LEARNING WHEN TO QUIT:
AN EMPIRICAL MODEL OF EXPERIMENTATION

Bernhard Ganglmair
Timothy Simcoe
Emanuele Tarantino

Learning When to Quit: An Empirical Model of Experimentation
Bernhard Ganglmair, Timothy Simcoe, and Emanuele Tarantino
NBER Working Paper No. 24358
February 2018
JEL No. D83,O31,O32

ABSTRACT

We study a dynamic model of the decision to continue or abandon a research project. Researchers
improve their ideas over time and also learn whether those ideas will be adopted by the scientific
community. Projects are abandoned as researchers grow more pessimistic about their chance of
success. We estimate the structural parameters of this dynamic decision problem using a novel
data set that contains information on both successful and abandoned projects submitted to the
Internet Engineering Task Force (IETF), an organization that creates and maintains internet
standards. Using the model and parameter estimates, we simulate two counterfactual policies: a
cost-subsidy and a prize-based incentive scheme. For a fixed budget, subsidies have a larger
impact on research output, but prizes perform better when accounting for researchers' opportunity
costs.

Bernhard Ganglmair
University of Texas at Dallas
Naveen Jindal School of Management
800 W. Campbell Rd (SM31)
Richardson, TX 75080
ganglmair@utdallas.edu

Timothy Simcoe
Boston University
Questrom School of Business
595 Commonwealth Avenue
Boston, MA 02215
and NBER
tsimcoe@bu.edu

Emanuele Tarantino
University of Mannheim
L7, 3-5, 68131
Mannheim, Germany
etaranti@gmail.com

1 Introduction

At some point, every inventor abandons a promising idea. The decision to quit reveals that either the costs of continuing a line of research have increased, or expectations of its importance have declined. Policies that help researchers quickly discern whether specific projects will bear fruit should increase the overall productivity of R&D. But it is difficult to predict how changes in the research environment will influence researchers' rate of learning, because their beliefs and opportunity costs are rarely observed.

This paper proposes a structural model of Bayesian learning in R&D, estimates the model using a unique data set that contains information on both successful and abandoned projects, and uses the estimates to explore links between the research environment and the implied rate of learning. While many studies have examined the behavior and productivity of individual scientists and engineers (e.g., Stephan, 1996; Azoulay et al., 2011), very few (perhaps none) analyze how individuals allocate the key research input of *time* in the face of substantial uncertainty about the quality of ideas. That is the question at the heart of this paper.

Our model of learning combines a two-armed bandit with a more traditional finite-horizon optimal stopping problem to capture two different phases of the research process. The two phases are separated by a *breakthrough* that resolves any uncertainty about whether the project will succeed. Before a breakthrough, the decision to continue or abandon a project reflects the standard trade-off between exploration and exploitation common to the literature on experimentation (Roberts and Weitzman, 1981; Moscarini and Smith, 2001). Specifically, we adopt a discrete-time version of Keller et al. (2005) where quitting corresponds to the safe arm, and continuing a line of research corresponds to the risky arm. After a breakthrough, researchers face a

simple optimal stopping problem, where they compare the marginal costs and benefits of continuing to "polish" their idea. This notion that R&D projects have both a fundamental and a more implementation-focused dimension of quality, with greater uncertainty and more learning about the former, is similar to the Ellison (2002) model of economics publishing.[1]

Empirically, our model can be formulated as a non-stationary dynamic discrete choice problem with an unobserved state variable (i.e., whether a breakthrough has occurred). In each period, researchers decide whether to continue or abandon their project, with payoffs realized if and only if the team stops *after* a breakthrough occurs. We develop a maximum likelihood estimator similar to Pakes (1986), which yields estimates of researchers' opportunity costs, their prior beliefs about the distribution of project quality, and the rate of learning. We provide sufficient conditions for the existence of a unique solution to the model, yielding point-identification of the learning and cost parameters.

To estimate the model, we use data obtained from the public archives of the Internet Engineering Task Force (IETF), an organization that develops technology standards and protocols for managing the internet. A unique feature of this data set is that we observe each revision of every proposal, including projects that are eventually abandoned. Because proposals must achieve *consensus* in order to be published by the IETF, that is how we interpret the breakthrough in our model.[2] Payoffs are measured using U.S. patent citations to published standards.

Whereas prior studies of the IETF, including Simcoe (2012), have emphasized

---

[1] In Ellison (2002), economics papers have "q" quality based on the importance of a paper's main contribution, and "r" quality that reflects generality, robustness checks, extensions, etc. Authors learn over time what weights to place on these dimensions in the review process.

[2] In other applications, the breakthrough might correspond to achieving some key technical milestone. We have retained the term breakthrough for consistency with the theoretical literature, and because achieving a consensus to publish represents a key moment in the life of any research project.

non-cooperative standardization, this paper treats each new submission as an independent research project. We view cooperative R&D as an important part of the standard setting process, although as we explain below, our learning model could also be interpreted as the reduced form of a more complex non-cooperative bargaining game.

Our baseline estimates suggest that roughly 60 percent of submissions to the IETF could generate a consensus, and that only about one third of those projects actually result in a publication. Estimates of a key parameter that governs the speed of learning indicate that the consensus arrival rate is around 17 percent per revision. Consistent with our priors about how the IETF works, we obtain higher estimates of initial quality, and faster rates of learning, for the sub-sample of proposals initiated inside an IETF Working Group. We consider a number of robustness checks and extensions to the model, including the addition of parametric heterogeneity in the rate of learning.[3]

Finally, we use the structural parameters from our model to conduct two counterfactual policy experiments. In the first experiment, we compare an R&D subsidy that lowers the per-period cost for all projects to a prize that increases the payoff for successful projects (holding fixed the total budget, in patent citations, of the two policies). Although both policies can increase research output, prizes perform better after accounting for the private costs of over-developing bad ideas. Our second counterfactual considers the cost of misaligned priors, and shows that within our model over-confidence is more costly than pessimism.

This study contributes to several streams of literature. The theory builds upon

---

[3]Results for models that introduce parametric heterogeneity are described in Appendix D. We find that learning is positively correlated with communication (in the form of emails to IETF listservs mentioning a specific proposal) and that there is a non-monotonic inverted-U relationship between learning and commerciality (the share of email addresses with a dot-com top-level domain).

early models of experimentation (Weitzman, 1979; Roberts and Weitzman, 1981; Moscarini and Smith, 2001) that were later extended to consider strategic experimentation (Bolton and Harris, 1999; Keller et al., 2005; Bonatti and Hörner, 2011).[4] To our knowledge, this is the first paper to estimate the parameters that drive Bayesian learning in such a model.

Our work also contributes to a broader literature that estimates structural models of learning. Applications to consumer behavior include Erdem and Keane (1996), as well as the various studies described in Ching et al. (2013) and Ching et al. (2017). Both Crawford and Shum (2005) and Dickstein (2014) consider learning in the context of matching patients to prescription drugs.[5] Notably, the latter paper formulates the problem as a multi-armed bandit. Doraszelski et al. (forthcoming) consider the learning that takes place as firms enter a new market and converge to a competitive equilibrium.

We believe our paper is the first to estimate a structural model of learning within the context of R&D.[6] There are, however, many studies that take a reduced-form approach to learning in R&D. Allen (1966) is one of the first papers to characterize how engineers allocate time and prioritize alternative problem-solving approaches, and his early analysis of the decision to abandon an idea bears many similarities to this paper. More recently, Howell (2017) and Gross (2017) have considered the relationship between feedback and learning. Krieger (2017) examines knowledge spillovers in pharmaceutical R&D, while Bennett and Snyder (2017) critique a large strategy

---

[4]Kerr et al. (2014) discuss the importance of experimentation in innovation and entrepreneurship. In the vein of the literature on strategic experimentation, Akcigit and Liu (2016) study investment decisions when firms do not know their competitors' dead-end results. Instead, Moscarini and Squintani (2010) analyze R&D competition when privately informed firms learn from exit of their opponents.

[5]Related, Chan and Hamilton (2006) study learning in a setting featuring patients deciding on when to quit a clinical trial.

[6]In a context that abstracts from learning, but related to R&D, Bhattacharya (2017) considers a procurement model of multi-stage innovation.

literature that attempts to draw inferences about learning from failure.

The balance of the paper is organized as follows: Section 2 describes our empirical context and the data set, and also provides several descriptive empirical results that motivate our theoretical modeling choices. Section 3 introduces the formal model of learning, and then describes our estimation strategy. Section 4 presents the baseline empirical results. Section 5 describes our counterfactual policy experiments, and Section 6 concludes.

2 The Internet Engineering Task Force

The IETF is a voluntary organization that contributes to the engineering and evolution of internet technologies. It is considered the principal entity engaged in the development of new internet standards (Hoffman, 2012). During the early 1990s, the IETF evolved from a small quasi-academic networking community into a high-stakes forum for technical decision-making (Simcoe, 2012), and it is now populated by researchers and engineers from public and private organizations, including firms, universities, and government agencies.

The IETF has played a major role in the technological development of the internet. The most famous IETF standard is the TCP/IP protocol used to route all internet traffic from source to destination. A more recent standard is the Session Initiation Protocol (SIP), which enables VoIP ("Voice over IP") services, and is used for video conferencing, instant messaging, file transfer, and online games. Other IETF standards include the Network Address Translation (NAT) protocol, which defines an interface between private and public networks (thereby dramatically expanding the supply of IP addresses), and the Dynamic Host Configuration Protocol (DHCP),

which distributes addresses among machines attached to a network.[7]

2.1   IETF Standards Development

Much of the IETF's activity takes place on email lists, and in a series of face-to-face meetings held three times each year. In both forums, IETF participants propose new standards and extensions to existing protocols, debate the merits of these proposals, and decide whether to collectively endorse them.

The IETF standard setting process begins when participants identify a problem and form a working group (WG) to consider solutions.[8] To prevent forum shopping and overlapping technical agendas, new working groups must be approved by an advisory board called the Internet Engineering Steering Group (IESG). Once a working group is formed, anyone can submit a proposal by posting it to a public repository. New proposals are called *Internet Drafts* (IDs). Although some IDs describe an entire protocol, such as SIP, NAT, or DHCP, the majority propose updates and extensions to the functionality of existing standards.

The authors of a proposal can decide whether they want their ID to be considered within a WG or as an individual submission (typically because there is no appropriate WG, or because the suitable WG is busy on other projects). After their submission, IDs are debated on the email discussion lists and at IETF meetings. Authors incorporate feedback from the community into their proposals. IDs are continually revised, and, as a statutory rule, an unpublished ID expires after six months of inactivity, leading to its removal from active online directories.[9] This process continues until either the IETF reaches a *rough consensus* in favor of publishing the ID, or the authors

---

[7]Table A.1 lists some of the more prominent standards created and then certified by the IETF.

[8]This synopsis draws on Hoffman (2012) and Simcoe (2012).

[9]Since an ID's removal does not necessarily imply that authors cannot resume the project, in our estimation sample construction, we take a more conservative approach to define an abandoned ID.

abandon their proposal.

While the IETF provides no formal definition, rough consensus is often described as the *dominant view* of the relevant working group and implies support from well over 51 percent of active participants. In practice, a working group chair decides whether consensus has been reached.[10] If the working group chair declares a consensus, there is a *last call* for comments from the WG, and the ID is submitted to the IESG. The IESG reviews the proposal and issues a second last call for comments from the entire IETF community. Any comments or formal appeals are reviewed by the IESG and may be referred back to the working group for resolution. If the IESG is satisfied that a consensus exists within the working group and sees no problem with the ID, it will be published as a *Request For Comments* (RFC).

The IETF publishes two types of RFCs: standards-track and nonstandards-track. Standards-track RFCs define new features and protocols, which progress in maturity from Proposed Standard to Draft Standard and then finally to Internet Standard.[11] Nonstandards-track RFCs either provide information that complements a standard, such as implementation guidelines and technical references (Informational RFCs), or describe new protocols that are not yet ready for standardization (Experimental RFCs). Although standards- and nonstandards-track RFCs go through an identical publication process, nonstandards do not place normative constraints on new products and therefore tend to be less controversial. In some of our empirical models, we treat nonstandards-track RFCs as a control sample where consensus occurs automatically.

In the formal model described below, authors learn over time whether their proposal will produce a consensus and, therefore, a payoff. This leads to a dynamic

---

[10]There is also an appeal process for the authors who feel that a working group chair, or the IESG, has taken the wrong choice on their ID.

[11]The IETF requirements for advancing protocols along the standards track to the level of Draft or Internet Standard include multiple independent implementations and significant experience in the field. We do not study these later stages of the standardization process here.

decision problem in which Bayesian updating helps rationalize the decision to abandon a proposal. There is a clear analogy to other kinds of research, and in practice, consensus standard setting involves a large element of cooperative R&D, as modeled, for instance, in Ganglmair and Tarantino (2014). At the same time, one could interpret the learning process as a reduced-form representation of noncooperative bargaining, such as the model of standardization in Simcoe (2012). Before introducing the formal model, however, we describe our data and present several stylized facts that it is meant to capture.

2.2 Data

Our primary data source is the online archives of the IETF. These archives contain the full text for each version of every Internet Draft submitted after July 1990, for both published RFCs and abandoned proposals, along with various pieces of bibliographic information.[12,13] From each version of every ID, we extract the names and affiliations of all authors in order to construct measures of team size, experience, and composition. We also download all of the archived IETF email lists and create proposal-specific measures of feedback by counting the number of messages that mention a specific Internet Draft. As a measure of *ex-post* commercial significance, we count the number of non-patent prior art citations from U.S. patents to an Internet Draft or associated RFC. The data appendix provides details on the process that we used to collect, clean, link, and merge these data into a panel where the unit of analysis is an ID-version.

---

[12]Our raw sample ends in September 2015. We drop IDs that are still considered active at that point. We further exclude IDs initiated before 1996 and in or after 2010 (to avoid selection on outcomes). For more details on the final sample construction, see the data appendix (in Appendix E).

[13]In most cases, the IETF file-naming convention allows us to identify the relevant working group, and track successive revisions within a project. For instance, the ID `draft-ietf-homenet-arch-02` (entitled "Home Networking Architecture for IPv6") was associated with the `homenet` WG and was succeeded by `draft-ietf-homenet-arch-03` and eventually published as RFC 7368. The data appendix (in Appendix E) describes how we link IDs to create a single project when there is a change in file name.

