
PSYCHOLOGY, APPLIED	46541	51560	-1.13608	2.234217	2.855539	3.850233	4.085334
NANOSCIENCE & NANOTECHNOLOGY	4397	14874	-3.62394	1.16082	2.237341	6.915534	1.955042
INFORMATION SCIENCE & LIBRARY SCIENCE	72266	63748	-11.2392	1.580817	2.660071	5.544014	5.72226
INSTRUMENTS & INSTRUMENTATION	115492	181878	-24.086	2.787618	4.829886	5.451392	7.188086
INTEGRATIVE & COMPLEMENTARY MEDICINE	3095	7881	-0.88364	2.805727	3.902754	2.486076	1.441379
MEDICAL ETHICS	439	1158	-4.21285	5.821887	7.687987	3.16517	3.085881
MEDICINE, LEGAL	12585	22593	-1.22212	1.697705	2.961414	2.76883	3.042602
LIMNOLOGY	3353	6520	-0.5103	3.69976	6.876978	3.840584	3.032577
LANGUAGE & LINGUISTICS	40824	30762	-31.3856	3.593858	5.850051	5.278308	6.607353
MANAGEMENT	50791	51592	0.178006	1.193978	1.185465	1.260554	1.083691
OPERATIONS RESEARCH & MANAGEMENT SCIENCE	76199	77403	-72.0302	2.155313	3.589418	5.24298	5.823889
MARINE & FRESHWATER BIOLOGY	127204	128773	-1.33599	1.593253	2.068659	2.30573	2.95343
MATERIALS SCIENCE, PAPER & WOOD	49188	46542	-108.376	3.020347	7.078623	9.403677	12.71609
MATERIALS SCIENCE, CERAMICS	69998	103693	-103.581	2.694172	5.560018	9.779149	9.835949
MATERIALS SCIENCE (*)	297343	441467	-0.67828	2.582538	3.512234	2.680163	2.518452
MATHEMATICS, APPLIED	161954	124049	-1.58451	1.368474	2.024366	2.771963	2.758212
MATHEMATICS, INTERDISCIPLINARY APPLICATIONS	15223	21040	-0.62897	1.069779	1.046987	1.258735	1.942032
MATHEMATICS	401466	151573	-101.649	2.061685	3.622335	4.802691	13.5362
SOCIAL SCIENCES, MATHEMATICAL METHODS	19398	20026	-3.51324	2.407264	5.34752	7.115991	7.283256
MEDICAL INFORMATICS	22485	48931	-0.51072	1.516367	2.504986	2.020617	2.601164
MECHANICS	161656	155521	-106.392	2.480578	4.923809	7.077821	6.984633
MEDICAL LABORATORY TECHNOLOGY	44032	102486	-0.66835	1.270161	1.6162	1.670253	2.17006
MEDICINE, GENERAL & INTERNAL (*)	634825	974625	-0.3341	2.863796	4.798553	3.518837	3.386252
METALLURGY & METALLURGICAL ENGINEERING	191967	260749	-101.742	2.081333	4.301263	8.446742	9.788057
MEDICINE, RESEARCH & EXPERIMENTAL (*)	205952	427646	-0.89542	1.149311	1.47091	1.851279	1.826935
MEDICINE, MISCELLANEOUS	8228	15350	-0.65544	1.273981	1.316807	1.353552	1.492202
MATERIALS SCIENCE, BIOMATERIALS	24558	53968	-0.43429	1.051001	2.393394	2.37866	2.385923
MATERIALS SCIENCE, CHARACTERIZATION & TESTING	38794	64702	0.166056	1.506171	1.392013	1.411742	1.498346
MATERIALS SCIENCE, COATINGS & FILMS	31710	56665	-104.508	2.038246	6.460435	9.697189	9.837055
MATERIALS SCIENCE, COMPOSITES	29450	50401	-0.21977	1.040203	2.36301	3.069476	3.048475
MATERIALS SCIENCE, TEXTILES	21402	26525	-94.8217	1.808954	2.689396	4.161946	5.830735
METALLURGY & MINING	61338	76644	-71.2969	2.09057	3.748177	4.849846	5.668076
METEOROLOGY & ATMOSPHERIC SCIENCES	119858	114722	-54.1493	1.881166	2.976587	3.641023	5.052802
MICROBIOLOGY (*)	261592	386844	-1.82718	1.565206	2.369825	2.711489	3.044738

MICROSCOPY	23463	43372	-0.95796	1.838093	2.207996	2.305621	1.839219
ROBOTICS	7908	14132	-87.2703	2.582649	4.582073	5.945097	7.181401
MINERALOGY	36354	36452	-101.65	1.542231	9.698319	6.722937	9.715099
MULTIDISCIPLINARY SCIENCES (*)	560795	817884	-53.9814	2.988387	3.847402	4.849886	9.394644
MYCOLOGY	29294	40550	-101.702	3.185076	6.576779	9.741299	14.08856
CLINICAL NEUROLOGY	94764	182902	-2.37273	2.150223	3.507206	4.041721	4.49843
NEUROSCIENCES (*)	367152	448300	-82.8752	2.701349	4.870093	6.271197	6.198934
NEUROIMAGING	1981	7038	-0.49119	1.236003	1.025677	1.03491	1.005523
NUCLEAR SCIENCE & TECHNOLOGY	119440	180278	-1.46646	1.53292	2.097975	2.334852	1.460812
NURSING	50398	69481	-0.27217	1.245188	1.396644	1.346325	1.46581
NUTRITION & DIETETICS	123318	201206	-0.58503	2.543186	2.923508	3.015739	3.762348
OBSTETRICS & GYNECOLOGY	146901	224216	-107.335	3.541546	9.241255	9.853125	13.79945
OCEANOGRAPHY	104654	114464	-109.428	2.829357	4.623968	8.637312	9.420201
REMOTE SENSING	8005	14085	-0.65029	1.475673	1.967162	2.158326	1.912324
OPHTHALMOLOGY	151008	166728	0.04379	1.586508	1.443012	1.743931	1.391209
OPTICS	174763	216048	-1.81373	2.446406	2.942854	2.988174	3.626803
ORTHOPEDICS	52198	88077	-2.62707	1.348083	1.976917	2.801114	3.386252
OTORHINOLARYNGOLOGY	78407	99340	-101.802	3.043512	6.77136	13.86179	13.87816
PARASITOLOGY	43507	56824	-74.5088	1.663378	2.638978	4.105554	4.926787
PATHOLOGY	161026	305497	-0.91265	1.052377	1.844887	2.11129	2.227165
PEDIATRICS (*)	201509	359821	-0.44897	1.945706	1.120142	1.013723	1.00541
PHARMACOLOGY & PHARMACY (*)	550309	825696	-3.1744	2.319399	3.654992	4.53888	4.971218
PHYSICS, APPLIED (*)	537307	630238	-97.0888	3.491701	6.557386	9.215971	6.084662
IMAGING SCIENCE & PHOTOGRAPHIC TECHNOLOGY	25049	36471	-153.622	2.944666	5.705571	9.866234	9.859701
PHYSICS, FLUIDS & PLASMAS	69521	72757	-0.92101	2.932152	2.118677	2.608255	2.731721
PHYSICS, ATOMIC, MOLECULAR & CHEMICAL	259231	229550	-60.144	1.40619	2.443456	3.178074	4.213507
PHYSICS (*)	557186	537062	-1.06275	1.580501	1.894491	2.118604	2.108821
PHYSICS, CONDENSED MATTER (*)	471424	456949	-3.99176	1.170461	1.901275	3.163874	3.381769
PHYSIOLOGY	235903	318082	-18.5029	1.65241	2.846007	3.610836	1.75699
PHYSICS, NUCLEAR	113070	141033	-5.0331	1.617981	2.415257	3.482565	4.553606
PHYSICS, PARTICLES & FIELDS	135223	129140	-101.378	2.257413	3.19878	5.321884	6.853666
PHYSICS, MATHEMATICAL	140205	124309	-44.6428	2.406473	4.438553	4.500399	8.126398
POLYMER SCIENCE	286550	300039	-0.77949	2.601121	2.545949	3.076675	3.375546
PSYCHOLOGY, MATHEMATICAL	15054	16220	-98.4653	2.964803	4.291518	6.435142	7.65669
RADIOLOGY & NUCLEAR MEDICINE (*)	294224	441907	-1.0803	1.514946	2.234817	2.04931	2.002945
RESPIRATORY SYSTEM	98649	190809	-3.12515	1.243395	2.054548	3.045224	3.605678
REPRODUCTIVE BIOLOGY	60054	102756	0.042524	1.048668	1.211628	1.221105	2.115809
RHEUMATOLOGY	58774	108186	-3.36201	5.021176	5.867865	7.030696	3.920862
SOCIAL SCIENCES, BIOMEDICAL	24596	40009	-1.07787	2.233682	3.443884	3.441012	1.882034
AGRICULTURE, SOIL SCIENCE	84041	89637	-0.2882	1.625668	1.66857	1.579224	1.453116
SPECTROSCOPY	161849	267898	-2.35357	1.016618	1.663252	1.598723	2.769249
SPORT SCIENCES	81203	128907	-102.624	3.265562	6.010354	7.824635	6.519372
STATISTICS & PROBABILITY	124351	90882	-3.95461	1.819751	3.061514	6.951111	6.033513
SURGERY (*)	513427	730294	-0.1173	1.604692	3.385807	3.025853	3.017754