Table 1: Summary Statistics

	Full Sample	Working Group	Standards-track Abandoned	Standards-track Published	Nonstandards-track
WG (%)	24.44	100.00	14.11	65.25	46.39
Team Size (Author Count)	2.28	2.45	2.22	2.43	2.48
Experience (max Projects)	15.01	15.69	13.50	21.87	16.84
Versions	3.55	5.60	2.09	9.33	6.71
Communication (Emails)	21.20	33.78	9.87	69.19	41.03
Published RFC (%)	23.97	56.10	0.00	100.00	100.00
Citations	2.99	8.30	0.76	12.23	7.19
N (Projects)	16,091	3,932	12,234	2,210	1,647
N (Versions)	57,179	22,025	25,511	20,622	11,046

Notes: This table reports sample means at the Internet Draft (proposal) level. Team Size is the number of authors of the initial draft; Experience (max Projects) is the number of IDs the most prolific author has completed before the initial draft of a given project; Communication is the average number of emails that reference a project; Citations is the number of U.S. patent prior art citations to a given project. For additional information on the construction of these variables, see the data appendix (in Appendix E).

Table 1 reports variable means for five groups of IETF proposals: the full sample, proposals initially submitted to a WG, abandoned IDs, Proposed Standards, and nonstandards-track RFCs. Note that in defining the last three samples, we are conditioning on outcomes. The first three rows of the table report means of predetermined variables, while variables below the solid line are realized during or after the publication process.

A number of interesting patterns can be seen in Table 1. First, although working group proposals constitute just one quarter of the full sample, about two thirds of standards and half of nonstandards-track RFCs originate within a WG. Second, the author-teams on individual and WG submissions have similar size and experience levels. In particular, even for non-WG proposals, the average team has an author with 15 prior submissions, suggesting that they will not differ in their rate of learning due to inexperience. There is, however, a strong positive correlation between experience and publication that can be seen by comparing the average experience across the

Published and Abandoned columns. Moving to the endogenous variables, we see that published IDs go through many more revisions than abandoned proposals and are therefore mentioned in many more emails. And finally, we observe a twelve-fold increase in U.S. patent citations for Published IDs, consistent with our assumption that payoffs are contingent on achieving a breakthrough.

2.3 Descriptive Analysis

To motivate our theoretical model of learning, we now present several stylized facts about the internet standard setting process. This part of the analysis has three objectives: (1) to establish that proposals are changing over time, (2) to characterize the hazard rates for success and abandonment, and (3) to link revisions with commercial significance.

2.3.1 Revision Distance

Figure 1 shows how projects change during the revision process.[14] To create the figure, we treat each version of a proposal as a "bag of words," calculate the cosine distance between that version and original submission, and then plot the mean distance (conditional on the number of revisions) for standards- and nonstandards-track proposals, respectively.[15]

Although it is hard to give a natural interpretation to this distance metric, the figure suggests that each revision takes the proposal further from the initial submission, with the rate of change rapid at first, and gradually diminishing. The shape of

---

[14]We use the terms "revision" and "version" interchangeably. The number of revisions, therefore, includes the very first version of the draft (and is thus equal to the total number of versions).

[15]Formally, the cosine distance of a version T from the initial version is defined as: $1 - \frac{x_T \cdot x_1}{\|x_T\| \|x_1\|}$, where x_T is the vector of word frequencies for version T and x_1 the vector for the initial (i.e., first) version. See Appendix E for more details.

Figure 1: Cosine Distance to First Draft
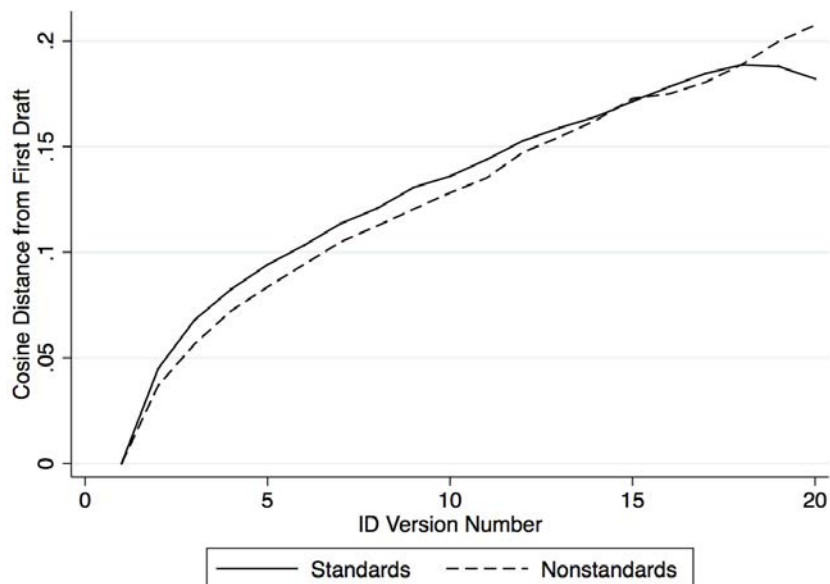
the plot is very similar for standards- and nonstandards-track proposals. In fact, if one conditions on nonstandards submitted to a WG, or those eventually published as an Experimental RFC, we cannot statistically distinguish between the two curves.[16]

2.3.2 Hazard Rates

Figure 2 plots the hazard rates (i.e., the probability of exit, conditional on survival to version t) for standards- and nonstandards-track proposals. We assume that standards-track proposals can experience three possible events in each period (abandon, publish, or continue), whereas nonstandards-track IDs have just two potential events (publish or continue).[17]

The most striking aspect of Figure 2 is the rapidly decreasing hazard of abandon-

---

[16]Specifically, we estimate a regression model that includes a dummy for each version-number interacted with a nonstandards-track indicator and cannot reject the hypothesis that the interaction coefficients are jointly zero.

[17]In practice, some standards-track proposals may morph into nonstandards, and it may initially be unclear which track an individual ID is on. Our interviews with IETF participants suggest these scenarios are infrequent.

Figure 2: Hazard Rates
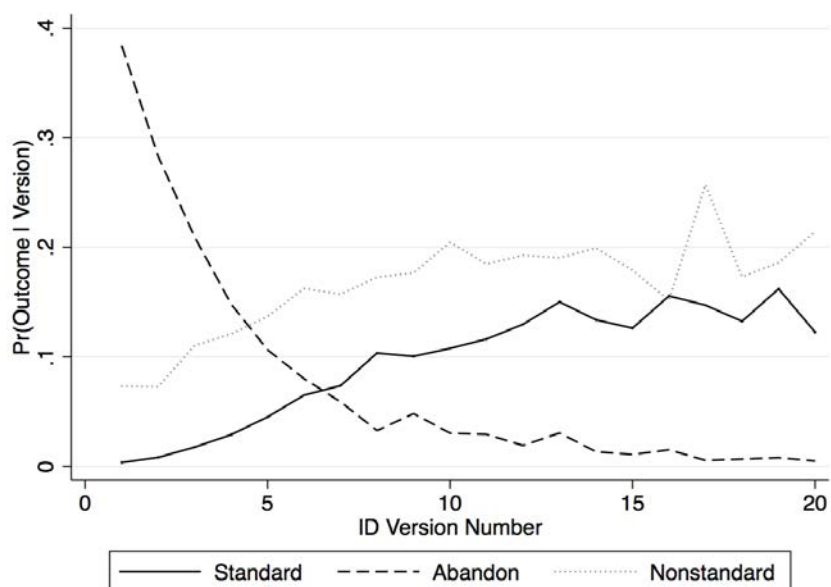
ment: almost 40 percent of proposals are never revised, but conditional on reaching a fifth version, the probability of abandonment falls below 10 percent. At the same time, the hazard of becoming a Proposed Standard climbs from zero on initial submission to about 10 percent by the eighth version. Together, these two hazard rates imply that the probability of publication increases sharply with the number of revisions. The increasing odds of publication are consistent with the arrival of "good news" over time, which will be a feature of our learning model.

Another feature of Figure 2 is that the standards-track publication hazard never climbs above 16 percent, even though the hazard of abandonment falls to nearly zero. This suggests that author-teams continue to revise proposals even after it becomes clear that they will be published, and it leads us to propose a two-phase model rather than a theory where consensus leads to immediate publication.

Finally, the hazard of nonstandards-track publication begins at around 8 percent (implying that some are published immediately) and climbs to around 20 percent af-

12

ter 10 revisions. Thus, although standards and nonstandards exhibit a similar degree of textual change (see Figure 1), nonstandards are more likely to be published in any given period. This provides some justification for our later assumption that there is no learning on the nonstandards-track—because lower commercial stakes imply immediate consensus—even though both tracks provide opportunities for improvement.

2.3.3 Revision and Citation

The final piece of our descriptive analysis examines the relationship between revisions and U.S. patent citations to published RFCs (our proxy for commercial impact).[18] Specifically, we estimate linear regressions of the form

$$Cites_i = \alpha_y + (\beta_1 + \beta_2 \ Nonstandard_i) * \log(Versions_i) + \varepsilon_i \tag{1}$$

where $Cites_i$ is a count of citations to published RFC i, and α_y are publication-year effects that control for right-censoring in the citation process. Figure 3 plots separate fitted values for standards- and nonstandards-track RFCs, assuming a publication year of 2000. The dotted lines correspond to estimates from a non-parametric model where we replace $\log(Versions_i)$ with dummies for each value of the variable $Versions_i$.

There are two main take-aways from this analysis. First, citations increase (at a declining rate) with the number of revisions. And second, citations to nonstandards-track RFCs are lower than for standards. We use fitted values from these regressions as our measure of expected payoffs in the structural model described below. Thus, in the absence of monetary measures of cost and benefit, citations will serve as the nu-

---

[18]Roach and Cohen (2013) provide a discussion of measurement issues associated with counting non-patent prior art references in U.S. patents, and suggests that these citations are a superior measure of knowledge reuse than more commonly used metrics based on citations to other patents.

Figure 3: Expected U.S. Patent Citations
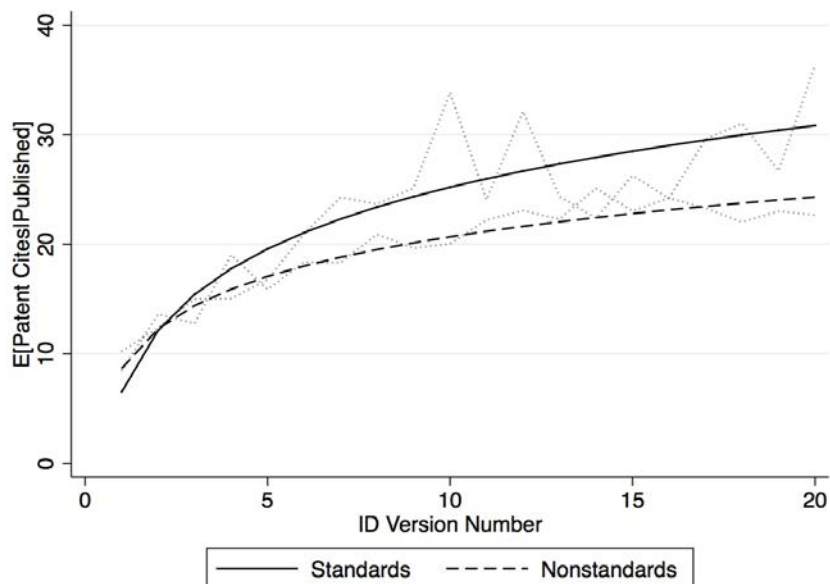
meraire, providing a unit of measure for teams' opportunity costs and the evaluation of counterfactual policies.

3 Bayesian Model of Experimentation and Learning

3.1 Overview

Figure 4 illustrates the basic structure of our model. At $t = 0$, a team of researchers is endowed with an idea that has one of two types, $\theta \in \{\text{good}, \text{bad}\}$. This type might represent technical or commercial feasibility, but in our application it stands for the possibility of reaching consensus. We model the team as a single agent that does not know its type but can learn through costly experimentation. With each experiment, there is a strictly positive probability that good ideas yield a *breakthrough*, whereas

14

Figure 4: Stylized Model of R&D Process

$-F_1$ $-F_2$ $-F_3$ $-F_\tau$ Breakthrough $\hat{\pi}(\tau+1)$ $\hat{\pi}(\tau+2)$ $\hat{\pi}(\cdot)$
$-F_{\tau+1}$ $-F_{\tau+2}$ $-F_{...}$

good

$t=1$ 2 3 τ $\tau+1$ $\tau+2$ ...

⟵ Abandon ⟶   ⟵ Publish ⟶

bad

Abandon

$t=1$ 2 3 4 ... ...

$-F_1$ $-F_2$ $-F_3$ $-F_4$ $-F_{...}$ $-F_{...}$ $-F_{...}$
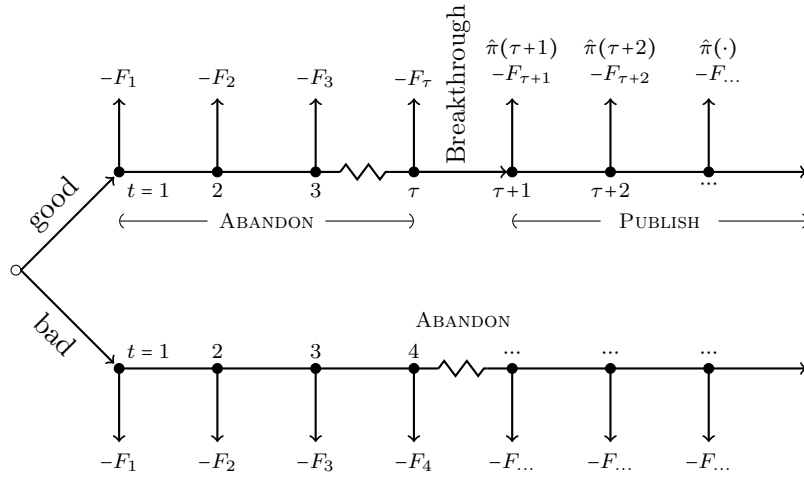
bad ideas always fail. Thus, teams that do not have a breakthrough grow increasingly pessimistic about their type. Experimentation continues until the team decides to stop.

Payoffs depend on whether a team's idea is published (after breakthrough) or abandoned (before breakthrough), and the number of revisions. More specifically, the gross benefits of publication are $\hat{\pi}(t)$, a project's scrap value is zero, and the cumulative (i.e., up to t) value of the stochastic revision costs is F_t. This leads to *ex-post* payoffs $\hat{\pi}(t) - F_t$ for published projects, and $-F_t$ for abandoned projects.

Before a breakthrough, this model is a discrete time version of the two-armed bandit problem analyzed by Keller et al. (2005). The decision to stop is the safe arm, and the decision to continue is the risky arm.[19] After a breakthrough, it becomes a simple optimal stopping problem, where the team compares the expected costs and benefits of revision.