TELECOMMUNICATIONS	142286	175662	-0.64046	1.163391	1.823244	1.87981	1.847082
TOXICOLOGY	161114	268963	-110.665	2.903586	6.575	9.788359	9.755397
TRANSPLANTATION	76017	162838	-0.40605	1.008517	1.014932	1.005497	1.059571
TRANSPORTATION	14086	18064	-0.65943	1.51927	1.93136	1.682693	1.722488
TRANSPORTATION SCIENCE & TECHNOLOGY	16333	27825	-1.68482	2.229477	3.27586	4.057626	3.867854
TROPICAL MEDICINE	46884	90580	-61.4714	2.194866	3.519515	5.036937	5.742102
UROLOGY & NEPHROLOGY	173411	276198	-101.824	1.344791	2.788448	4.732565	6.368334
VETERINARY SCIENCES	290205	329564	-1.17844	1.277963	1.903698	2.318916	2.872017
PERIPHERAL VASCULAR DISEASE	166912	321039	-1.18122	1.065837	1.674582	1.60451	2.870398
VIROLOGY	106596	181497	-0.8237	1.054476	1.350384	1.828472	2.109886
WELDING TECHNOLOGY	731	970	-0.71149	2.128149	1.810453	1.940154	1.31893
MINING & MINERAL PROCESSING	42977	59286	-76.9837	1.853841	2.712801	3.597141	4.741283
WATER RESOURCES	126854	170368	-1.94337	1.484653	2.123702	2.74445	1.883943

Table S2. Parameter Estimates by USPTO Technology Class.

Class	Patent Count	Inventor Count	ρ	β_2	β_3	β_4	β_5
Abrading	7759	7702	-0.63368	1.236918	1.824313	3.29396	2.890066
Abrasive tool making process, material, or composition	11799	13416	-54.7664	1.251522	1.598658	1.341272	1.568552
Acoustics	1533	2103	-0.19928	1.155329	1.33583	1.592249	1.627349
Active solid-state devices (e.g., transistors, solid-state diodes)	3831	5014	-4.06663	2.221702	1.749012	3.418743	4.161153
Adhesive bonding and miscellaneous chemical manufacture	47090	44348	-47.6256	1.48111	3.2331	1.818421	1.488163
Advancing material of indeterminate length	23695	34418	-0.07987	1.359777	1.657051	0.635945	0.974664
Aeronautics and astronautics	1681	2338	0.294249	0.951224	1.375352	1.315511	2.265253
Agitating	9075	10967	-0.35227	1.65265	2.405514	2.376994	3.192983
Alloys or metallic compositions	5413	7069	-0.60799	3.804806	3.145336	5.337454	2.219632
Ammunition and explosives	2800	5089	-0.12374	1.365291	1.229887	1.371477	1.165385
Amplifiers	5324	6330	-8.76693	1.861596	3.335205	4.791252	4.665764
Amusement devices	8654	8574	-0.59592	1.549401	0.980617	3.809878	2.518983
Amusement devices: games	7929	8241	-142.013	1.608597	1.482251	1.623082	2.133018
Amusement devices: games	1134	1288	-24.0694	1.421812	1.620854	4.098825	1.074822
Amusement devices: toys	4096	4388	0.014178	1.015087	1.291573	1.318855	2.750065
Animal husbandry	6251	5780	-0.0106	0.894581	1.045318	1.36843	1.396785
Apparel	7939	8448	-0.42907	1.58851	1.480865	1.947323	2.871392
Apparel apparatus	9213	9485	-0.78604	1.380776	2.416262	2.700716	2.149333
Article dispensing	1418	1406	-0.08367	1.768463	1.2651	1.637136	2.164341
Automatic temperature and humidity regulation	3387	4373	-9.21724	1.306834	3.60431	2.004164	0.767263
Baths, closets, sinks, and spittoons	2346	2877	0.308811	1.86286	1.815633	1.997117	2.287569
Batteries: thermoelectric and photoelectric	7092	7589	-4.398	2.036227	0.310633	4.589291	1.563893
Bearings	2813	3651	0.24973	1.006025	1.007166	1.183597	2.190758
Beds	7099	7887	-20.1668	1.369139	2.287073	3.604323	5.810538
Binder device releasably engaging aperture or notch of sheet	7888	7866	-0.94769	1.281715	1.477571	1.940571	4.044317
Bleaching and dyeing; fluid treatment and chemical modification of textiles and fibers	839	799	-0.45462	1.502827	1.429779	2.312601	3.481053
Bookbinding: process and apparatus	519	613	-4.09857	1.699645	2.135816	3.917212	6.00996
Books, strips, and leaves	6478	7895	-25.8533	1.898675	3.194364	4.999255	6.183186
Boot and shoe making	507	525	-3.63844	1.538129	2.518052	5.365485	8.088689
Boots, shoes, and leggings	998	1083	0.213543	0.539623	1.266242	1.355039	2.741
Boring or penetrating the earth	4884	4133	-0.37283	1.008835	1.224782	1.307842	1.185325
Bottles and jars	5899	5696	0.206524	1.233454	1.66601	1.57284	1.992722

Brakes	4218	3923	-5.40069	2.231272	3.137836	1.661913	3.41293
Bridges	7809	8394	-0.83753	1.389642	1.229038	1.614866	2.563043
Brushing, scrubbing, and general cleaning	939	987	0.049495	1.849199	1.88912	2.308262	2.726023
Buckles, buttons, clasps, etc.	11264	12037	-0.02327	1.284649	0.764362	0.906765	1.542127
Buoys, rafts, and aquatic devices	6874	6835	-0.74685	1.127579	1.707595	1.303998	2.03069
Butchering	2110	2260	-1.40493	1.22048	1.875117	3.471693	1.699148
Card, picture, or sign exhibiting	2624	2554	-0.04745	1.003648	1.479275	1.38129	0.409822
Catalyst, solid sorbent, or support therefor: product or process of making	5965	6424	-1.4975	2.39408	3.152861	2.258621	4.512941
Chain, staple, and horseshoe making	11936	14844	-0.78981	0.179131	0.162204	0.59612	1.354772
Chairs and seats	662	682	0.220497	1.807797	2.20294	2.204603	3.059556
Check-actuated control mechanisms	11000	11292	-0.04684	1.130677	2.86602	0.883223	2.206866
Chemical apparatus and process disinfecting, deodorizing, preserving, or sterilizing	1615	1753	0.024936	1.911585	1.855422	1.69661	1.605961
Chemistry of hydrocarbon compounds	14110	22939	-43.0184	1.815564	2.665054	3.332778	3.1783
Chemistry of inorganic compounds	6669	7193	-0.81609	1.451959	1.719384	1.263619	1.146585
Chemistry: analytical and immunological testing	16145	23586	-35.8661	2.337746	1.694313	2.249723	0.515299
Chemistry: electrical and wave energy	8529	14016	0.076707	1.092108	1.246681	1.329046	2.464237
Chemistry: electrical current producing apparatus, product, and process	14213	20749	0.271643	0.530524	0.8877	0.864059	0.709811
Chemistry: fertilizers	13468	14908	-31.5009	1.55343	0.946759	1.311415	2.298057
Chemistry: fischer-tropsch processes; or purification or recovery of products thereof	1322	1864	-0.25735	1.582417	2.241695	3.702456	0.824247
Chemistry: molecular biology and microbiology	909	1178	0.054811	1.363454	1.20915	3.351857	1.877779
Chemistry: natural resins or derivatives; peptides or proteins; lignins or reaction products thereof	49270	65155	-1.66526	1.478073	2.754201	0.356372	3.330928
Chucks or sockets	10117	17316	-0.25766	2.125672	2.216566	2.4081	0.852771
Classifying, separating, and assorting solids	1357	1407	-0.19863	1.085304	1.396587	1.328319	2.921136
Cleaning and liquid contact with solids	5857	7813	-8.1379	1.943031	2.994628	2.088029	3.648099
Cleaning compositions for solid surfaces, auxiliary compositions therefor, or processes of preparing the compositions	8209	12299	0.530167	1.407039	1.557716	1.728097	3.348594
Closure fasteners	9514	10261	-0.34974	1.367351	2.482471	3.181402	3.074615
clutches and power-stop control	5176	5343	-0.06308	1.019106	1.187891	1.397043	1.687259
Coating apparatus	8801	14008	0.366637	1.471426	1.820473	1.722242	1.720846
Coating implements with material supply	3806	3995	-0.01009	1.01585	0.839811	1.753797	1.128139
Coating processes	21454	36876	-57.7091	1.656781	3.034859	3.712963	0.99862
Coded data generation or conversion	10364	12391	-1.0119	1.883153	1.798222	2.637482	4.333135
Coherent light generators	11185	11942	0.187753	1.574833	1.545867	1.477267	1.334201
Coin handling	603	578	-0.61452	1.758974	1.331954	2.202614	3.539976

Colloid systems and wetting agents; subcombinations thereof; processes of	1629	2862	-0.78225	1.449004	2.994848	1.655311	2.038112
Combinatorial chemistry technology: method, library, apparatus	1049	1696	-19.6902	1.20091	2.716421	4.705593	5.643591
Combustion	5166	5935	-0.29537	1.007231	1.608181	1.409472	2.184782
Communications, electrical: acoustic wave systems and devices	5593	6464	-0.15667	1.043313	1.586362	1.519694	2.545278
Communications: directive radio wave systems and devices (e.g., radar, radio navigation)	10012	12001	-3.06201	1.246941	2.226745	3.559681	2.496052
Communications: electrical	26439	35234	-0.89312	1.158224	1.543946	4.290208	1.637288
Communications: radio wave antennas	9712	10821	-1.06536	1.907636	4.073926	1.681786	3.213153
Compositions	13483	20520	-103.031	1.395796	3.030443	2.940001	5.441692
Compositions: ceramic	6105	8320	-0.9235	1.114784	1.354911	1.378361	1.828212
Compositions: coating or plastic	10895	16000	0.444764	1.844311	1.963938	2.31893	3.46195
Compound tools	783	874	-0.38832	2.434809	1.783441	4.385523	5.249833
Computer graphics processing and selective visual display systems	25531	28635	-0.72931	1.546761	1.394016	3.086275	0.652646
Computer-aided design and analysis of circuits and semiconductor masks	5598	7482	-0.53736	1.006527	1.302942	1.35093	3.252926
Concentrating evaporators	517	872	-3.66178	1.488071	2.279836	5.053053	5.961675
Conveyors, chutes, skids, guides, and ways	590	716	-4.07976	1.579096	2.228516	5.21467	7.473022
Conveyors: fluid current	1691	2278	-3.98069	1.265718	3.282175	3.545386	0.950875
Conveyors: power-driven	10814	12170	0.26445	1.803901	1.949012	2.615008	3.13909
Crop threshing or separating	852	902	0.012138	1.719056	1.52301	1.612772	1.845833
Cryptography	3640	4582	0.352674	1.714472	1.69279	3.016879	2.92894
Cutlery	7292	7239	-0.52398	1.070203	2.179595	0.305431	2.293922
Cutters, for shaping	1585	1513	-0.42302	1.508838	2.120583	0.951758	1.723343
Cutting	8037	9197	-0.25946	1.945693	2.282979	2.70318	4.37172
Cutting by use of rotating axially moving tool	3174	3537	-21.4302	1.709043	2.342727	1.531944	1.960538
Data processing: artificial intelligence	2952	4301	-1.58477	1.51004	1.686933	2.905473	1.847866
Data processing: database and file management or data structures	13089	19661	0.222121	1.151643	1.137282	1.190881	2.840054
Data processing: financial, business practice, management, or cost/price determination	7427	12220	-0.7253	1.855772	2.212391	1.87067	2.403236
Data processing: generic control systems or specific applications	13057	22360	-27.8085	1.475721	2.371143	2.396424	3.243236
Data processing: measuring, calibrating, or testing	12846	22931	0.274596	1.523364	1.966796	2.198008	3.265539
Data processing: presentation processing of document, operator interface processing, and screen saver display processing	8464	13462	-17.6279	1.525125	1.998018	2.154861	3.534933
Data processing: software development, installation, and management	4552	7135	-0.26558	1.494457	1.689856	1.33109	0.470454

Data processing: speech signal processing, linguistics, language translation, and audio compression/decompression	7675	8295	-101.65	2.171848	3.582398	0.649875	1.364793
Data processing: structural design, modeling, simulation, and emulation	3773	7118	-0.52451	2.420651	1.413365	1.622139	2.541098
Data processing: vehicles, navigation, and relative location	13705	18044	-1.06716	1.694272	1.598579	1.667201	2.449097
Demodulators	931	1289	0.84486	1.39146	2.307789	2.689603	3.009633
Dentistry	7700	6509	-11.8474	2.29589	3.344301	1.652339	4.918932
Deposit and collection receptacles	716	832	0.155034	1.575289	3.157945	5.214231	2.288176
Dispensing	14122	14966	-35.1874	1.906471	0.983529	0.679231	1.379045
Distillation: apparatus	1286	1776	-0.82825	1.549862	2.255866	2.781112	2.090873
Distillation: processes, separatory	1967	3021	-1.30269	1.589499	3.095527	2.737303	2.439005
Drug, bio-affecting and body treating compositions	77571	73700	-4.0375	1.848277	2.527427	4.693021	5.860491
Drug, bio-affecting and body treating compositions	42145	52548	-0.47091	1.164109	1.379043	1.836029	5.384632
Drying and gas or vapor contact with solids	5917	7843	0.011194	2.535911	3.296676	4.31734	3.656191
Dynamic information storage or retrieval	15170	11568	-0.82649	1.225379	2.504555	0.740194	1.855349
Dynamic magnetic information storage or retrieval	22576	18961	-0.83166	1.277494	1.894022	4.226385	1.785112
Dynamic optical information storage or retrieval	2341	2686	-0.41125	1.763406	3.497068	2.303187	2.596459
Earth boring, well treating, and oil field chemistry	1832	2026	-2.01133	1.348715	2.321201	3.021407	1.333943
Earth working	3495	3367	-0.04595	1.085111	1.658219	2.249412	2.963423
Education and demonstration	5499	7199	-0.41034	2.213617	1.468252	2.035755	2.85551
Electric heating	22549	29640	-59.3928	2.268137	3.331122	3.238237	1.987603
Electric lamp and discharge devices	12660	14325	-50.9357	1.211104	1.515397	1.924804	1.895994
Electric lamp and discharge devices: systems	12202	13663	-0.87435	1.39484	1.779673	1.324439	4.239555
Electric lamp or space discharge component or device manufacturing	2407	3877	-0.31537	1.574742	1.874112	2.347498	1.697249
Electric power conversion systems	7926	8626	-0.56579	1.923828	2.543979	4.726626	3.326357
Electric resistance heating devices	2266	3139	-4.89894	3.003422	4.186823	2.651283	0.856971
Electrical audio signal processing systems and devices	7600	8231	0.441733	0.655121	0.776817	0.76194	1.500739
Electrical computers and digital data processing systems: input/output	13268	18231	-1.47657	1.755094	2.440706	1.011176	4.00636
Electrical computers and digital processing systems: interprogram communication or interprocess communication (ipc)	1729	3365	-2.23736	1.984037	3.10491	2.21399	1.800055
Electrical computers and digital processing systems: memory	14310	16590	-0.63075	1.232374	1.69294	0.828155	1.508397
Electrical computers and digital processing systems: multicomputer data transferring	13638	22444	-60.183	1.79611	2.114731	1.672023	2.074769
Electrical computers and digital processing systems: processing architectures and instruction processing (e.g., processors)	6708	7146	-0.96486	1.196326	1.525296	1.334955	1.411928