When fitting the model to data, we estimate $\hat{\pi}(t)$ using citations data and assume that this function is known to both the team and the econometrician. Based on

---

[19] Many mathematicians refer to this type of problem, with one "safe" arm, as a one-armed bandit, or degenerate two-armed bandit problem.

publication outcomes, the econometrician also knows if a breakthrough occurred, but not when. The key structural parameters introduced below are the prior probability of having a good idea, p, the breakthrough arrival rate, b, and the (deterministic part of) opportunity costs of experimentation, F.[20] When the hazard of breakthrough is constant, there is a simple closed-form expression for the team's posterior beliefs (Heidhues et al., 2015) that simplifies estimation. We assume a finite horizon and solve the model using backwards induction, which allows us to characterize all possible states in which a breakthrough has occurred. In that sense, our estimation strategy is similar to Pakes (1986), except that where he takes costs as known and estimates the distribution of benefits, we do the reverse.

3.2 Model

A team of risk-neutral agents initiates a project with type $\theta \in \{\text{good}, \text{bad}\}$.[21] The research environment is characterized by three state-variables. The project's latent type θ is unobserved to both the team and the econometrician. The integer t represents the number of *versions* created and is observed by both team and econometrician. Finally, the indicator variable σ_t denotes that the team has achieved a breakthrough; it is observed by the team and revealed to the econometrician when the project is either published or abandoned after $T \leq \overline{T}$ versions.

In each period $t \geq 1$, the team decides whether to revise its project and advance to period $t + 1$. Revisions are costly, but increase the expected benefits of a successful project, as explained below. Revisions also provide information about θ. Specifically,

---

[20]The per-period costs of experimentation are the sum of a deterministic cost and a random cost shock.

[21]We assume the initial version of the project comes at zero cost and can therefore ignore the team's entry decision.

we assume the team knows the following transition probabilities for σ:

$$\Pr(\sigma_{t+1} = 1 | \sigma_t = 0, \theta = \text{good}) = b \tag{2}$$

$$\Pr(\sigma_{t+1} = 1 | \sigma_t = 0, \theta = \text{bad}) = 0 \tag{3}$$

$$\Pr(\sigma_{t+1} = 1 | \sigma_t = 1, \theta = \text{good}) = 1. \tag{4}$$

Equation (2) states that good projects have a constant per-period breakthrough arrival rate b; Equation (3) states that bad projects never experience a breakthrough; and Equation (4) states that breakthrough is an absorbing state—once a project is revealed to be good, that information cannot be withdrawn.[22]

Now consider the team's beliefs, $\Pr(\theta = \text{good} | t, \sigma_t) \equiv \hat{p}(t, \sigma_t)$. At the start of the research process, we assume the team's priors over project type are $\hat{p}(0, 0) = p$, with $0 < p < 1$. After a breakthrough, the team knows its type, so $\hat{p}(t, 1) = 1$. Before a breakthrough, Equation (2) and repeated application of Bayes' rule show that:

$$\hat{p}(t, 0) = \frac{p (1 - b)^t}{(1 - p) + p (1 - b)^t}. \tag{5}$$

For $0 < b < 1$, the function $\hat{p}(t, 0)$ is strictly decreasing in t, so the team becomes more pessimistic with time as its revisions fail to yield a breakthrough.

To see how learning influences a team's decision to revise or quit, we must place some additional structure on that choice. Let $\hat{\pi}(t)$ represent the expected quality of an idea, and $F_t \equiv \sum_{k=0}^{t-1} (F(k, \sigma_k) + \varepsilon_k)$ the cumulative cost of t versions, where $F(t, \sigma_t)$ is the deterministic portion of per-stage opportunity costs, and ε_t is a random cost

---

[22]Note that discrete time with endogenous costly revisions (which we interpret as a substantive change to the underlying idea) make this a model of *active* learning, in the spirit of Ericson and Pakes (1995), even though the stochastic process described by Equations (2)–(4) is quite general and could also be assumed to operate independently of the team's control.

shock.[23] The *ex-post* payoff to a team that stops after T versions is therefore

$$\pi(T, \sigma_T) = \sigma_T \hat{\pi}(T) - F_T. \tag{6}$$

We assume the team does not discount future payoffs.[24] Because the benefits $\hat{\pi}(T)$ are conditional on breakthrough, we use the terms "published" and "abandoned" for projects terminated when $\sigma_T = 1, 0$, respectively. The sequence of steps in each period t is:

- Step $t.1$: The team observes whether there was a breakthrough in $t - 1$, and it observes the private cost shock ε_t.

- Step $t.2$: Given the state of the project, σ_t, the team's beliefs, $\hat{p}(t, \sigma_t)$, and the marginal cost of a revision, $F(t, \sigma_t) + \varepsilon_t$, the team decides whether to stop or continue. Stopping implies a permanent decision to publish or abandon the project.

We are now ready to analyze the team's dynamic decision problem.

3.3   Recursive Characterization

The team's objective is to choose a contingent plan of actions that maximizes its payoff. In each period, stopping yields the payoffs in Equation (6). Alternatively, the team can create a revision and preserve the option to continue the project. The

---

[23]Recall that we assume the initial version comes at zero cost so that $F(0,0) = \varepsilon_0 = 0$.

[24]Magnac and Thesmar (2002) showed that the discount rate and payoffs in a dynamic decision problem are not jointly identified, and it is common practice to fix the former parameter. In principle, we could add a per-revision discount rate to $\hat{\pi}(t)$. However, setting the discount factor to unity seems a reasonable choice in our empirical setting, where lags in *ex-post* deployment of the standard are likely to be more consequential than those in initial protocol development.

resulting value function is

$$V(t, \sigma_t) = \max\left\{\sigma_t \hat{\pi}(t), E_{\varepsilon,\sigma}(V(t+1, \sigma_{t+1})) - F(t, \sigma_t) - \varepsilon_t\right\}, \tag{7}$$

where $E_{\varepsilon,\sigma}(\cdot)$ is the expectation with respect to the realization of breakthrough and future cost shocks, which we assume to be independent from each other. As can be seen in (7), the value function depends on whether breakthrough has been reached in the project's development process. The two state variables, t and σ_t, evolve according to a first-order Markov process. Because σ_t is not observed by the econometrician, the team's decision problem is non-stationary. This introduces serial correlation in the controlled stochastic process generating the value functions.[25] For this reason, we find it useful to solve the team's recursive decision problem under the assumption that the time horizon is finite, with the value of the potential standard dropping to zero after $\overline{T}$ revisions.[26] This will allow us to account for all possible states of the problem (i.e., the stages when a breakthrough could have occurred).

The optimal decision rule consists of a series of cut-points $\{\bar{\varepsilon}_t^\sigma\}_{t=1}^T$, such that in each period the team will revise the project if and only if $\varepsilon_t \leq \bar{\varepsilon}_t^\sigma$. We assume i.i.d. cost shocks with cumulative distribution G, and for notational convenience define the state-contingent continuation probabilities $G^\sigma(t) \equiv G(\bar{\varepsilon}_t^\sigma)$. To find the optimal cut-points, we first consider the stopping problem following a breakthrough, when $\sigma_t = 1$, and then turn to the two-armed bandit problem that precedes a breakthrough, when $\sigma_t = 0$.

---

[25]We refer to the stochastic process generating $\{V(t)\}_{t=1}^\infty$ as "controlled" because, although it is inherently random, it is also affected by the team's decision to continue.

[26]To grasp how strong this assumption is, note that only the projects with no breakthrough by $\overline{T}$ are affected. In our empirical analysis, the probability that a project receives cost shocks that lead to $\overline{T}$ revisions without a breakthrough is infinitesimal at estimated parameter values.

Post-breakthrough. Suppose a breakthrough occurred at some $t < \overline{T}$. A team that reaches the terminal period will stop and obtain payoffs of $\hat{\pi}(\overline{T}) > 0$.[27] In $\overline{T} - 1$, the value function is $V(\overline{T} - 1, 1) = \max\{\hat{\pi}(\overline{T} - 1), \hat{\pi}(\overline{T}) - F(\overline{T} - 1, 1) - \varepsilon_{\overline{T}-1}\}$ and the team will revise the project if and only if

$$\varepsilon_{\overline{T}-1} \leq \bar{\varepsilon}^1_{\overline{T}-1} \equiv \hat{\pi}(\overline{T}) - \hat{\pi}(\overline{T} - 1) - F(\overline{T} - 1, 1).$$

Otherwise the team stops and publishes the project. This implies that the value function in period $\overline{T} - 2$ is $V(\overline{T} - 2, 1) = \max\{\hat{\pi}(\overline{T} - 2), E_\varepsilon(V(\overline{T} - 1, 1)) - F(\overline{T} - 2, 1) - \varepsilon_{\overline{T}-2}\}$, where $E_\varepsilon(\cdot)$ is the expectation with respect to future cost shocks only, given that uncertainty over σ realized in $\tau < t$, and

$$E_\varepsilon(V(\overline{T} - 1, 1)) = G^1(\overline{T} - 1)\left[\hat{\pi}(\overline{T}) - F(\overline{T} - 1, 1) - \varepsilon_{\overline{T}-1}\right] + \left(1 - G^1(\overline{T} - 1)\right)\hat{\pi}(\overline{T} - 1).$$

The team continues solving this problem backwards through $t = 1$. Hence, in each $t \in [1, \overline{T})$ and given any continuation value function $\hat{V}(t + 1, 1)$, the team chooses to

$$\begin{cases} \text{Stop} & \text{if } \varepsilon_t > \bar{\varepsilon}^1_t \\ \text{Continue} & \text{if } \varepsilon_t \leq \bar{\varepsilon}^1_t. \end{cases}$$

The value function is then

$$V(t, 1) = \begin{cases} \hat{\pi}(t) & \text{if } \varepsilon_t > \bar{\varepsilon}^1_t \\ E_\varepsilon(V(t + 1, 1)) - F(t, 1) - \varepsilon_t & \text{if } \varepsilon_t \leq \bar{\varepsilon}^1_t, \end{cases} \tag{8}$$

---

[27]The process is force-terminated in $\overline{T}$ so that there are no decisions and $G^1(\overline{T}) = G^0(\overline{T}) = 0$.

and the post-breakthrough cost cut-points are defined as

$$\bar{\varepsilon}_t^1 \equiv E_\varepsilon(V(t+1,1)) - \hat{\pi}(t) - F(t,1). \tag{9}$$

Pre-breakthrough. Before a breakthrough, the team faces a two-armed bandit problem. In particular, they must choose between stopping to obtain a fixed (nil) retirement value, and running a costly experiment that may reveal their project's type. At any stage t where $\sigma_t = 0$, the value function is

$$V(t,0) = \max\{0, E_{\varepsilon,\sigma}(V(t+1,\sigma_{t+1})) - F(t,0) - \varepsilon_t\}, \tag{10}$$

where $E_{\varepsilon,\sigma}(\cdot)$ is defined above. The project's continuation value is

$$E_{\varepsilon,\sigma}(V(t+1,\sigma_{t+1})) = b\hat{p}(t,0)E_\varepsilon(V(t+1,1)) + (1 - b\hat{p}(t,0))E_{\varepsilon,\sigma}(V(t+1,0)). \tag{11}$$

If the team continues in t, it incurs costs $F(t,0) + \varepsilon_t$ and expects continuation payoffs $E_{\varepsilon,\sigma}(V(t+1,\sigma_{t+1}))$ as in Equation (11), where breakthrough arrives with probability $b\hat{p}(t,0)$ and the post-breakthrough value function $V(t+1,1)$ is given by Equation (8).

To find the sequence $\{V(t,0)\}_{t=1}^{\overline{T}}$, we start in the terminal period, $\overline{T}$, where, if $\sigma_{\overline{T}} = 0$, the process ends and the team's payoffs are 0. In $\overline{T} - 1$, the team will revise the project if its expected payoffs are nonnegative, or

$$\varepsilon_{\overline{T}-1} \leq \bar{\varepsilon}_{\overline{T}-1}^0 \equiv b\hat{p}(\overline{T}-1,0)\hat{\pi}(\overline{T}) - F(\overline{T}-1,0).$$

The team then solves this problem backwards through $t = 1$.

By analogy to the pre-breakthrough phase, for each $t \in [1,\overline{T})$ and given any continuation value function $V(t+1,0)$, the team chooses to abandon the project if

and only if $\varepsilon_t > \bar{\varepsilon}_t^0$. The value function is then

$$
V(t,0) = \begin{cases} 0 & \text{if } \varepsilon_t > \bar{\varepsilon}_t^0 \\ E_{\varepsilon,\sigma}(V(t+1,\sigma_{t+1})) - F(t,0) - \varepsilon_t & \text{if } \varepsilon_t \le \bar{\varepsilon}_t^0, \end{cases}
$$

and the cost cut-points are defined as

$$
\bar{\varepsilon}_t^0 \equiv E_{\varepsilon,\sigma}(V(t+1,\sigma_{t+1})) - F(t,0). \tag{12}
$$

Comparing Equations (9) and (12) shows how a breakthrough influences the team's decision. Before a breakthrough, the marginal benefits of continuation are $E_{\varepsilon,\sigma}(V(t+1,\sigma_{t+1}))$, whereas afterwards they are $E_\varepsilon(V(t+1,1)) - \hat{\pi}(t)$. The first expression begins large but declines with t as the team grows more pessimistic. Although the first term in the second expression is larger than pre-breakthrough expected benefits, the marginal gains are net of $\hat{\pi}(t)$, because after a breakthrough the team is only "polishing" an idea that it knows will be published. Hazards of quitting and publication will reflect both processes, which in turn depend on costs $F(t,\sigma_t)$, benefits $\hat{\pi}(t)$, and the speed of learning as governed by (b,p).

3.4 Likelihood Function

To derive the likelihood function, assume that $\hat{\pi}(\cdot)$ and $G(\cdot)$ are known, and consider a project published after T versions. In order to account for the fact that a breakthrough could have occurred at any $\tau = 0, ..., T-1$ (and is observed by the team at the beginning of any $\tau+1$), it is helpful to define a function $\rho(\tau, T)$ as the probability of breakthrough

in $t = \tau$ and publication in $t = T$. This function is equal to

$$\rho(\tau, T) = b \left(1 - b\right)^{\tau} \left(1 - G^1(T)\right) \prod_{j=0}^{\tau} G^0(j) \prod_{k=\tau+1}^{T-1} G^1(k). \tag{13}$$

Summing over all periods when a breakthrough could occur (and accounting for the prior probability of a project being good), we can write the likelihood of publication in T as

$$\Pr(T, \sigma_T = 1) = p \sum_{\tau=0}^{T-1} \rho(\tau, T). \tag{14}$$

The log-likelihood for projects published in period T is then equal to

$$LL(T, \sigma_T = 1 | b, p, \boldsymbol{F}) = \log(p) + \log\left(\sum_{\tau=0}^{T-1} \rho(\tau, T)\right), \tag{15}$$

where $\boldsymbol{F}$ denotes a vector of cost parameters.