Electrical computers and digital processing systems: support	8780	13111	0.254776	0.957998	1.175116	1.697631	3.410814
Electrical computers and digital processing systems: virtual machine task or process management or task management/control	1840	3518	-0.90192	3.264452	4.052807	4.790879	2.000149
Electrical computers: arithmetic processing and calculating	7128	8330	-1.88023	2.282361	3.427182	4.387114	5.21062
Electrical connectors	32353	27648	-4.13067	1.752225	1.930778	3.760966	4.308385
Electrical generator or motor structure	17655	20338	-0.66835	1.548988	2.058864	2.071739	1.76027
Electrical pulse counters, pulse dividers, or shift registers: circuits and systems	2289	3175	-0.52023	1.09632	1.687868	2.145467	2.591989
Electrical resistors	2810	4409	-0.18911	1.899386	2.677896	1.858348	1.827809
Electrical transmission or interconnection systems	5820	9288	-0.04497	1.035524	1.343361	1.346959	1.965634
Electricity: battery or capacitor charging or discharging	4277	5395	0.038917	1.646346	2.33872	3.206299	3.82378
Electricity: circuit makers and breakers	9536	11137	-11.6327	2.328758	3.115115	0.34071	3.11226
Electricity: conductors and insulators	11564	17071	-1.73746	1.407338	0.348694	1.432795	3.366129
Electricity: electrical systems and devices	27056	36058	0.304715	0.584305	1.538243	2.229047	3.385242
Electricity: electrothermally or thermally actuated switches	2250	2342	-1.25907	1.580918	2.848561	2.238661	1.044552
Electricity: magnetically operated switches, magnets, and electromagnets	5332	6639	-0.47343	1.077089	1.283278	1.862064	1.25088
Electricity: measuring and testing	28643	35281	-4.58812	1.622443	2.092453	4.780315	5.208762
Electricity: motive power systems	13575	16835	0.080379	0.801316	1.039082	1.012662	1.402391
Electricity: motor control systems	805	1181	-4.976	1.545813	3.101229	1.909986	2.172166
Electricity: power supply or regulation systems	5596	6676	-0.02238	1.098141	0.586997	0.745894	1.650206
Electricity: single generator systems	1234	1540	-1.61968	1.490069	3.017881	4.209023	1.294787
Electrolysis: processes, compositions used therein, and methods of preparing the compositions	10359	15616	-6.44307	1.643733	2.203045	2.892108	1.99023
Electronic digital logic circuitry	8334	8144	-0.97923	1.562433	2.422597	1.822042	2.960091
Electrophotography	18277	14274	-1.1715	1.672901	2.049252	0.612915	1.547871
Elevator, industrial lift truck, or stationary lift for vehicle	2928	2881	-0.09654	2.231049	3.31714	4.316007	2.546559
Elongated-member-driving apparatus	2871	2589	-0.01006	1.296748	2.910271	3.64497	0.535697
Endless belt power transmission systems or components	4230	4129	-0.31811	0.968778	1.465568	3.597998	2.919219
Envelopes, wrappers, and paperboard boxes	4821	4481	0.300484	1.833349	1.811146	3.477671	4.011945
Error detection/correction and fault detection/recovery	16913	24217	-0.68026	1.462188	1.666795	4.116977	2.18012
Etching a substrate: processes	4457	8431	-0.13927	1.873367	0.244562	0.79334	1.563238
Excavating	2755	3019	-0.43559	2.747785	2.570707	1.017476	0.999599
Exercise devices	8179	7192	0.26739	0.732362	0.729831	1.541146	2.316346

Expanded, threaded, driven, headed, tool-deformed, or locked-threaded fastener	4792	4700	-0.01852	1.253749	1.597844	3.48002	3.971474
Expansible chamber devices	3090	3932	-0.35267	1.157543	3.007927	1.857127	1.123514
Explosive and thermic compositions or charges	1646	1870	-37.7858	1.286186	1.871426	2.051577	1.300753
Fabric (woven, knitted, or nonwoven textile or cloth, etc.)	4326	7372	-0.03198	1.028495	1.229661	1.783404	2.407748
Facsimile and static presentation processing	14399	14774	-0.61367	1.535723	1.223076	2.066184	2.16602
Fences	1366	1477	-0.97542	1.134656	1.592876	3.989432	0.954468
Fire escape, ladder, or scaffold	3892	4103	0.03855	0.749422	1.388203	1.306595	1.525038
Fire extinguishers	1434	1592	-0.80825	1.29087	3.015301	4.003239	1.05074
Firearms	3518	3101	0.178975	1.519664	0.642362	0.622637	2.037806
Fishing, trapping, and vermin destroying	6979	7121	0.145516	0.985861	1.50281	1.572449	1.959476
Flexible bags	2229	2302	-1.08869	2.736022	3.200786	3.740809	2.188092
Flexible or portable closure, partition, or panel	3974	3757	-0.69863	1.002617	1.251413	1.558469	1.819271
Fluent material handling, with receiver or receiver coacting means	6196	8096	-0.00828	1.169379	1.225538	1.378043	1.163733
Fluid handling	20388	23163	-0.98391	1.288695	2.002131	0.571142	2.56022
Fluid reaction surfaces (i.e., impellers)	4858	5813	0.15716	1.115777	1.242181	1.640451	2.252799
Fluid sprinkling, spraying, and diffusing	12307	13691	-17.3184	2.563985	3.233624	3.738374	2.11953
Fluid-pressure and analogous brake systems	5260	4336	-0.70159	1.158668	1.376736	1.92697	3.822299
Food or edible material: processes, compositions, and products	17716	23344	-0.64108	1.457542	1.972078	2.023657	2.374311
Foods and beverages: apparatus	6766	7667	0.133703	1.924877	2.768477	2.577879	5.201977
Foundation garments	640	673	-1.43778	0.636771	0.633986	0.566172	0.90152
Freight accommodation on freight carrier	1675	1867	-0.44052	1.029019	2.247373	1.732498	2.732136
Friction gear transmission systems or components	517	377	-3.73281	1.459665	2.128779	6.817158	7.443231
Fuel and related compositions	3158	3807	-4.45232	3.577349	3.780931	2.251997	1.3822
Furnaces	3193	4440	-0.69948	1.480493	3.689231	3.160428	1.607462
Games using tangible projectile	12958	11202	-4.67647	1.841152	2.639632	4.444803	6.176363
Gas and liquid contact apparatus	3530	4391	-0.57589	1.483745	2.4751	2.30727	2.538021
Gas separation	3497	4937	0.273067	1.401159	1.608603	1.405837	0.7519
Gas separation: apparatus	3736	5721	-0.63013	1.04002	1.494053	1.319928	1.763777
Gas separation: processes	4847	7462	-0.50543	1.061131	1.582528	1.300665	1.82708
Gas: heating and illuminating	1395	1920	-0.47291	1.105214	2.420639	2.398568	2.315396
Gear cutting, milling, or planing	2598	3458	-0.35316	1.163087	1.458008	0.354622	1.30933
Geometrical instruments	9132	10716	-0.04183	0.956381	1.010594	1.254416	2.236394
Glass manufacturing	6347	7450	-0.53551	1.320908	1.284017	1.376554	1.981635
Handling: hand and hoist-line implements	4973	6011	-0.05583	1.001009	0.430937	2.356277	2.237158
Harness for working animal	536	523	-4.4078	1.608373	2.354123	3.984543	3.927368

Harvesters	5656	5310	-0.01321	1.025036	1.266239	1.527269	2.620686
Hazardous or toxic waste destruction or containment	1519	2808	-0.39301	1.767905	2.932407	3.84088	2.047314
Heat exchange	10247	13147	0.028658	1.830226	2.0206	2.131591	3.368624
Heating	3284	4858	-1.30759	1.701973	1.602551	0.865344	0.678591
Heating systems	1052	1455	-0.25527	1.287977	0.472885	2.557075	0.77158
High-voltage switches with arc preventing or extinguishing devices	1749	2087	-0.43933	1.531254	2.601869	3.255293	3.669017
Horizontally supported planar surfaces	3332	3976	0.030271	1.804727	2.675445	2.795554	4.011688
Horology: time measuring systems or devices	4621	4491	-0.04552	0.979239	1.363535	1.436226	1.348197
Hydraulic and earth engineering	9508	10645	0.1337	0.989648	0.753739	0.719743	1.209944
Illumination	16014	16523	-1.51241	1.213523	1.530689	4.049798	1.702606
Image analysis	19004	21484	-39.7713	1.485812	0.281751	3.804737	1.74413
Imperforate bowl: centrifugal separators	1374	1389	-0.79514	1.095172	1.862272	1.729059	1.655998
Implements or apparatus for applying pushing or pulling force	3562	4008	0.090807	1.3791	1.289372	1.300504	1.610129
Incremental printing of symbolic information	20217	16406	-0.67861	1.580277	1.777982	1.120725	1.597802
Induced nuclear reactions: processes, systems, and elements	5777	6766	-0.76548	1.497494	1.468321	2.102325	2.429988
Inductor devices	2768	4202	0.186372	1.066764	2.451769	4.290465	1.245761
Industrial electric heating furnaces	1501	2279	-8.48681	2.081006	1.752623	2.808021	2.039887
Information security	2184	3936	0.014618	1.748206	2.317085	4.310074	1.614923
Interactive video distribution systems	2875	4292	0.022639	1.637676	3.242934	4.29917	2.242335
Internal-combustion engines	35783	31000	-4.03587	1.632849	2.266696	3.861931	3.989697
Interrelated power delivery controls, including engine control	4016	4226	-2.61884	1.543372	1.946548	1.724619	2.791769
Jewelry	940	973	-0.46298	1.057091	2.603517	3.014275	2.374919
Joints and connections	6321	8192	-0.05534	1.08654	1.400447	1.543161	1.671662
Land vehicles	25236	25409	-0.52081	1.096457	1.56392	4.262045	1.381324
Land vehicles: bodies and tops	11400	13774	-1.16603	1.663157	2.6194	2.705595	3.906842
Land vehicles: wheels and axles	1951	2171	-4.97621	1.785459	2.157953	0.423658	1.714836
Liquid crystal cells, elements and systems	10212	9458	-78.617	2.186714	2.452918	1.276692	2.632026
Liquid heaters and vaporizers	2445	3163	-8.14158	1.135623	3.168653	4.227973	1.134574
Liquid purification or separation	28346	34930	-4.33047	1.925067	2.381693	4.558878	5.252402
Locks	6357	5876	0.19876	1.048967	1.220282	1.326104	2.661091
Lubrication	1428	2018	-1.65872	1.532959	1.367578	3.26003	2.352372
Machine element or mechanism	13781	16053	-19.5048	1.656842	3.374341	2.908798	2.479895
Manufacturing container or tube from paper; or other manufacturing from a sheet or web	3650	4204	0.322592	1.007827	1.436219	1.304783	2.97336
Marine propulsion	3925	3269	-0.69064	1.873866	2.558314	2.352269	4.043077

Material or article handling	13478	16695	0.45901	1.839461	1.571854	1.766621	1.611789
Measuring and testing	39929	52778	-4.29537	1.693824	2.219957	4.60795	5.465871
Mechanical guns and projectors	2386	1969	0.073394	1.181257	3.49494	1.535884	2.531569
Metal deforming	11489	14261	-0.12466	1.896467	1.804385	2.337146	4.29609
Metal founding	6820	8917	-0.16305	1.224877	1.292049	1.604397	1.557766
Metal fusion bonding	6576	10231	0.060417	1.456525	2.367801	3.972997	3.52707
Metal tools and implements, making	807	976	-1.43558	0.860278	1.423667	1.245958	3.51446
Metal treatment	9689	15175	0.155144	0.702284	0.957569	0.747776	1.413569
Metal working	29446	44304	-5.12021	2.29895	2.490629	4.101147	4.04488
Metallurgical apparatus	3406	5564	0.023166	1.399263	2.640387	4.090598	2.57472
Mineral oils: processes and products	7489	7113	-87.736	1.687519	3.267057	4.056012	4.970381
Mining or in situ disintegration of hard material	2333	2575	-2.12184	3.063363	3.619422	4.122306	3.88072
Miscellaneous active electrical nonlinear devices, circuits, and systems	19752	20106	-0.55932	1.413023	1.982583	2.203516	1.409722
Miscellaneous hardware (e.g., bushing, carpet fastener, caster, door closer, panel hanger, attachable or adjunct handle, hinge, window sash balance, etc.)	4748	5085	-9.48067	1.808954	1.368676	1.304494	3.218095
Modulators	832	1158	-4.28648	2.159023	1.925272	0.528346	3.842758
Motion video signal processing for recording or reproducing	6149	5901	-0.54349	1.288733	1.260322	1.769166	2.302032
Motor vehicles	11310	14120	-51.922	2.072154	3.554029	3.857556	5.573194
Motors: expansible chamber type	3587	3899	0.221586	1.076503	1.205754	1.954942	2.946373
Movable or removable closures	4852	5599	-0.79608	1.014641	1.383994	1.639943	2.797376
Multicellular living organisms and unmodified parts thereof and related processes	5612	6469	-0.45517	1.723313	0.411569	0.555794	1.049868
Multiplex communications	29093	33903	-1.02702	1.217275	3.663647	4.415035	2.625042
Music	8217	6180	-42.3539	1.580364	2.796817	3.740925	4.840851
Optical communications	5566	6865	-0.01343	0.663916	0.666407	0.792742	1.968753
Optical waveguides	21474	23430	-1.08032	1.847738	0.622621	3.873897	1.747136
Optical: systems and elements	30699	28991	-4.04792	1.931035	1.994874	4.615576	3.282432
Optics: eye examining, vision testing and correcting	5348	5152	-0.11127	0.72221	1.30279	1.655997	1.80244
Optics: image projectors	3104	3483	-3.16041	2.451444	0.845509	3.341373	1.406128
Optics: measuring and testing	20508	26736	-0.59224	1.345811	1.771371	1.454261	1.106978
Optics: motion pictures	960	1075	-0.75908	2.262838	2.541522	4.338085	2.089746
Ordnance	3645	4111	0.213718	1.021723	0.716959	1.64384	2.873481
Organic compounds -- part of the class 532-570 series	10499	13898	0.243985	1.331935	1.322128	1.641738	2.594584
Organic compounds -- part of the class 532-570 series	8874	16026	-1.4491	1.717536	1.20444	3.28474	2.962882
Organic compounds -- part of the class 532-570 series	8349	13237	0.333659	0.740674	0.913082	1.105588	1.464456
Organic compounds -- part of the class 532-570 series	8199	12548	-0.85899	1.36651	2.05295	2.733002	3.796176