Now consider the likelihood of abandoning a project after T versions:

$$\Pr(T, \sigma_T = 0) = \left(1 - p\right) \prod_{k=0}^{T-1} G^0(k) \left(1 - G^0(T)\right) + p \left(1 - b\right)^T \prod_{k=0}^{T-1} G^0(k) \left(1 - G^0(T)\right). \tag{16}$$

The first term accounts for bad projects where $\sigma_t = 0$ for all t, and the second term accounts for good projects that have no breakthrough. Hence, the log-likelihood for abandonment in a given T is equal to

$$LL(T, \sigma_T = 0 | b, p, \boldsymbol{F}) = \sum_{k=0}^{T-1} \log(G^0(k)) + \log(1 - G^0(T)) + \log\left(\left(1 - p\right) + p \left(1 - b\right)^T\right). \tag{17}$$

Given data on revisions and publication outcomes (T_i, σ_{T_i}), we can now write the

23

log-likelihood for a sample of N projects indexed by i:

$$LL(b, p, \boldsymbol{F}) = \sum_{i=1}^{N} LL(T_i, \sigma_{T_i} | b, p, \boldsymbol{F}), \tag{18}$$

and search for the parameter vector $(b, p, \boldsymbol{F})$ that maximizes this object.[28]

3.5 Identification

To illustrate how the data and model jointly identify the structural parameters, we consider a two-period example.[29] The team observes whether there was breakthrough at $t = 0$ and decides whether to stop or continue in $t = 1$, with a revision providing the possibility of breakthrough before the process ends at $\overline{T} = 2$. Let $S_t > 0$ and $A_t > 0$ represent the number of projects published and abandoned, respectively, in period t, and define N_t^σ as the (unobserved) number of projects in state σ at the start of each period. We normalize total projects to one, so that $N_1^1 + N_1^0 = 1 = \sum_t (S_t + A_t)$.

As a starting point, note that the model and data place bounds on the values of p and b. We must have $p \geq S_1 + S_2$ because breakthrough, and thus publication, requires a good idea. We also know that $b > 0$, because if $b = 0$, then there would be no breakthroughs, and therefore no publication. Similarly, if $A_2 > 0$, then we must have $b < 1$, because the marginal benefits of revision are strictly negative when $b = 1$ and $\sigma_1 = 0$. These bounds suggest, at an intuitive level, that p is identified by variation in the total share of published projects, while b reflects variation in the timing of the two outcomes (though timing, of course, also depends on costs).

---

[28]We have ignored the possibility that projects could be censored. However, the likelihood can easily accommodate right-censored IDs, because, for any value of the parameters, we know the probability of a breakthrough by t, and the probability of a series of cost-shocks that imply continuation to that period. Appendix C provides a formal derivation of the log-likelihood function when right-censored projects are included.

[29]The arguments that we make in the two-period example generalize to $\overline{T}$ periods. We provide formal proofs of such a $\overline{T}$-period extension in Appendix B.

Without further assumptions, we can say no more about the values of the structural parameters. Specifically, we can prove that:

Theorem 1. *For any $b \geq \underline{b}$, there exists a unique $p(b)$ and a unique value of $F(1,1)$ and $F(1,0)$, both functions of b, such that for all t, the probability of publication (abandonment) in period t equals S_t (A_t).*

Proof. For any choice of p and b, we can solve for N_1^1 (the share of projects with a breakthrough at $t = 0$, observed at the outset of stage $t = 1$) in two ways. First, our assumptions about the learning process imply that $N_1^1 = pb$. And second, rational expectations imply that

$$N_1^1 = S_1 + S_2 - \frac{\hat{p}(1,0)b}{1 - \hat{p}(1,0)b} A_2, \tag{19}$$

where the last term represents the number of breakthroughs at $t = 1$, given posteriors $\hat{p}(1,0) = \frac{p(1-b)}{1-pb}$. Combining these two solutions for N_1^1 yields the following condition:

$$\beta(p,b) \equiv pb + \frac{\hat{p}(1,0)b}{1 - \hat{p}(1,0)b} A_2 = S_1 + S_2. \tag{20}$$

Because $\beta(p,b)$ is continuous for all p,b on the unit interval, and strictly increasing in p, there is a unique $p(b)$ that solves (20) for any $b \geq \underline{b}$, where $\underline{b}$ is a sharp lower bound for b given by $\beta(1,\underline{b}) = S_1 + S_2$. This establishes the first part of the claim.

To see that for each b there is a unique couple $\{F(1,1), F(1,0)\}$, note that knowledge of N_t^σ allows us to construct the conditional choice probabilities. Combining this information with observed payoffs $\hat{\pi}(t)$, we have:

$$\Pr(\text{stop}|\sigma_1 = 1) = S_1/N_1^1 \quad = \quad 1 - G(\hat{\pi}(2) - \hat{\pi}(1) - F(1,1)) \tag{21}$$

$$\Pr(\text{stop}|\sigma_1 = 0) = A_1/N_1^0 \quad = \quad 1 - G(\hat{p}(1,0)b\hat{\pi}(2) - F(1,0)). \tag{22}$$

Given a strictly monotone $G(\cdot)$ and $p = p(b)$, these equations can be inverted to find the unique solution for $F(t, \sigma_t)$ as a function of b and the data (A_t, S_t).[30]   Q.E.D.

Theorem 1 says that for any admissible value of b, we can choose the other structural parameters in a manner that rationalizes the observed data. Thus, more information is needed to achieve point-identification. The cost functions are a natural place to look for two reasons. First, we have placed no restrictions on the cost functions up to this point. And second, there are substantive arguments for assuming pre-breakthrough and post-breakthrough equality of the cost functions. For example, the costs of revision may primarily be opportunity costs. Alternatively, breakthroughs may depend on the "fundamental" quality of the underlying idea, whereas revisions—both before and after a breakthrough—improve other dimensions of quality, such as exposition and generality. These observations lead us to make:

Assumption 1. *Revision costs are independent of σ_t, $F(t, 1) = F(t, 0) = F(t)$,*

which leads immediately to the following result:

Theorem 2. *Let $G(\cdot)$ be the logistic cdf, and make Assumption 1. If $p \geq 1 - A_1^2$, then there exists a unique solution to Equations (20), (21), and (22).*

Appendix B proves this theorem, and then gives the conditions under which the result extends to the $\overline{T}$-period model.[31] We provide some intuition in what follows. Solving Equations (21) and (22) for F and setting the two expressions equal (based on Assumption 1) leads to the following condition:

$$b\hat{p}(1, 0) + \frac{\log(\frac{N_1^1 - S_1}{S_1}) - \log(\frac{N_1^0 - A_1}{A_1})}{\hat{\pi}(2)} = \frac{\hat{\pi}(2) - \hat{\pi}(1)}{\hat{\pi}(2)}. \tag{23}$$

---

[30]In Claim B.3, Appendix B, we provide the proof for any $\overline{T} \geq 2$.
[31]Specifically, Theorem 2 corresponds to Claims B.4 and B.5. The general $\overline{T}$-period case is considered in Claim B.6.

The right side of this equation is a constant, and the conditions in Theorem 2 are sufficient to establish that as b increases, the expression on the left side crosses $\frac{\hat{\pi}(2)-\hat{\pi}(1)}{\hat{\pi}(2)}$ from below exactly once. If b is large, however, the left side of (23) can be non-monotonic, possibly leading to multiple solutions. The economic intuition is that when b is large, absence of breakthrough leads to pessimism, causing players to quit. But high costs can also lead to quitting. For sufficiently high b, there may be solutions that exhibit higher b with lower p and F, and vice versa. But for all of the estimates we report below, the parameter estimates are consistent with a unique solution.[32] Moreover, while the model can be identified by assuming that costs are equal in a single period, in practice we assume that Assumption 1 holds for all t and exploit the over-identifying restrictions for estimation.[33]

Although our model is identified under Assumption 1, some readers may prefer more clarity regarding the specific source of variation that identifies the costs relative to the learning parameters. This leads us to consider a second identification strategy that exploits nonstandards-track RFCs as a unique institutional feature of the IETF standardization process. Specifically, we estimate a series of models that make

Assumption 2. *Nonstandards-track projects have the same cost function $F(t)$ as standards-track projects, but are always published (i.e., $p = b = 1$).*

Given this second assumption, we can estimate $\boldsymbol{F}(b)$ in an initial step (relying only on the nonstandards-track proposals), and then feed the resulting cost estimates into the likelihood function to estimate b and p. This approach has costs and benefits relative to relying exclusively on Assumption 1. The primary benefit is clarifying the relationship between data and parameter estimates. Under Assumption 2,

---

[32]Although the sufficient condition in Theorem 2 is not satisfied by our estimates, the estimates do satisfy a necessary condition provided in Appendix B. That condition requires that $p \geq (2b-1)/b^2$.

[33]In one robustness check of the empirical analysis we assume equality up to a constant, so $F(t,1) = F(t,0) + \kappa$.

the cost function is identified by the payoffs and distribution of stopping times for nonstandards-track RFCs, while b and p are identified by the payoffs and stopping times for the standards-track. The obvious cost is the use of additional (strong) assumptions about the relationship between the two tracks.

On a substantive level, we think Assumption 2 is reasonable. Recall that we interpret breakthrough as the arrival of a consensus. In Section 2, we described how standards- and nonstandards-track proposals are very similar in terms of content and publication process, but have rather different commercial implications. The similar substance justifies our assumption that cost functions are equal across the two tracks (e.g., see Figure 1). The different commercial implications justify our assumption that $p = b = 1$ for nonstandards. While standards-track RFCs provide a focal point for implementation in products, nonstandards-track RFCs are classified as Informational (i.e., complements to Proposed Standards) or Experimental (i.e., protocols that are not mature enough to merit a full recommendation). Because nonstandards do not produce winners and losers, IETF participants have few incentives to delay or prevent their success, and consensus is more or less automatic. For example, Figure 2 shows that some nonstandards-track projects are published immediately, without revision.

3.6 Estimation

With data, model, and identification strategy in hand, we now consider estimation. Our model is a finite-horizon non-stationary dynamic discrete choice problem. Previous studies have used simulation to evaluate the likelihood for such models (e.g., Pakes, 1986; Fernandez-Villaverde and Rubio-Ramirez, 2007). We have derived an explicit solution that exploits observation of the "hidden" state σ_t at the time when a proposal is either published or abandoned, along with the particularly simple form

of state-transitions in our model.

We make several parametric assumptions when estimating the model. In particular, we set $\overline{T} = 25$, adopt a quadratic specification of revision costs $F(t)$, and assume that the structural cost shocks ε_t have a logistic distribution with a variance of 1. Finally, we assume that expected payoffs (measured in patent citations) are log-linear in the number of versions, as in Equation (1). Given these assumptions, estimation proceeds as follows:

1. Estimate the log-linear patent citation model in Equation (1), and set $\hat{\pi}(t)$ equal to the predicted values for RFCs published in the year 2000. Do this separately for standards and nonstandards.

2. If using Assumption 2, then estimate $\boldsymbol{F}$ for the sample of nonstandards-track RFCs, based on nonstandards-track estimates of $\hat{\pi}(t)$ from Step 1.

3. Iteratively search for values of $(b, p, \boldsymbol{F})$ that maximize the log-likelihood, where each iteration consists of two steps:

   (a) Starting in period $\overline{T}$, recursively compute the sequence of cut-points $\{\bar{\varepsilon}_t^\sigma\}_{t=1}^{\overline{T}-1}$, along with the associated probabilities $G^\sigma(t)$ and continuation values $V(t, \sigma)$.

   (b) Form $LL(b, p, \boldsymbol{F})$, retaining estimates of $\boldsymbol{F}$ from Step 2 if using Assumption 2.

The first step in this procedure yields estimates of agents' beliefs. That is, we assume authors know $\hat{\pi}(t)$ and decide whether to revise a proposal given that knowledge. This leads to a question about what sample we should use to estimate beliefs. A rational expectations assumption suggests using the realized distribution of citations for projects in the estimation sample. On the other hand, if the author-teams have adaptive expectations, we might use a sample of RFCs published before the project is

29

started (in principle, this sample could be different for each observation). In practice, we find that our citation model estimates are relatively stable, and so use the realized citation distribution for projects in the estimation sample below.

We refer to the approach that uses nonstandards-track RFCs and Assumption 2 as the three-step estimator, and the approach that relies only on Assumption 1 (skipping over Step 2) as the two-step estimator. Our main estimates are based on the two-step approach, and for those models we compute 95% confidence intervals based on 1,000 bootstrap replications.

4   Empirical Results

This section presents and discusses parameter estimates. Our baseline specification uses the two-step estimator for the full sample of projects submitted from 1996 through 2009.[34] We perform a variety of robustness checks, based on different samples and assumptions, and conclude by discussing further analyses (presented in Appendix D) that add parametric heterogeneity to the learning process through p and b.

4.1   Baseline Estimates

Table 2 presents the estimates for our baseline model and a variety of robustness checks. The top panel reports the learning parameters b and p, and the bottom panel reports the estimated costs of a revision (measured in patent citations) at $t = 1$, $t = 10$, and $t = 20$. For each parameter, we report either a 95% bootstrapped confidence interval [in brackets] or a standard error based on the estimated covariance matrix

---

[34]Limiting our sample to projects initiated until (and including) 2009 avoids issues of right-censoring. Eliminating all projects initiated before 1996, on the other hand, may introduce left-censoring. However, the IETF was relatively small and young at that time, so our results can be interpreted as applying to the "mature" IETF.

(in parentheses). The first five columns employ a two-step estimation approach based on Assumption 1. The last two columns are from a three-step estimator that adds Assumption 2.

The results for our baseline model are in Column (1). Ex ante, IETF members anticipate that three out of five projects have the potential to achieve a breakthrough, based on our estimate of $p = 0.59$. Conditional on having a good idea, the estimated breakthrough arrival rate is $b = 0.17$. The bottom panel shows that the implied costs of revision are decreasing and convex. We interpret these fitted values as opportunity costs, including the foregone value of closing a given project and submitting a first version of another project. Our results then suggest that switching to another project is less attractive as the number of revisions increases.[35] To our knowledge, these are the first structural estimates of the parameters guiding learning in two-armed bandit experimentation models.[36]

One way to understand the magnitude of these estimates is to consider the share of good projects that actually get published, given the speed of the learning process and the (stochastic) opportunity costs of continuing a line of research. For our baseline parameter estimates, the "success rate" for good projects is 31%. We believe this statistic provides some sense of the real economic costs of information discovery, and the potential benefits from policies that increase the rate of learning.[37]

Figure 5 plots the hazard of publication and abandonment for the raw data (left-

---

[35]Figure 1 provides a possible explanation (outside our model) for the estimated shape of the cost function. The graphs in the figure capture a decreasing rate of textual changes: later versions are increasingly different from the initial draft, but develop in this way at a decreasing rate. If smaller textual changes come at lower cost because less effort is exerted (if, for instance, later revisions are primarily minor edits whereas earlier revisions reflect more substantive changes in the text), then the depicted pattern in Figure 1 comports with the estimated costs in Table 2.