Organic compounds -- part of the class 532-570 series	7273	11566	0.409653	1.651827	1.716746	1.94389	2.545181
Organic compounds -- part of the class 532-570 series	7202	11766	-1.08147	1.322243	2.244056	2.034113	4.283244
Organic compounds -- part of the class 532-570 series	6983	11783	-0.35225	0.877764	1.240409	2.949407	2.750827
Organic compounds -- part of the class 532-570 series	6264	9962	-0.54609	1.22664	1.372333	2.439327	1.594063
Organic compounds -- part of the class 532-570 series	5846	9754	-0.47437	1.002102	0.408103	1.462562	1.742395
Organic compounds -- part of the class 532-570 series	5420	7469	-0.33923	1.97948	2.899328	1.428399	1.43573
Organic compounds -- part of the class 532-570 series	4703	6717	-0.7143	1.129992	1.41673	1.420698	2.945517
Organic compounds -- part of the class 532-570 series	4655	7295	-0.55457	0.958838	1.238533	1.325514	2.316061
Organic compounds -- part of the class 532-570 series	2657	2719	-0.30788	1.052352	1.710449	1.651378	1.873993
Organic compounds -- part of the class 532-570 series	2233	3826	-0.20186	1.317429	1.593946	3.580613	2.100385
Organic compounds -- part of the class 532-570 series	1978	2451	-1.30695	1.575263	1.8226	3.943454	2.480264
Organic compounds -- part of the class 532-570 series	1645	2608	0.087841	1.06184	1.343503	3.360078	2.022887
Oscillators	5676	6989	0.150775	1.898548	4.004128	4.571547	3.618716
Package and article carriers	5458	5905	-0.03142	1.259165	1.203432	1.300409	4.106798
Package making	10362	11273	-7.18618	1.169504	1.890905	3.241483	3.74237
Paper making and fiber liberation	6650	7799	-0.45861	1.091464	1.338094	1.628527	2.386326
Perfume compositions	900	971	-0.9223	2.183285	3.708136	1.781943	3.898694
Photocopying	6615	7029	-0.86044	1.114135	2.017503	2.35341	2.632611
Photography	14769	8870	-1.09981	1.3593	2.195551	2.022381	2.126904
Pipe joints or couplings	7759	8740	0.121274	1.916572	2.249974	2.223127	3.835544
Pipes and tubular conduits	4046	5442	-0.09983	1.581721	2.503107	2.482951	4.235385
Planetary gear transmission systems or components	4753	4521	-0.85973	1.017162	1.223545	1.34423	1.297711
Plant husbandry	4147	4668	-0.77059	1.008335	1.771843	1.438594	2.306168
Plant protecting and regulating compositions	7219	6200	-0.67901	1.338476	2.225835	2.303716	2.253031
Planting	1315	1439	-2.32844	2.626896	3.738039	3.150587	4.180486
Plastic and nonmetallic article shaping or treating: processes	23606	37309	-1.24611	1.448964	2.643738	4.244204	2.255103
Plastic article or earthenware shaping or treating: apparatus	12322	15791	-1.31706	1.225604	1.603163	2.087203	2.230359
Powder metallurgy processes	2021	3688	-0.73816	2.010594	3.473707	3.750889	1.136139
Power plants	19856	21586	-0.81741	1.071687	1.837244	2.922444	2.692575
Presses	3165	3574	-18.6891	2.244671	3.677139	2.909683	1.929226
Prime-mover dynamo plants	2218	2996	-0.79131	1.008246	2.405451	2.734238	1.113906
Printed matter	1897	2195	-0.30495	1.059126	1.28671	1.50353	2.036256
Printing	9863	9600	0.197768	1.569026	1.752305	1.706798	3.114007
Prosthesis (i.e., artificial body members), parts thereof, or aids and accessories therefor	9547	9133	-65.4146	0.527434	2.374976	2.729663	2.575527
Pulse or digital communications	21372	21897	-0.44956	1.063402	1.70393	1.702889	1.011914
Pumps	11791	13911	0.21694	1.653209	1.67396	1.694511	2.11236

Radiant energy	32267	40560	-4.614	1.744357	2.241702	4.23838	3.684977
Radiation imagery chemistry: process, composition, or product thereof	34957	28688	-2.11155	1.015356	3.533009	3.240149	2.587077
Railway rolling stock	2177	2231	-2.33155	2.523343	3.648722	1.92512	1.836949
Railway switches and signals	805	956	0.312825	2.113782	2.276024	3.284042	1.266128
Railways	2467	2717	-0.38301	1.120138	1.926318	2.195547	1.743861
Railways: surface track	766	863	-51.8922	1.872604	2.471728	2.828321	5.799744
Receptacles	10524	12403	-0.06671	1.132419	1.803716	1.35729	1.316323
Record receiver having plural interactive leaves or a colorless color former, method of use, or developer therefor	2744	2662	0.09442	1.221678	1.363705	2.183236	2.83059
Recorders	1385	2109	-2.59869	1.261696	2.515776	1.276384	2.363365
Refrigeration	18922	20859	-0.65406	1.255319	0.810777	2.379667	0.713351
Registers	10378	12614	-101.662	1.384054	2.839599	0.681507	2.867494
Resilient tires and wheels	4189	3759	0.014984	1.485033	2.383673	2.34171	3.524293
Road structure, process, or apparatus	3371	3579	0.142568	1.46115	1.200158	1.459606	1.952382
Roll or roller	845	1179	-0.98408	0.968728	1.537503	1.997492	2.551148
Rotary expansible chamber devices	4305	4286	-0.16524	1.23139	1.361794	1.485392	3.149522
Rotary kinetic fluid motors or pumps	6438	8314	0.216061	1.219399	1.342193	1.324715	2.093325
Rotary shafts, gudgeons, housings, and flexible couplings for rotary shafts	3075	3216	0.024714	1.240424	2.254725	3.657014	2.176902
Seal for a joint or juncture	6110	7345	-0.05478	1.157465	1.213106	1.315783	1.185697
Semiconductor device manufacturing: process	55195	47005	-4.06031	1.93241	1.667277	3.054512	5.325361
Severing by tearing or breaking	1012	1273	-0.28624	1.505322	2.330584	1.720133	1.500007
Sewing	5203	3910	-0.23786	1.532285	2.811426	2.711801	1.530603
Sheet feeding or delivering	7297	8186	0.358799	0.697417	0.620684	0.761647	1.544869
Sheet-material associating	1529	1795	-1.72924	1.262403	2.527443	1.295083	1.462313
Ships	7864	8548	-0.01472	0.89051	0.88309	1.495341	2.451395
Signals and indicators	1966	2471	-0.73281	2.402527	2.200516	0.371681	0.798817
Single-crystal, oriented-crystal, and epitaxy growth processes; non-coating apparatus therefor	3819	5818	-0.35008	1.581851	1.691856	1.714578	2.3194
Solid anti-friction devices, materials therefor, lubricant or separant compositions for moving solid surfaces, and miscellaneous mineral oil compositions	4992	5098	-1.01972	1.052363	1.59761	1.78916	1.26388
Solid material comminution or disintegration	6230	7703	-0.17957	1.601648	1.524147	1.635628	2.238804
Special receptacle or package	16620	18798	-20.2415	2.370117	0.343531	3.253028	0.685011
Specialized metallurgical processes, compositions for use therein, consolidated metal powder compositions, and loose metal particulate mixtures	6764	10873	0.23021	1.807462	1.809857	2.355272	3.561437
Spring devices	3606	4086	-0.08488	1.037518	1.311422	1.335836	2.558224
Static information storage and retrieval	26978	16709	-3.87834	1.749012	2.0912	3.045954	3.576336

Static molds	1716	2138	-0.92361	1.571413	4.145186	4.698728	1.697086
Static structures (e.g., buildings)	21684	22413	-0.78821	1.108134	1.467342	0.595368	0.939114
Stock material or miscellaneous articles	57013	80239	-4.1065	1.8291	2.393148	4.281499	5.482027
Stone working	1029	1249	-51.8921	1.268751	2.85355	3.746073	1.792659
Stoves and furnaces	7628	8175	0.4561	0.884502	0.551582	1.584807	2.213736
Sugar, starch, and carbohydrates	803	1286	-0.27869	1.199417	1.122038	0.541967	1.250511
Superconductor technology: apparatus, material, process	2125	2793	-5.59081	2.040208	3.262736	4.108501	2.589709
Supports	15064	17673	-98.433	1.891733	1.747147	1.726823	2.284383
Supports: cabinet structure	5256	6548	-0.11279	1.001352	1.3516	1.609942	2.945231
Supports: racks	6217	6806	0.016706	1.353639	2.265333	2.780712	2.563344
Surgery	27900	29504	-4.48185	1.756791	2.056848	3.438234	4.35837
Surgery	25082	22754	-0.53879	1.14468	1.500728	1.61274	2.268233
Surgery	21029	17461	-0.83738	1.578148	3.279441	2.543427	1.600082
Surgery	10202	11109	-9.25602	2.821513	0.619276	3.307718	3.161891
Surgery: kinesitherapy	2741	3071	-0.03054	1.255782	1.299387	3.920633	2.635215
Surgery: light, thermal, and electrical application	8800	7018	0.066258	0.515626	0.803751	0.971488	1.313182
Surgery: splint, brace, or bandage	3893	4103	0.267503	1.012216	1.200118	1.354948	2.999298
Synthetic resins or natural rubbers -- part of the class 520 series	23016	30615	-0.98247	1.125351	1.490969	1.952733	0.798927
Synthetic resins or natural rubbers -- part of the class 520 series	20198	25073	-1.53592	1.142885	1.968288	2.781701	1.32617
Synthetic resins or natural rubbers -- part of the class 520 series	16193	20154	-0.9381	1.229576	1.908856	3.335897	2.654603
Synthetic resins or natural rubbers -- part of the class 520 series	10977	14570	-12.2262	2.695866	2.790572	0.591903	3.333297
Synthetic resins or natural rubbers -- part of the class 520 series	7745	12689	-0.96349	1.323656	1.73167	1.29057	1.642532
Synthetic resins or natural rubbers -- part of the class 520 series	6898	9005	0.377079	1.561554	1.695516	1.661391	2.653845
Synthetic resins or natural rubbers -- part of the class 520 series	2425	4111	-0.85044	1.045196	1.65809	2.933733	2.389907
Telecommunications	24945	28696	-0.35085	1.136122	1.448294	1.590359	1.556167
Telegraphy	915	1214	-0.31254	1.831839	2.427313	1.564285	1.013995
Telephonic communications	17342	21630	-1.20118	1.237786	1.900582	3.02158	1.248116
Television	24039	24315	-95.6167	1.589368	2.503092	3.131743	4.464968
Tent, canopy, umbrella, or cane	2375	2198	-0.56529	1.525887	2.763637	4.270359	1.700829
Textiles: fiber preparation	1748	1524	0.126822	1.006032	3.191999	3.414352	2.594452
Textiles: fluid treating apparatus	1918	2280	-0.74752	1.088481	0.206715	1.818896	2.556427
Textiles: ironing or smoothing	828	1016	-1.04566	1.743562	2.018699	1.644002	1.912501
Textiles: knitting	2649	2217	-0.26277	1.167918	2.820537	0.580886	2.178877

Textiles: manufacturing	1416	1776	-0.93391	1.110504	2.509563	2.701379	0.637005
Textiles: spinning, twisting, and twining	4271	4006	-0.44694	1.72038	1.899031	2.12067	2.966946
Textiles: weaving	3463	2734	0.129703	1.05968	0.779607	1.302465	2.471634
Thermal measuring and testing	3681	5796	-0.05654	1.012726	1.208562	1.392452	2.873182
Tobacco	3402	3285	-0.50347	1.71005	1.272045	1.90069	1.949288
Toilet	3647	3709	0.030151	1.554877	1.526892	1.717347	3.498846
Tool changing	836	1212	0.252497	1.792715	2.072388	2.225088	2.417853
Tool driving or impacting	2370	2736	-2.12922	2.762708	3.216753	1.072998	2.435414
Tools	6162	5802	-15.1912	1.90441	0.617201	0.68184	2.990621
Traversing hoists	1126	1309	-1.16113	1.021133	1.926331	2.708974	3.496876
Trunks and hand-carried luggage	817	735	-0.1114	1.249752	0.937676	4.739527	4.158667
Turning	2053	2532	-0.84329	1.030941	1.927632	2.366983	1.412156
Typewriting machines	7402	8340	-1.0476	1.318567	1.697194	3.671797	3.95283
Valves and valve actuation	7235	9275	0.036156	1.94094	1.927768	2.059845	3.733555
Vehicle fenders	1006	1424	-0.72214	1.783199	2.123557	2.363091	2.118786
Ventilation	3638	4751	0.04705	1.002434	1.268502	1.406377	0.612066
Wave transmission lines and networks	8625	9682	-0.56979	0.961622	0.995994	1.278483	1.356733
Weighing scales	2589	3010	-1.05217	1.039263	3.353881	3.990741	1.608908
Wells	12972	10637	-79.9942	0.986606	2.473516	4.128449	2.76344
Wheel substitutes for land vehicles	820	819	-4.39953	2.733829	3.119915	0.631715	2.834134
Winding, tensioning, or guiding	13032	13183	0.255256	1.038445	0.840959	1.205727	1.93675
Wireworking	973	1245	-19.4382	2.166712	0.813478	3.012424	0.79966
Woodworking	3234	3190	-0.12268	1.491146	2.034801	2.659029	2.676031
Work holders	2891	3349	-1.04125	2.484105	1.426677	4.479091	4.268794
X-ray or gamma ray systems or devices	8858	9182	-0.76172	1.033983	1.579865	0.858757	2.365469

Table S3. Computational run times. We present empirical run times for estimating the model parameters as a function of the number of papers and number of authors in the field.

Field Code	Team size Average	Paper count	Author count	Running Time (hours)
PN	1.81	161308	121976	20.08
BP	1.07	73261	44471	18.46
UQ	1.31	30749	32424	15.2
WY	1.49	31773	34914	8.45
BI	1.34	22597	18827	8.27
YY	1.36	26408	26901	7.51
NM	1.27	20094	17519	7.38
EY	1.44	20483	19483	7.29
YJ	1.08	20868	14217	6.03
QJ	2.32	20660	24003	3.28
VS	1.91	14968	15873	5
OR	1.02	11847	9235	3.04
AF	1.86	1702	2317	2.32
IX	2.59	1816	3337	1.63
BD	2.89	582	1259	1.17

Table S4. Parameter convergence for 20 different fields. (**A**) Four different sets of initial conditions for testing the algorithm convergence. (**B**) For each initial condition, final parameter estimates averaging across the 20 fields. **(C)** For each field, final parameter estimates taking arithmetic mean and standard error across the different initial conditions. We see broad consistency in the final parameter estimates regardless of the different initial conditions.

Table S4A.

Initial Condition	ρ	β_2	β_3	β_4	β_5
1	2	.1	.1	.1	.1
2	4	.5	.4	.3	.2
3	-2	5	4	3	2
4	-4	10	10	10	10

Table S4B.

Final Results	$mean(\frac{1}{1-\rho})$	$mean(\beta_2)$	$mean(\beta_3)$	$mean(\beta_4)$	$mean(\beta_5)$
1	0.1942	1.5609	2.1874	4.3405	5.1453
2	0.1939	1.5604	2.1866	4.3430	5.1403
3	0.1945	1.5589	2.1877	4.3384	5.1384
4	0.1944	1.5614	2.1865	4.3400	5.1421

Table S4C.

Field Code	Number of Authors	$mean\left(\frac{1}{1-\rho}\right)$	$se\left(\frac{1}{1-\rho}\right)$	$mean(\beta_2)$	$se(\beta_2)$
PQ	1580	0.00474	0.00273	1.55205	0.01085
LQ	571	0.00116	0.00336	1.47234	0.00891
RZ	561	0.00386	0.00608	2.45250	0.03608
EY	1728	0.19140	0.03048	1.32522	0.03213
FF	528	0.11234	0.02998	1.18652	0.03329
XE	1105	0.11342	0.01269	3.27651	0.02063
AF	1329	0.26132	0.03320	1.91949	0.03304
PT	491	0.66194	0.07786	1.68077	0.02809
MQ	216	0.65029	0.07178	1.17054	0.00373
JO	3267	0.06357	0.02470	1.44762	0.02209
AE	485	0.37075	0.04931	3.50253	0.02383
YY	3883	0.43000	0.01239	3.41019	0.05204
DB	2895	0.17036	0.03803	2.00119	0.02496
PO	895	0.71089	0.08148	3.39815	0.04238
BI	1641	0.01105	0.00866	2.78181	0.02147
YE	455	0.08832	0.02587	2.91858	0.04192
SR	489	0.05050	0.01937	1.09944	0.01700
AM	371	0.13211	0.02276	2.27725	0.03859
PI	813	0.00568	0.00610	2.29859	0.03980
JI	979	0.20832	0.01888	2.51502	0.04338

Table S5. Parameter estimation using alternative outcomes. In the main text, we define the outcome variable, y, as the number of citations a work receives in the first 8 years after publication (papers) or application (patents). For 20 different field communities (see Table S4C) we consider parameter estimates of $\hat{\rho}$ and $\hat{\beta}_2 \dots \hat{\beta}_5$ when defining alternative outcome measures, including (a) a logarithm citation count and (b) an indicator variable for high impact works.