[36]Bray et al. (2016) study the choice between competing arms with a quasi-experimental approach in the context of the Italian judiciary.

[37]Appendix Figure D.1 plots *ex-ante* project value (at the time of the initial draft and before the first continuation decision is made) against b to illustrate the benefits of faster learning.

Table 2: Baseline and Robustness Results

	2-Step Estimator					3-Step Estimator	
Specification	Full Sample (1)	WG Sample (2)	Area Specific (3)	Calendar Time (4)	Phase Specific (5)	Full Sample (6)	WG Sample (7)
Parameter Estimates b and p (Standards-Track Projects)							
Rate of Learning: b	0.17	0.35	0.10	0.20	0.26	0.30	0.35
	[0.16, 0.20]	[0.32, 0.37]	(0.003)	(0.003)	(0.006)	[0.27, 0.34]	[0.28, 0.99]
Quality Prior: p	0.59	0.75	0.59	0.49	0.42	0.45	0.32
	[0.51, 0.73]	[0.65, 0.84]	(0.008)	(0.003)	(0.007)	[0.28, 0.59]	[0.30, 0.53]
log-likelihood/Project	-2.138	-3.145	-2.070	-2.175	-2.124	-2.383	-5.711
Projects (standards-track)	14,444	3,168	14,444	14,444	14,444	14,444	3,168
Cost Estimates (Standards-Track Projects; Nonstandards-Track Projects in Models (6) and (7))							
Costs at $t = 1$	2.35	4.44	1.34	2.72	2.20	3.10	2.18
	[1.95, 3.25]	[3.19, 6.13]	(0.015)	(0.029)	(0.028)	[2.03, 4.06]	[0.68, 3.73]
Costs at $t = 10$	1.25	1.44	1.08	0.63	0.58	0.71	0.71
	[1.12, 1.81]	[1.13, 2.06]	(0.010)	(0.005)	(0.023)	[0.56, 0.86]	[0.48, 0.96]
Costs at $t = 20$	0.51	0.77	0.46	0.37	0.07	0.67	0.52
	[0.48, 0.65]	[0.63, 0.98]	(0.015)	(0.009)	(0.010)	[0.50, 0.85]	[0.27, 0.83]
Post-Breakthrough Cost Shift κ					0.59		
					(0.019)		
log-likelihood/Project						-1.189	-1.074
Projects (nonstandards-track)						1,647	764

Notes: This table presents ML estimates for the baseline model and results from a set of sensitivity analyses. In Columns (1) and (2), we report full-sample and WG-sample results for our 2-step estimator. In Columns (3), (4), and (5), we report full-sample results for our 2-step estimator for the model versions with area-specific patent citations, calendar time, and breakthrough-phase-specific costs. In Columns (6) and (7), we report full-sample and WG-sample results for our 3-step estimator. For these latter models, the top panel presents estimates for b and p from Step 3 of our estimation procedure (using data on standards-track projects), given the cost estimates from Step 2 in the bottom panel (using data on nonstandards-track projects). Estimates for project values $\hat{\pi}(t)$ are the predicted values from the citation model in Equation (1). For the models in Columns (1), (2), (6), and (7), we report in brackets the 95% confidence interval from 1,000 bootstrap replications. The reported coefficient is the coefficient from a single ML estimation on the complete sample. For the other models, we report coefficient and standard errors (in parentheses) from a single ML estimation on the complete sample; standard errors for cost estimates are calculated using the delta method. For the model in Column (5), the costs for $t = 1$, $t = 10$, and $t = 20$ are for the pre-breakthrough phase; post-breakthrough costs are increased by the amount of the post-breakthrough cost shift κ.

Figure 5: Hazard Rates – Data and Simulation
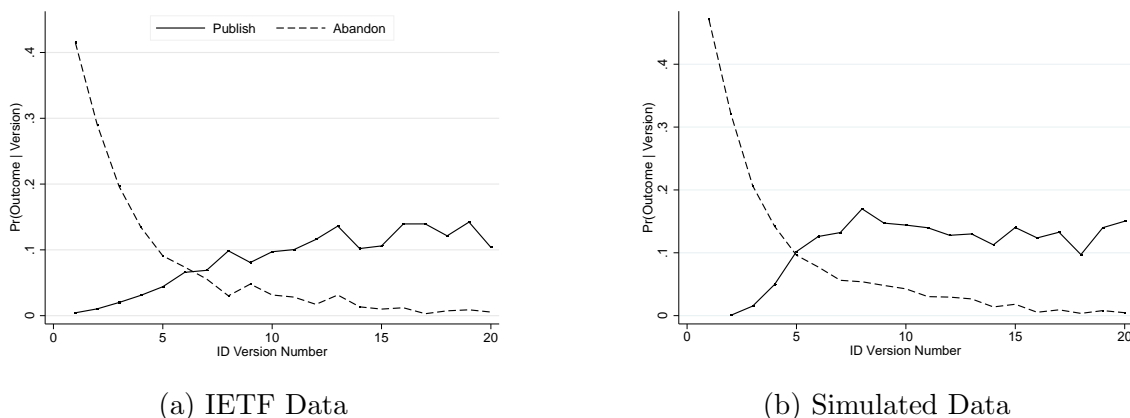
(a) IETF Data	(b) Simulated Data

Notes: Panel (a) plots the hazard rates for the published (solid line) and abandoned (dashed line) projects (versions 1 through 20) for the IETF data sample of 14,444 standards-track projects. See also Figure 2. Panel (b) plots the hazard rates for published and abandoned projects for a simulated sample of 14,444 projects, with the values for parameters b, p, and $\boldsymbol{F}$ from the model in Column (1) of Table 2.

hand panel) and for a simulated data set based on the parameter estimates from our baseline model (right-hand panel). Although the actual publication hazard exhibits more steady growth between $t = 1$ and $t = 13$, the model can apparently capture the basic features of the IETF data reasonably well.

4.2 Robustness Analysis

Columns (2) through (7) in Table 2 present the results for additional models that relax or change our baseline assumptions. The results generally show that estimates from our preferred baseline model are consistent across alternative specifications and sub-samples, at least to the degree of precision afforded by the data. Further robustness checks and sensitivity analyses are reported in Appendix A.[38]

---

[38] Figures A.1 and A.2 plot the implied hazard rates for additional models in Table 2. Columns (1), (2), and (6) in Table A.2 correspond to Columns (1), (6), and (7) in Table 2. All models in Table A.2 use an extended sample (including RFCs initiated between 2010 and 2015) to estimate and predict patent citations $\hat{\pi}(t)$. All but the first model in the table use the three-step estimation approach. Comparisons (within Table A.2) of the alternative models in Columns (3) through (6) are therefore with the model in Column (2) and Column (8) with Column (7). Further descriptions can be found

4.2.1 Subsample of Working Group Projects ("WG Sample")

We begin by evaluating estimates for the sub-sample of IETF projects initiated within a working group.[39] While anyone can start a project outside of a working group, the threshold for the working group to initiate a project is expected to be higher, leading to the higher publication rates that we observed in Table 1.

Column (2) of Table 2 presents estimates for the WG sample. Both learning parameters, b and p, increase relative to the full sample (though we cannot reject the hypothesis that p is equal in the two samples). The estimated costs of revision are initially higher than for the full sample, but decline faster, and achieve a comparable level by $t = 10$. One possible explanation for higher point estimates of b and p is that the additional attention and feedback from the working group leads to a faster consensus.[40]

4.2.2 Heterogeneity of Project Values ("Area Specific")

In Column (3) of Table 2, we introduce ex-ante heterogeneity in the expected payoffs. That is, we continue to assume that players have shared expectations regarding $\hat{\pi}(t)$ but allow those expectations to differ across the seven broad Technology Areas defined by the IETF.[41] To implement this idea, we add technology-area fixed effects to our model of expected citations (Step 1), and use the area-specific estimates of $\hat{\pi}(t)$

---

in the table notes.

[39] As we stress in the next subsection, an important property of this subsample of projects is that we cannot reject the hypothesis that the distance measures featured by nonstandards-track projects are not statistically distinguishable from the one of standards-track projects. This result supports the assumption that the two types of projects feature comparable fixed costs.

[40] We provide some additional support for this explanation in Section 4.3 below, when we consider the impact of communication (as a proxy for attention) on learning.

[41] The IETF technology areas correspond roughly to the various layers in the engineering "protocol stack" as described in Simcoe (2012). From top to bottom, those layers/areas are: Applications, Realtime Applications and Infrastructure, Transport, Internet, and Routing. The IETF also recognizes two areas that cut across the various layers: Operations and Security. For more details on technology areas, see the data appendix (in Appendix E).

when evaluating the likelihood. The estimates in Column (3) of Table 2 show that accounting for ex-ante technological heterogeneity in expected project values has a mixed effect on the estimates for b and p. While the point estimates for b are lower and outside the confidence interval in Column (1), the point estimates for p fall well within the interval.[42] Thus, adding technological heterogeneity in (rationally expected) payoffs does not change the priors about project types but suggests that learning is slower than indicated by our baseline model.

4.2.3 Definition of Time ("Calendar Time")

Our baseline specification assumes that costs are determined by the number of revisions t, and does not include discounting for time. In reality, some teams submit revisions faster than others. If this reflects variation in the amount of revision, so that faster resubmissions imply less substantive change, our model would fail to measure the costs of revision accurately. To address this concern, we estimate a version of the model where the data are re-shaped so that t is measured in calendar-quarters instead of proposal revisions. We set $\overline{T}$ equal to 6 years (24 quarters) and assume that a project is abandoned in the quarter of its last version if that revision is followed by 2 years of inactivity.

Column (4) of Table 2 presents the results of the calendar-time model. In general, the p parameter should not vary with our choice of time-units, while b should change. Because the average time between versions is roughly three months, however, we expect b to have a similar magnitude to the version-time results. In practice, the calendar-time estimate of $p = 0.49$ falls just below the lower bound of the 95% confidence interval in the baseline model, and the two estimates for b are statistically

---

[42]We switch from bootstrap to ordinary confidence intervals at this point, because adding heterogeneity to the model raises the computational costs associated with the bootstrap computations.

indistinguishable.

4.2.4   Pre- and Post-Breakthrough Specific Costs ("Phase Specific")

For the model in Column (5) of Table 2, we relax Assumption 1 by allowing the pre- and post-breakthrough revision costs to differ by a constant κ, so that $F(t,1) = F(t,0) + \kappa$. This specification is meant to address concerns that the cost of "polishing" (post-breakthrough) is small or even negligible compared to the costs of "experimentation" (pre-breakthrough). Contrary to this hypothesis, our estimates suggest that the costs of revision increase after a breakthrough, with $\kappa = 0.59$. In this specification, p is a bit lower than the baseline estimates, b is somewhat higher, and the costs (though not directly comparable) appear similar.

4.2.5   Three-Step Estimation

The last two models in Table 2 present the results for our three-step estimation approach. For this, we use Assumption 2 and estimate the costs (in Step 2) using only data on nonstandards-track projects while setting $p = b = 1$. In Step 3, we take the results from Step 2 and estimate b and p using data on standards-track projects. We present bootstrapped 95% confidence intervals in brackets.

The bottom panel of the table shows that using the nonstandards-track RFCs to estimate $F(t)$ leads to a cost function that starts higher, and declines more steeply than in our baseline model. Given this cost function, estimates of p are somewhat smaller, though not statistically different, from our baseline model. The estimated breakthrough arrival rate, however, almost doubles. Interestingly, using Assumption 2 also produces estimates of b and p that are more similar across the full and WG samples, shown in Columns (6) and (7), respectively.

4.3 Parametric Heterogeneity

Until now, we have treated projects as ex-ante identical (with the exception of the area-specific model in the third column of Table 2). It is straightforward, however, to incorporate some heterogeneity into the model by specifying parameters p, b, or $F(t)$ as functions of observable characteristics. We now briefly summarize a variety of specifications that explore heterogeneity in the rate of learning, with full details provided in Appendix D.

To study heterogeneity in learning, we let $b \equiv b(x)$, where x measures four different project and author-team characteristics: communication, author experience, commerciality, and team size. Communication is the number of emails per revision that specifically mention a focal project. Commerciality is the share of those emails that originate from a dot-com domain. Experience is the maximum number of prior IETF submissions by an author-team member, and team size is an author count. Note that our proxies for experience and team size are pre-determined, whereas the proxies for communication and commerciality are endogenous to the revision process because they are based on emails sent during the revision process. We provide more details on how we construct all of these explanatory variables in Appendix D.

While heterogeneity gives rise to differential patterns in terms of authors' rate of learning, the qualitative features of our results remain similar to the ones discussed in the baseline analysis of Section 4. The findings for $b(x)$ are broadly consistent with our prior expectations about the learning process. We find that more communication increases the estimated value for b. We also find that b increases with the size and experience of author-teams. Finally, we find that commerciality, as proxied by the share of corporate IETF members who contribute to the discussion of a project, has a non-monotonic impact on learning—implying a lower speed of learning (b) for large

values of the ratio.

5 Counterfactual Analysis

We now consider two counterfactual experiments based on estimates from our baseline model (i.e., the first column of Table 2). The first experiment compares two policies meant to stimulate research: a subsidy that lowers the cost of revisions, and a prize for publication of an RFC. The second experiment examines the costs of misaligned priors by simulating an over-/under-confident team that believes p is above/below its true value.

5.1 Prizes and Subsidies

Subsidies and prizes are both used as innovation policy instruments.[43] Examples include the R&D tax credit and the patent system. Much of the theoretical literature on optimal innovation rewards considers trade-offs between *ex-ante* incentive and *ex-post* market power when the prize is a patent (e.g., Gilbert and Shapiro, 1990; Hopenhayn et al., 2006). We use our model to study a different question: how do prizes and subsidies influence the decision to continue exploring a risky idea? Specifically, we simulate the impact of prizes and subsidies on both research output and the ex-ante value of a project, holding fixed the total budget allocated to each policy.