Outcome measure	ρ	β_2	β_3	β_4	β_5
Logarithm	0.19	1.42	1.47	2.5561	2.67
Home Run Indicator	0.14	2.14	2.68	2.3609	2.21

Note: The logarithm measure is $y = \log(c + 1)$, where c is the number of citations received in the first 8 years. We add 1 to include observations that receive zero citations. The "home run" measure is an indicator, $y \in \{0,1\}$, where $y = 1$ if the citations received, c, are in the upper 20th percentile in that field and year, and $y = 0$ otherwise.

Table S6. Parameter regression results for the relationship between the estimated values of ρ and β_2. Each observation is a field in the respective domain. Papers are considered in the first three columns and patents in the final three columns, using three different regression models. A statistically significant negative relationship occurs for all three regression models and in both the paper and patenting domains.

	Papers			Patents		
Regression Model	(1)	(2)	(3)	(1)	(2)	(3)
Coefficient	-0.121***	-0.350***	-1.897***	-0.294***	-0.427***	-1.057***
Standard Error	(0.020)	(0.048)	(0.254)	(0.051)	(0.074)	(0.172)
R-squared	0.17	0.23	0.24	0.08	0.08	0.09
Observations	182	182	182	384	384	384

Note: Regression model (1) examines $\frac{1}{1-\rho} = \theta\beta_2 + \epsilon$; regression model (2) examines $\frac{1}{1-\rho} = \theta\ln(\beta_2) + \epsilon$; and regression model (3) examines $ln\left(\frac{1}{1-\rho}\right) = \theta\ln(\beta_2) + \epsilon$. *** indicates statistical significance with p<0.01.

Table S7. Additional, author-order based indices for citation sharing in team papers, as collected in (28). The parameter r_i represents the integer position of the i^{th} author in the author list, N is the length of the author list (i.e., the team size), $\varphi = 1.618$ is the golden number used in the Golden p-index, and $0 < \alpha < 1$ is an arbitrary constant used in the Arithmetic Index (we set $\alpha = 0.5$).

Index	Description
First Author	All citations are credited to the first author
Lp-index (Linear Productivity)	$\frac{2}{r_i(N+1)}$
Golden p-index	$\begin{cases} 1 & N = 1 \\ \varphi^{2r_i - 1} & N \geq 2 \\ \varphi^{2r_i - 2} & N \geq 2 \ AND \ r_i = N \end{cases}$
Trueba-Guerrero Index	$\frac{2N - r_i + 2}{N(N+1)} \times \frac{2}{3}$
Proportional Index	$\frac{2 \times (1 - \frac{r_i}{N+1})}{N}$
Geometric Index	$\frac{2^{N-r_i}}{2^N - 1}$
Arithmetic Index	$\frac{1}{N} + \frac{N - 2 \times r_i + 1}{2} \times \alpha$
Harmonic Index	$\frac{\frac{1}{r_i}}{\sum_{j=1}^{N} \frac{1}{j}}$

Table S8. The prediction of the citation impact outcome for out-of-sample solo-authored papers based on featured metrics in the main text and the additional author-order based metrics described in Table S7. Predictive accuracy is measured in field-specific regressions (see main text and SI methods), with the median R^2 value presented below as a summary statistic.

Index	Median of regression R^2
Trueba-Guerrero Index	0.08
Geometric Index	0.08
Arithmetic Index	0.10
Golden p-index	0.10
Harmonic	0.11
Lp-index	0.14
Proportional Index	0.14
all	0.09
solo	0.20
pp	0.20
$\hat{a}$	0.31

Table S9. Crosswalk from NAS fields to WOS fields

NAS Field	WOS Field(S)
Systems Neuroscience	Neuroscience
Biochemistry	Biochemistry And Molecular Biology
Physics	(Physics, Patricles And Fields) (Physics, Atomics, Molecular) (Physics, Multidiciplinary) (Physics, Condenced Matter) (Physics, Applied) (Physics, Fluids, Plasma) (Physics, Mathematical)(Physics, Nuclear)
Economic Sciences	Economics
Engineering Sciences	(Engineering And Technology) (Engineering, Manufacturing) (Engineering, Environmental) (Engineering, Biomedical) (Engineering, Petroleum) (Engineering, Aerospace)(Engineering, Electric And Electronics) (Engineering, Multidisciplinary) (Engineering, Marine) (Engineering, Mechanical) (Engineering, Chemical) (Engineering, Chemical) (Engineering, Ocean) (Engineering, Geological) (Engineering, Industrial) (Engineering, Civil)
Medical Genetics, Hematology, and Oncology	(Genetics & Heredity) (Hematology) (Oncology)
Chemistry	(Chemistry, Applied) (Chemistry, Multidisciplinary) (Chemistry, Inorganic, Nuclear) (Chemistry, Physical) (Chemistry, Medicinal) (Chemistry, Analytical) (Chemistry, Organic)
Psychological and Cognitive Sciences	(Psychology, Clinical) (Psycholog, Educational) (Psychology, Biological) (Psychology, Experimental) (Psychology, Psychoanalysis)
Microbial Biology	(Microbiology) (Biotechnology And Applied Microbiology)
Animal, Nutritional, and Applied Microbial Sciences	(Agriculture, Dairy & Animal Sciences) (Nutrition & Dietetics)
Computer and Information Sciences	(Computer Science & Ai) (Computer Science, Hardware, Architecture) (Computer Science, Interdisciplinary) (Computer Application, Cybernetics) (Computer Science, Cybernetics) (Computer Science, Information Systems) (Computer Science, Software) (Computer Science, Theory & Methods)
Anthropology	Anthropology
Applied Physical Sciences	(Physics, Applied), (Physics, Fluids & Plasmas),(Physics, Multidisciplinary),(Physics, Condensed Matter)
Biophysics and Computational Biology	(Biophysics) (Mathematical & Computational Biology)
Mathematics	(Mathematics, Applied) (Mathematics)(Mathematics, Misc)
Physiology and Pharmacology	(Physiology) (Pharmacology & Pharmacy)
Immunology and Inflammation	(Immunology)
Applied Mathematical Sciences	(Mathematics, Applied), (Mathematics), (Mathematics, Miscellaneous)
Astronomy	Astronomy & Astrophysics
Evolutionary Biology	Evolutionary Biology
Geology	Geology
Geophysics	Geochemistry & Geophysics
Cellular and Developmental Biology	(Developmental Biology)(Cell Biology)
Cellular and Molecular Neuroscience	(Neurosciences), (Multidisciplinary)
Medical Physiology and Metabolism	Endocrinology & Metabolism
Plant, Soil, and Microbial Sciences	(Plant Sciences) (Transplantation) (Soil Sciences)
Genetics	Genetics & Heredity
Environmental Sciences and Ecology	(Environmental Sciences) (Environmental Studies) (Ecology)
Social and Political Sciences	(Social Sciences, Mathematical Models) (Social Sciences, Biomedical) (Social Sciences, Intedesciplinary) (Politics & Policy) (Political Science)

Table S10. Summary of matching algorithm outcomes for NAS members.

Type	Fraction of NAS Members
Unique Author ID and $\hat{a}$ available	0.45
Several Author ID	0.15
Less than 10 Publications	0.08
Unique Author ID and $\hat{a}$ not available	0.21
No match found – (with name, field, and/or affiliation)	0.07
Author ID does not have more than 30% of his/her papers in the listed primary/secondary field	0.02

Table S11. Median rank of NAS members in their corresponding cohort based on numerous different productivity indices, including all those in the main text and the additional measures defined in Table S7.

Index	Median rank (per-paper impact rank)	Median Rank (per-paper impact and paper count)
Trueba-Guerrero Index	0.73	0.78
Proportional Index	0.81	0.84
Arithmetic Index	0.79	0.82
Geometric Index	0.82	0.82
Harmonic	0.76	0.79
Golden p-index	0.74	0.78
Lp-index	0.82	0.84
First Author	0.76	0.77
h-index	--	0.83
i10 - index	--	0.74
solo	0.84	0.86
all	0.94	0.98
pp	0.92	0.98
$\hat{a}$	0.97	0.99