We model "prizes" as a small increase in publication payoffs, δ_p, so the benefits of an RFC become $\hat{\pi}(t) + \delta_p$. Each project's scrap value remains zero. We model subsidies as a small decrease in revision costs, δ_s, leading to the new cost function

---

[43]We use the term *prize* for a contingent benefit paid on successful completion of a project. In the literature, the term is also used for a reward paid to the winner of a contest (e.g., Murray et al., 2012; Galasso et al., 2016).

Figure 6: Prizes and Subsidies

(a) Additional Citations

(b) Project Value
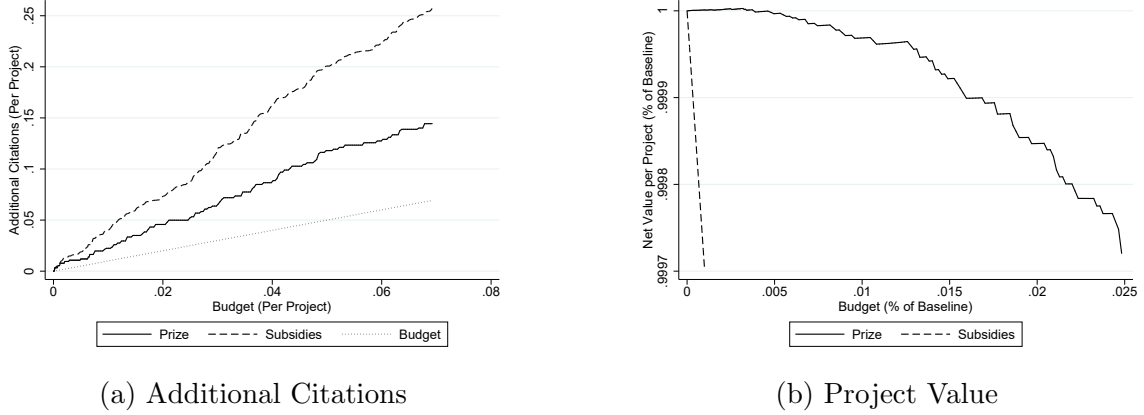
Notes: This figure depicts results of our counterfactual analysis in which we introduce (1) a prize δ_p for a published RFC and (2) cost subsidies (per version) δ_s. We simulate $N = 14,444$ projects (our estimation sample size) for varying levels of a fixed budget, using parameter estimates from the baseline model in Column (1) of Table 2. Varying the fixed budget, we determine the respective publication prize δ_p and the respective per-period subsidies δ_s such that the total expenditure does not exceed the budget. Panel (a) plots the number of additional citations (per project) stemming from the introduction of a publication prize (solid) or per-version subsidies (dashed), for all projects. On the x-axis, we have the per-project budget. Panel (b) plots the total net value of projects, measured as expected (ex-ante) per-project value as percentage of the baseline ex-ante project value (where "baseline" means without prize or subsidies). On the x-axis, we have the budget as percentage of the baseline ex-ante value.

$F(t) - \delta_s$. To compare the two policies, we vary either δ_p or δ_s (holding the other structural parameters constant), and simulate outcomes in a sample of N projects. For each simulation, we calculate the total cost of the policy along with private costs and benefits, all measured in terms of expected patent citations.

Figure 6 shows the results of our counterfactual experiment. Panel (a) plots the increase in gross citations as a function of the policy's expenditure. Panel (b) plots the total expected value of a project (i.e., citations net of budget and private opportunity costs, expressed as share of the pre-policy baseline) against the expenditure (also expressed as a share of baseline).

From Panel (a) we observe that both prizes and subsidies lead to an increase in research output, with an elasticity larger than unity. Moreover, for a given budget, the subsidy produces more new citations than the prize. This picture changes, however, when we examine Panel (b), which shows that after accounting for researchers'

39

opportunity cost, subsidies produce a decline in the *ex-ante* value of a project. Prizes smaller than 2% of the pre-policy expected value of an idea can increase the expected value of a project, but larger prizes do not.[44]

Why do the two policies produce such different results? Both prizes and subsidies provide incentives for pre-breakthrough exploration. But the prize-based incentive declines over time as researchers grow more pessimistic, while the subsidy-based incentive remains constant. After a breakthrough, prizes have no impact on the decision to revise, whereas subsidies continue to encourage development of the idea. Thus, while both policies provide "early" exploration incentives, only subsidies provide incentives for "late" exploration and further refinement. In this sense, the subsidy creates a larger distortion than the prize. It explains why subsidies generate more new citations in Panel (a), and also why the net benefits of a prize are greater in Panel (b). While the early incentive can lead to more breakthroughs, much of the late incentive is wasted on bad ideas and post-breakthrough refinements that would not be pursued by a team that fully internalized the revision costs.[45]

5.2 Misaligned Priors

A number of researchers have considered the impact of behavioral biases on innovation outcomes. For instance, Allen (1966) suggests that engineers are prone to a type of sunk-cost fallacy and become over-committed to a particular solution. On the

---

[44]In Panel (b), the x-axis corresponds to the budget value in percentage of baseline project net value. All budget values up to 0.37% of baseline value (or equivalently, prizes up to 2% of the ex ante expected value of an idea) increase the expected net value of a project relative to pre-policy values. To clarify our exercise in Panel (b) and the ensuing discussion, consider the following example. Suppose we have 100 projects, and the expected pre-policy value of a project is 10. Suppose the absolute budget is 100. The budget (as a % of baseline values) is then 10%. If, for this budget (and the respective prize), 25 of the 100 projects are published, then, with a total budget of 100, paid to 25 successful projects, the absolute prize is 4, that is, 40% of the pre-policy value of a project.

[45]This result is similar in spirit to Akcigit et al. (2016), who are concerned with the cross-sectional inefficiencies produced by a uniform (cross-industry) subsidy in terms of investment value and research allocation.

other hand, Galasso and Simcoe (2011) and Hirshleifer et al. (2012) find that proxies for CEO over-confidence are positively correlated with corporate innovation. In our model, all research teams have rational expectations and are doing as well as they can. One reason to study learning, however, is that, in reality, no one knows precisely how likely it is that they have a good idea. Our second counterfactual exercise allows for this possibility and studies the cost of misaligned priors, which take two forms: overconfidence and pessimism.

Suppose a team can have biased perceptions of their own skills in choosing or developing new projects. Specifically, if p^* is the true (estimated) probability of a good idea, and p represents the team's subjective beliefs, then an over-confident team has $p > p^*$ while an under-confident (or pessimistic) team has $p < p^*$. In our model, over-confident researchers will pursue unpromising lines of research for too long, and pessimistic teams will cut loose their good ideas too early. To see which behavior is more costly, we vary p—holding p^* and the other structural parameters constant at estimated values—and use simulated choice to calculate the expected payoffs under misaligned priors. The results are shown in Figure 7.

Panel (a) graphs the expected value of a new project as a function of the team's subjective priors. This function achieves its maximum at $p = p^*$, as it must. The interesting feature of Panel (a) is the strong asymmetry in the costs of under- versus over-confidence. Pessimistic teams manage to capture much of a project's expected value, whereas over-confident teams fare much worse.

The intuition for the asymmetry in Panel (a) is made clear in Panel (b), which plots time to both consensus and project completion as a function of p. At low values of p, the average duration is increasing slowly for both good and bad ideas. When p exceeds p^*, average duration increases faster, but the expected duration for good projects levels off because almost all good ideas have been harvested. The average

Figure 7: Over-Confidence and Pessimism

(a) Individual Project Value

(b) Project Length
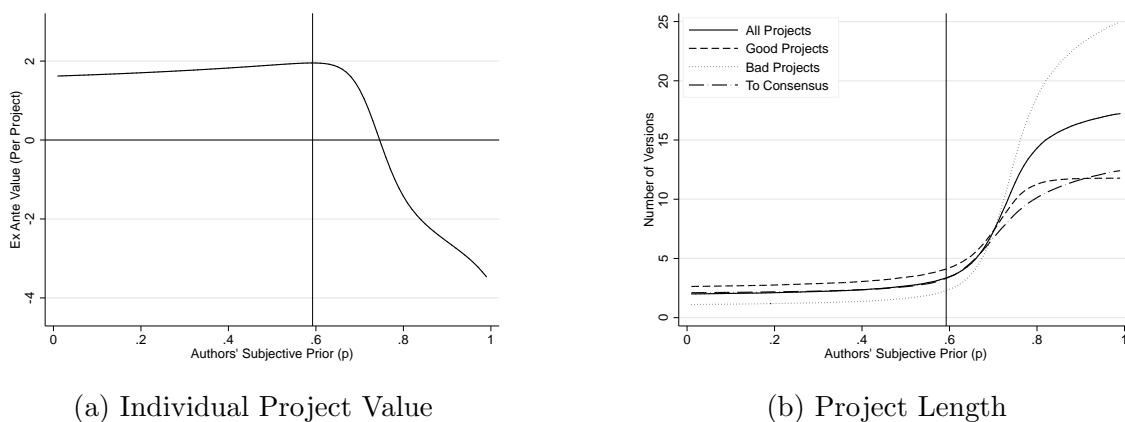
Notes: This figure depicts results of our counterfactual analysis in which we vary the author-team's subjective prior belief p (given a true quality prior of p^*, depicted by the vertical line). Graphs are based on $N = 14{,}444$ simulated standards-track projects (for each $p \in \{0.01, 0.02, \ldots, 0.99\}$) using parameter estimates from the baseline model in Column (1) of Table 2. Panel (a) plots a project's ex-ante value against the author-team's p when the true value is p^*. Panel (b) plots the average project length and the average duration to breakthrough.

duration for bad ideas, on the other hand, continues to increase, eventually surpassing the duration of good projects. In economic terms, pessimistic teams may miss a few opportunities but can still do fine as long as breakthroughs arrive relatively quickly. Over-confident teams are more likely to publish their good ideas but can also persist in costly exploration of "dry wells." In our model, for the parameter estimates obtained using the IETF data, over-confidence is more costly.

6   Concluding Remarks

In this paper, we study a dynamic discrete choice model of innovation in which researchers learn over time about the quality of their ideas. Our model combines a two-armed bandit process wherein researchers face a trade-off between exploration and exploitation, with a traditional optimal stopping problem where they compare the marginal costs and benefits of refining their ideas. In this model, a single failure

does not immediately terminate a line of research, but it does make researchers more pessimistic about the quality of their idea. The resulting process of "learning when to quit" rationalizes the observation that many ideas are abandoned, even though all must (by revealed preference) have a positive expected value when initially pursued. We believe this framework could potentially be applied to study the micro-economics of R&D management in a variety of settings.

We estimate the structural parameters of the model using a unique data set that contains information on every revision of both successful and abandoned projects submitted to the Internet Engineering Task Force (IETF), along with citations to successful (published) standards that we use to calibrate the expected payoffs. In this empirical context, learning is associated with discovery that a proposed standard has achieved "consensus" within the IETF community, and will therefore be published with the group's official recommendation.

Our baseline estimates suggest that 59% of proposals are capable of generating a consensus, and that within that population, about 17% of active projects learn that they will succeed in a given period. While these estimates vary somewhat, they are relatively robust to a wide range of alternative assumptions and measurement strategies. At this rate of learning, the model implies that around 31% of the "good ideas" (i.e., proposals that might lead to consensus) are ultimately published, while over two-thirds are abandoned due to a combination of uncertainty and opportunity cost.

We use the data, model, and parameter estimates to perform a pair of counterfactual experiments. One counterfactual compares the innovation-promoting effects of a *prize* that awards more citations to a successful project and a *subsidy* that lowers the costs of research. We find that both schemes can produce a net increase in innovation, but that, for a given budget, the subsidy produces a larger increase. However, if one

43

also accounts for the private costs, which include providing incentives for researchers to continue pursuing low-value projects, only the prize might actually increase total welfare. The second counterfactual examines the costs of misaligned priors, and shows that over-confidence would be more costly than pessimism in this setting.

To our knowledge, this is first paper to estimate a structural model of learning in the context of R&D. In practice, however, standardization combines elements of collaborative R&D with competition to benefit from adopting one's own technology. One direction for future research might be to move beyond our reduced-form characterization of the non-cooperative bargaining elements of standardization towards a dynamic-game model. Another opportunity is to apply this type of framework to other settings, such as scientific review and publication, or open-source software development. A key challenge for that agenda will be finding other data sets that contain information on both successful and abandoned ideas, as well as the underlying revision process.

Appendix

A   Additional Tables and Figures

Table A.1: Examples of IETF Internet Standards

	Description	RFC	Year
RTP	Real-time Transport Protocol	3550	2003
SIP	Session Initiation Protocol	3261	2002
HTTP	Hypertext Transfer Protocol – HTTP/1.1	2616	1999
IPV6	Internet Protocol, Version 6 (IPv6) Specification	2460	1998
DHCP	Dynamic Host Configuration Protocol	2131	1997
POP3	Post Office Protocol – Version 3	1939	1996
NAT	Network Address Translator	1631	1994
FTP	File Transfer Protocol	959	1985
TCP	Transmission Control Protocol	793	1981
IP	Internet Protocol	791	1981

Notes: RFC means "Request for Comments," Year is the year in which the standard was certified by the IETF.

Figure A.1: Hazard Rates – Data and Simulation (More Results)
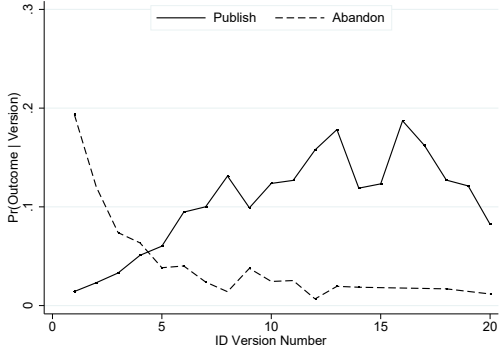
(a) IETF Data: WG Sample
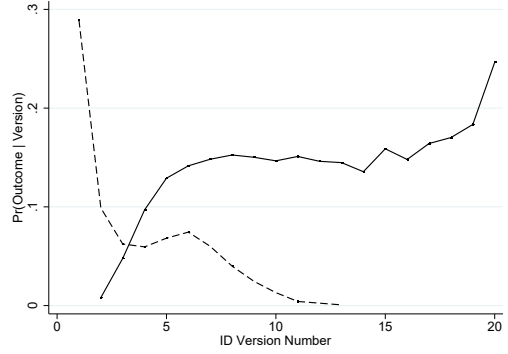
(b) Simulation: WG Sample
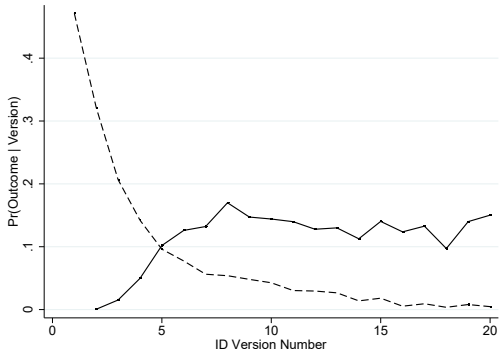
(c) Simulation: Baseline
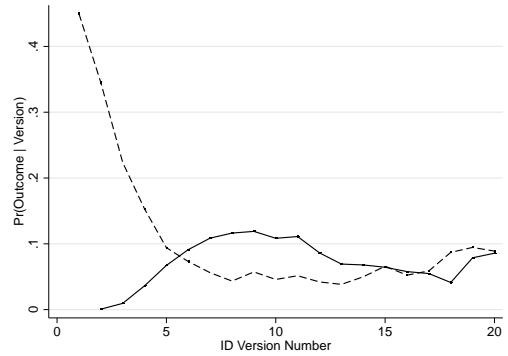
(d) Simulation: Area Specific
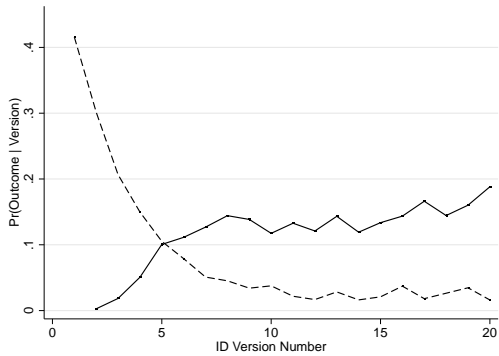
(e) Simulation: Phase Specific (κ)
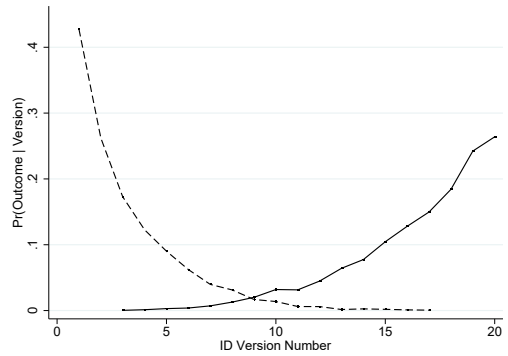
(f) Simulation: 3-Step Estimator

Notes: Panel (a) plots the hazard rates for the published (solid line) and abandoned (dashed line) projects (versions 1 through 20) for the IETF data sample of 3,168 standards-track projects (WG sample). Panel (b) plots the hazard rates for published and abandoned projects for a simulated sample of 3,168 projects, with the values for parameters b, p, and $\boldsymbol{F}$ from the model in Column (2) of Table 2. Panels (c), (d), (e), and (f) plot the hazard rates from simulated samples for the baseline model in Column (1) and the models in Columns (3), (5), and (6) in Table 2.

46

Figure A.2: Hazard Rates – Data and Simulation (Calendar Time)
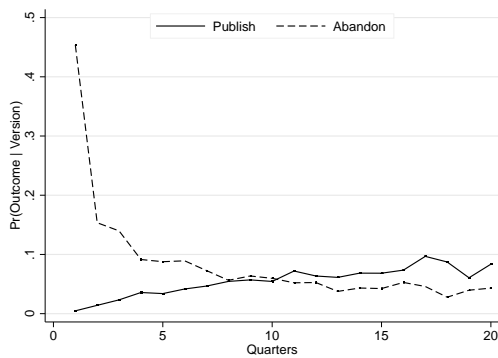
(a) IETF Data
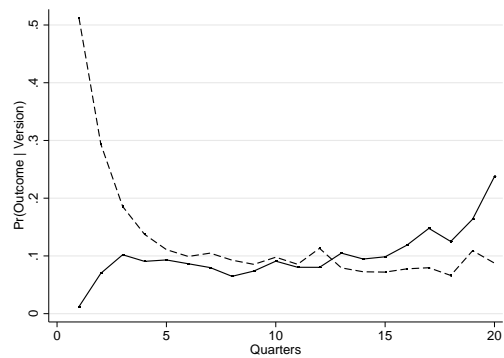
(b) Simulated Data

Notes: Panel (a) plots the hazard rates for the published (solid line) and abandoned (dashed line) projects (quarters 1 through 20) for the IETF data sample of 14,444 standards-track projects. Panel (b) plots the hazard rates for published and abandoned projects for a simulated sample of 14,444 projects, with the values for parameters b, p, and F from the model in Column (4) of Table 2.

Table A.2: More Robustness Results

Specification	Baseline 2-Step (1)	Baseline 3-Step (2)	Only Experimental (3)	Censored Projects (4)	RFC Citations (5)	Deadline $\overline{T}=50$ (6)	WG 3-Step (7)	First Project (8)
Step 3: Parameter Estimates b and p (Standards-Track Projects)								
Rate of Learning: b	0.18	0.28	0.33	0.31	0.30	0.23	0.34	0.25
	(0.004)	(0.004)	(0.002)	(0.004)	(0.008)	(0.004)	(0.011)	(0.026)
Quality Prior: p	0.51	0.39	0.54	0.39	0.29	0.37	0.43	0.44
	(0.004)	(0.004)	(0.002)	(0.004)	(0.008)	(0.004)	(0.011)	(0.026)
log-likelihood/Project	-2.111	-2.231	-2.242	-1.992	-2.106	-2.274	-3.533	-3.508
Projects (standards-track)	14,444	14,444	14,444	21,779	14,444	14,444	3,168	226
Step 2: Cost Estimates (Nonstandards-Track Projects)								
Costs at $t=1$	2.25	2.58	3.79	2.52	1.55	2.15	2.12	1.95
	(0.026)	(0.053)	(0.151)	(0.050)	(0.047)	(0.037)	(0.081)	(0.175)
Costs at $t=10$	0.98	0.63	0.77	0.62	0.48	0.71	0.70	0.71
	(0.007)	(0.008)	(0.019)	(0.008)	(0.009)	(0.007)	(0.015)	(0.031)
Costs at $t=20$	0.46	0.59	0.62	0.61	0.44	0.39	0.51	0.56
	(0.012)	(0.031)	(0.051)	(0.029)	(0.029)	(0.008)	(0.043)	(0.106)
Costs at $t=30$						0.43		
						(0.024)		
log-likelihood/Project		-1.176	-1.271	-1.148	-1.153	-1.235	-1.072	-0.980
Projects (nonstandards-track)		1,647	262	1,999	1,647	1,647	764	179

Notes: This table presents ML estimates for the baseline models in Column (1) (using the two-step estimation approach) and Columns (2) and (7) (using the three-step estimation approach; for the full sample and the WG sample) and the results from further sensitivity analyses in Columns (3) through (6) and Column (8). In (3) ("Only Experimental"), we use only Experimental RFCs for nonstandards-track projects; in (4) ("Censored Projects"), we include active projects and those completed projects that were initiated in or after 2010; in (5) ("RFC Citations"), we use citations by other RFCs for project values estimates $\hat{\pi}(t)$; in (6) ("Deadline $\overline{T}=50$"), we extend the forced deadline to $\overline{T}=50$; in (8) ("First Project"), we consider only working group projects that were the first project of their respective working group. For models (2) through (8), the top panel presents estimates for b and p from the third step of our estimation procedure (using data on standards-track projects), given the cost estimates from the second step in the bottom panel (using data on nonstandards-track projects). Estimates for project values $\hat{\pi}(t)$ are based on the extended sample, including RFCs initiated in year 2010 or after. Standard errors for the parameter estimates are reported in parentheses. Standard errors for cost estimates are calculated using the delta method.

B Model Identification

In this appendix, we show the conditions under which we achieve point-identification of the parameters in our empirical model. We proceed in two steps. We first show that, without further restrictions, the statistical features of the data, combined with our model, are not enough to uniquely identify all our parameters. We then make an assumption on the fixed costs that renders point-identification feasible.

B.1 No Further Restrictions

We consider a recursive solution of the empirical model with finite-time horizon $\overline{T}$, given $p \in (0,1)$, $b \in (0,1)$, and $G(\varepsilon)$ strictly monotone. Moreover, to ease notation, we let $\hat{p}(t) \equiv \hat{p}(t,0)$ represent posterior beliefs in period t (before a breakthrough). The number of parameters to identify is $2 \times \overline{T}$: p, b and $F(t, \sigma_t)$, with $t = 1, \ldots, \overline{T} - 1$, and $\sigma_t = 0, 1$. We start by constructing the expressions for the choice probabilities (of publication and abandonment) in each t. Then, we analyze whether these expressions are sufficient to identify all the parameters in our model.

For the construction of the choice probabilities, we compute the number of projects in the pre- and post-breakthrough phase in each stage t of the model. We start with the count of projects in the pre-breakthrough phase. Recall that we have normalized the total number of projects to one so that $N_t^0 + N_t^1 = 1$, where N_t^0 and N_t^1 have been defined in Section 3.5.

Claim B.1. *The number of projects in the pre-breakthrough phase in any $t \in \left[1, \overline{T} - 1 \right]$ is given by*

$$N_t^0 = A_t + \sum_{j=t}^{\overline{T}-1} \frac{A_{j+1}}{\prod_{k=t}^{j} \left(1 - \hat{p}(k)b \right)}. \tag{B.1}$$

Proof. To derive (B.1), we proceed recursively, starting from $\overline{T} - 1$:

$$
\begin{aligned}
N_{\overline{T}-1}^0 &= A_{\overline{T}-1} + A_{\overline{T}} \frac{1}{1 - \hat{p}(\overline{T} - 1)b} \\
N_{\overline{T}-2}^0 &= A_{\overline{T}-2} + N_0^{\overline{T}-1} \frac{1}{1 - \hat{p}(\overline{T} - 2)b} \\
&= A_{\overline{T}-2} + \frac{A_{\overline{T}-1}}{1 - \hat{p}(\overline{T} - 2)b} + \frac{A_{\overline{T}}}{\left[1 - \hat{p}(\overline{T} - 2)b \right] \left[1 - \hat{p}(\overline{T} - 1)b \right]}.
\end{aligned}
$$

By iteration of this rule, we obtain Equation (B.1) in the claim.   Q.E.D.

We now define the number of projects in the post-breakthrough phase in each stage t.

Claim B.2. *The number of projects in the post-breakthrough phase in any $t \in \left[1, \overline{T} - 1\right]$ is given by*

$$N_t^1 = \sum_{j=t}^{\overline{T}} S_j - \sum_{j=t}^{\overline{T}-2} N_{j+1}^0 \frac{\hat{p}(j)b}{1 - \hat{p}(j)b} - A_{\overline{T}} \frac{\hat{p}(\overline{T}-1)b}{1 - \hat{p}(\overline{T}-1)b}. \tag{B.2}$$

Proof. For the derivation of (B.2), we proceed recursively, starting from $\overline{T} - 1$:

$$
\begin{aligned}
N_{\overline{T}-1}^1 &= S_{\overline{T}} + S_{\overline{T}-1} - A_{\overline{T}} \frac{\hat{p}(\overline{T}-1)b}{1 - \hat{p}(\overline{T}-1)b} \\
N_{\overline{T}-2}^1 &= S_{\overline{T}-2} + N_{\overline{T}-1}^1 - N_{\overline{T}-1}^0 \frac{\hat{p}(\overline{T}-2)b}{1 - \hat{p}(\overline{T}-2)b} \\
N_{\overline{T}-3}^1 &= S_{\overline{T}-3} + N_{\overline{T}-2}^1 - N_{\overline{T}-2}^0 \frac{\hat{p}(\overline{T}-3)b}{1 - \hat{p}(\overline{T}-3)b},
\end{aligned}
$$

which, after rearranging, yields

$$N_{\overline{T}-3}^1 = \sum_{j=\overline{T}-3}^{\overline{T}} S_j - \sum_{j=\overline{T}-3}^{\overline{T}-2} N_{j+1}^0 \frac{\hat{p}(j)b}{1 - \hat{p}(j)b} - A_T \frac{\hat{p}(T-1)b}{1 - \hat{p}(T-1)b}.$$

By iteration of this process, we obtain Equation (B.2) in the claim. Q.E.D.

The formulas for $N_t^{\sigma_t}$, $\sigma_t = 0, 1$, allow us to construct the $2\left(\overline{T} - 1\right)$ expressions for the conditional choice probabilities in each t:

$$\frac{S_t}{N_t^1} = 1 - G(\hat{\pi}(t+1) - \hat{\pi}(t) - F(t,1)) \tag{B.3}$$

$$\frac{A_t}{N_t^0} = 1 - G(\hat{p}(t)b\hat{\pi}(t+1) - F(t,0)), \tag{B.4}$$

with $t = 1, \ldots, \overline{T} - 1$. Given any guess for b and p and the strict monotonicity assumption on $G(\cdot)$, the expressions in (B.3) and (B.4) can be inverted to solve for the costs in each t. That is, for each t, $F(t,1)$ and $F(t,0)$ are uniquely identified by (B.3) and (B.4), yielding expressions that depend on p, b and data.

To explore identification of p and b, we start exploiting the additional condition stemming from the property that, by construction, $pb = N_1^1$. We then use (B.2) to compute the number of post-breakthrough phase projects in $t = 1$, so that

$$pb = N_1^1 = \sum_{j=1}^{\overline{T}} S_j - \sum_{j=1}^{\overline{T}-2} N_{j+1}^0 \frac{\hat{p}(j)b}{1 - \hat{p}(j)b} - A_{\overline{T}} \frac{\hat{p}(\overline{T}-1)b}{1 - \hat{p}(\overline{T}-1)b}. \tag{B.5}$$

Claim B.3. *If $b \geq \underline{b} \in (0,1)$, there exists a unique function $p(b)$ that satisfies Equation*

(B.5), *with $p(b) \in (\sum_t S_t, 1)$ and $t = 1, ..., \overline{T}$.*

Proof. To prove the claim, we proceed by induction. First, we show the claim for $\overline{T} = 2$. For this case, let

$$\beta_2(p, b) \equiv pb + \frac{\hat{p}(1)b}{1 - \hat{p}(1)b} A_2 = S_1 + S_2. \tag{B.6}$$

The right-hand side takes values within the unit interval. At the left-hand side of the equation, function $\beta_2(p, b)$ is continuous for all $p, b \in (0, 1)$. Moreover, it is strictly increasing in p. This is because $\hat{p}(1)b/(1 - \hat{p}(1)b)$ is increasing in $\hat{p}(1)$, and $\hat{p}(1)$ and pb are both increasing in p. Finally, $\beta_2(p, 1) \to p$, $\beta_2(0, b) \to 0$ and $\beta_2(1, b) \to (A_2 b)/(1 - b) + b$, increasing in b, with

$$\lim_{b \to 0} \beta_2(1, b) = 0 \quad \text{and} \quad \lim_{b \to 1} \beta_2(1, b) = \infty.$$

Thus, there exists a lower bound for b, $\underline{b}_2 \in (0, 1)$, such that $\beta_2(1, b) \geq S_1 + S_2$ for all $b \geq \underline{b}_2$.

We now show that the claim holds for $\overline{T} = 3$. More specifically, $\overline{T} = 3$ implies that

$$N_1^1 = S_1 + S_2 + S_3 - A_2 \frac{\hat{p}(1)b}{1 - \hat{p}(1)b} - A_3 \frac{b[\hat{p}(1) + \hat{p}(2) - b\hat{p}(1)\hat{p}(2)]}{[1 - \hat{p}(1)b][1 - \hat{p}(2)b]},$$

so that

$$\beta_3(p, b) \equiv pb + A_2 \frac{\hat{p}(1)b}{1 - \hat{p}(1)b} + A_3 \frac{b[\hat{p}(1) + \hat{p}(2) - b\hat{p}(1)\hat{p}(2)]}{[1 - \hat{p}(1)b][1 - \hat{p}(2)b]} = \sum_{t=1}^{3} S_t. \tag{B.7}$$

As for (B.6), the right-hand side takes values within the unit interval. At the left-hand side of the equation, $\beta_3(p, b)$ is continuous for all $p, b \in (0, 1)$. Function $\beta_3(p, b)$ is strictly increasing in p, because pb, $\hat{p}(1)b/(1 - \hat{p}(1)b)$, and

$$b[\hat{p}(1) + \hat{p}(2) - b\hat{p}(1)\hat{p}(2)] / [1 - \hat{p}(1)b][1 - \hat{p}(2)b]$$

are all strictly increasing in p. Finally, $\beta_3(p, 1) \to p$, $\beta_3(0, b) \to 0$ and $\beta_3(1, b) \to (A_2 b)/(1 - b) + b - A_3 + A_3/(1 - b)^2 = \beta_2(1, b) - A_3 + A_3/(1 - b)^2$. Hence, the same conclusions as for $\overline{T} = 2$ apply (i.e., there exists a lower bound for b, $\underline{b}_3 \in (0, 1)$, such that $\beta_3(1, b) \geq S_1 + S_2 + S_3$ for all $b \geq \underline{b}_3$).

Reiterating this analysis, we can invoke the principle of induction to conclude that there exists a lower bound for b, denoted by $\underline{b}_{\overline{T}} \equiv \underline{b} \in (0, 1)$, such that $\beta_{\overline{T}}(1, b) \geq S_1 + S_2 + \ldots + S_{\overline{T}}$ for all $b \geq \underline{b}$. Hence, provided $b \geq \underline{b}$, a unique value of $p(b)$ exists, with $p(b) \in (\sum_t S_t, 1)$ and $t = 1, ..., \overline{T}$. Q.E.D.

The main implication of the claims above is that, while we can pin down the per-period costs from the choice probabilities, we can only recover an implicit rela-

51

tionship between b and p, as given by $p(b)$. Therefore, our structural parameters $(p, b, F(t, \sigma_t))$ are not uniquely determined by model and data alone, and we need to impose additional restrictions to separately identify b and p.

B.2 Additional Restriction for Identification

In line with the main text, the assumption we make is that, in $t = 1$, the per-period costs are the same pre- and post-breakthrough:

Assumption B.1. *Pre- and post-breakthrough revision costs are equal in $t = 1$:* $F(1, 1) = F(1, 0) = F(1)$.

Then, in any $t > 1$, the costs in the pre- and post-breakthrough phases, denoted by $(F(t, 0), F(t, 1))$, can be recovered by inverting (B.3) and (B.4). Moreover, by Assumption B.1, we can obtain $F(1)$ by inverting (B.3) evaluated in $t = 1$:

$$k_1 \equiv G^{-1}(1 - S_1/N_1^1) = \hat{\pi}(2) - \hat{\pi}(1) - F(1). \tag{B.8}$$

Then, from (B.4) evaluated at $t = 1$, we obtain

$$k_2 \equiv G^{-1}(1 - A_1/N_1^0) = b\hat{p}(1)\hat{\pi}(2) - F(1).$$

Substituting $F(1)$ from (B.8), and rearranging, yields

$$b\hat{p}(1) - (k_2 - k_1) = \frac{\hat{\pi}(2) - \hat{\pi}(1)}{\hat{\pi}(2)}. \tag{B.9}$$

In the next claim, we provide the sufficient conditions such that a value of $b(p)$ solving (B.9) exists and is unique.

Claim B.4. *Let $G(\varepsilon) = \frac{1}{1+\exp(-\varepsilon)}$. If $p \geq \underline{p} \equiv 1 - A_1^2$, then there exists a unique value of $b(p) \in (S_1/p, (1 - A_1)/p)$ such that*

$$b\hat{p}(1) - \frac{k_2 - k_1}{\hat{\pi}(2)} = \frac{\hat{\pi}(2) - \hat{\pi}(1)}{\hat{\pi}(2)}, \tag{B.10}$$

with $0 < S_1/p < (1 - A_1)/p < 1$ for all $p \geq \underline{p}$.

Proof. Under the Logistic distribution,

$$k_1 = \log\left(\frac{N_1^1 - S_1}{S_1}\right) \quad \text{and} \quad k_2 = \log\left(\frac{N_1^0 - A_1}{A_1}\right). \tag{B.11}$$

Thus, both sides of (B.10) are continuous for all $N_1^1 - S_1 = pb - S_1 > 0$, $N_1^0 - A_1 = 1 - pb - A_1 > 0$ and $1 > p, b > 0$. After using $N_1^1 = pb$ and $N_1^0 = 1 - pb$, Equation (B.10)

can be rewritten as

$$\alpha(p,b) \equiv b\hat{p}(1) - \left[\log\left(\frac{1 - bp - A_1}{A_1}\right) - \log\left(\frac{bp - S_1}{S_1}\right)\right]\frac{1}{\hat{\pi}(2)} = \frac{\hat{\pi}(2) - \hat{\pi}(1)}{\hat{\pi}(2)}. \quad \text{(B.12)}$$

Under our assumption of a strictly increasing and positive function $\hat{\pi}(\cdot)$, the right-hand side is a scalar that (strictly) lies within the unit interval. Moreover,

$$\lim_{b \to S_1/p} \alpha(p,b) = -\infty \quad \text{and} \quad \lim_{b \to (1-A_1)/p} \alpha(p,b) = +\infty.$$

Taken together, the continuity of α and its behavior in the limits imply that a solution exists for all $b \in (S_1/p, (1 - A_1)/p)$. We now determine the sufficient condition guaranteeing that such solution is unique. Specifically, this boils down to studying the monotonicity of $\alpha(p,b)$ with respect to b.

We then take the derivative of $\alpha(p,b)$ with respect to b and find that

$$\frac{\partial \alpha(p,b)}{\partial b} = \frac{p\left[1 - b\left(2 - bp\right)\right]}{(1 - bp)^2} + \frac{p\left(1 - A_1 - S_1\right)}{(1 - bp - A_1)\left(bp - S_1\right)\hat{\pi}(2)}.$$

Because $1 - A_1 - S_1 > 0$, $1 - pb = N_1^0 > A_1$ and $pb = N_1^1 > S_1$, the second term in the expression for $\frac{\partial \alpha(p,b)}{\partial b}$ is strictly positive. Using this fact, we provide below the sufficient restrictions under which $\frac{\partial \alpha(p,b)}{\partial b} > 0$.

To begin with, we note that if the first term of $\frac{\partial \alpha(p,b)}{\partial b}$ is weakly positive, or $p \geq (2b - 1)/b^2$, then $\alpha(p,b)$ is strictly monotone in b, with $(2b-1)/b^2$ increasing in b. Hence, if p is larger than this threshold evaluated at $b \to (1 - A_1)/p$ (the upper bound for the values of b such that (B.10) is well-defined), it is larger for all relevant values of b. Doing so, we find that $p \geq 1 - A_1^2$ is sufficient to guarantee that a solution to (B.10) is unique.[46] Moreover, $p \geq 1 - A_1^2$ implies that $0 < S_1/p < (1 - A_1)/p < 1$. \hfill Q.E.D.

This claim gives us the sufficient conditions for the existence of a unique $b(p)$, meaning that, thanks to Assumption B.1, point-identification can be achieved by jointly solving for p and b from $p = p(b)$ and $b = b(p)$. To conclude the analysis, we proceed as follows. First, we prove that, given Claim B.4, a unique (constrained) solution exists for p and b when $\overline{T} = 2$. That is, we obtain point-identification in $\overline{T} = 2$. Second, we give the additional condition under which this result on point-

---

[46]Were this condition violated (so that $p < (2b - 1)/b^2$), monotonicity holds if and only if the second term in $\frac{\partial \alpha(p,b)}{\partial b}$ is large enough. Since that term is strictly increasing in A_1, this boils down to finding the lower bound on A_1 such that the whole expression is strictly positive. We obtain that:

$$\frac{\partial \alpha(p,b)}{\partial b} \geq 0 \iff A_1 \geq \underline{A}_1 \equiv \frac{(1 - bp)^2\left(1 - S_1\right) + (1 - bp)\left[1 - b\left(2 - pb\right)\right]\left(bp - S_1\right)\hat{\pi}(2)}{(1 - bp)^2 + \left[1 - b\left(2 - bp\right)\right]\left(bp - S_1\right)\hat{\pi}(2)},$$

with $\underline{A}_1 < \left(0, N_1^0\right)$ for sufficiently large values of $\hat{\pi}(2)$. Hence, if $p < (2b - 1)/b^2$, $\alpha(p,b)$ is strictly increasing in b for all $A_1 > \underline{A}_1$.

identification extends to the case of a general $\overline{T}$. To close the model, the value of p and b can then be plugged into the expressions for the fixed costs.

Model Solution for $\overline{T} = 2$ We first show that a unique (constrained) solution exists in this case.

Claim B.5. *Let $\overline{T} = 2$ and $G(\varepsilon) = \frac{1}{1+\exp(-\varepsilon)}$. Then, there exists a unique couple (b^*, p^*) provided $p^* > 1 - A_1^2$, $b^* \geq \underline{b}$ and $p^* b^* \in (S_1, 1 - A_1)$.*

Proof. To begin with, we show two properties of $p(b)$ (Claim B.3) that are useful for what follows. Namely, we prove that $p'(b) < 0$ and $p''(b) > 0$ under $p > \underline{p}$.

After taking the total derivative of Equation (B.6) in the proof of Claim B.3, we find that

$$\frac{dp}{db} = p'(b) \equiv -\frac{\frac{\partial \beta(p,b)}{\partial b}}{\frac{\partial \beta(p,b)}{\partial p}} = -\frac{p}{b} \frac{1 + \frac{A_2[1 - b(2 - bp)]}{[1 - b(2 - bp)]^2}}{1 + \frac{A_2(1 - b)}{[1 - b(2 - bp)]^2}} < 0$$

for all $1 - b(2 - bp) > 0$ (which is implied by $p > 1 - A_1^2$, see the proof of Claim B.4). Moreover, we find that

$$\frac{dp'(b)}{db} = p''(b) \equiv p \frac{A_2^2\{1 - b[2 - b(2 - p)]\} + B^4 + A_2 B\{2 - b[2 + p(4 - bC)]\}}{b^2[A_2(1 - b) + B^2]^2} > 0,$$

where $B \equiv 1 - b(2 - b)p$, $C \equiv 9 + 3b^2 p - 2b(3 + p)$, $[A_2(1 - b) + B^2] > 0$, and the sign of the numerator is strictly positive under $1 - b(2 - bp) > 0$, or $p > 1 - A_1^2$. Importantly for what follows, $p'(b) < 0$ and $p''(b) > 0$ imply that $bp(b)$ is strictly decreasing in b.

We now plug $p(b)$ into Equation (B.12). We obtain

$$\frac{bp(b)(1 - b)}{(1 - p(b)) + p(b)(1 - b)} - \left[\log\left(\frac{1 - bp(b) - A_1}{A_1}\right) - \log\left(\frac{bp(b) - S_1}{S_1}\right)\right] \frac{1}{\hat{\pi}(2)} = \frac{\hat{\pi}(2) - \hat{\pi}(1)}{\hat{\pi}(2)},$$

where the left-hand side corresponds $\alpha(p(b), b)$ (see the proof of Claim B.4). Our goal is to establish that this equation has a unique solution in b.

We now study the behavior of $\alpha(p(b), b)$ over the support of b. Recall from Claim B.4 that function α is continuous for all $N_1^1 - S_1 = p(b)b - S_1 > 0$ and $N_1^0 - A_1 = 1 - p(b)b - A_1 > 0$. Moreover,

$$\lim_{bp(b) \to S_1} \alpha(p(b), b) = -\infty \quad \text{and} \quad \lim_{bp(b) \to (1 - A_1)} \alpha(p(b), b) = +\infty,$$

where the monotonicity of $p(b)b$ in b, established above, implies that this expression can be inverted to compute the lower and upper bounds on b such that $bp(b) \in (S_1, 1 - A_1)$. This then implies that a solution exists for all b such that $bp(b) \in (S_1, 1 - A_1)$.

Finally, we take the derivative of $\alpha(p(b), b)$ with respect to b and find that

$$\frac{d\alpha(p(b), b)}{db} = \frac{(1-b)(p(b) + bp'(b)) - bp(b)}{1 - bp(b)} + \frac{(1-b)(bp(b))(p(b) + bp'(b))}{(1 - bp(b))^2}$$

$$+ \frac{(1 - A_1 - S_1)(p(b) + bp'(b))}{\hat{\pi}(2)(1 - bp(b) - A_1)(bp(b) - S_1)}.$$

This expression is negative if $p(b) + bp'(b) < 0$, which holds true by the convexity of function $p(b)$. This then yields the strict monotonicity of $\alpha(p(b), b)$. Combined with the analysis above that guarantees existence of a solution for b, it implies that the solution is unique. We denote this solution by b^*.

To obtain the corresponding value of p^*, we plug b^* into $p(b)$ defined in Claim B.3, and obtain p^* provided $b^* \geq \underline{b}$. Finally, the couple (p^*, b^*) must satisfy $p^* b^* \in (S_1, 1 - A_1)$.   Q.E.D.

The claim proves that, given Claim B.4, the two-period model discussed in Section 3.5 is point-identified under constraints.

Model Solution for any $\overline{T} > 2$ We now provide the sufficient conditions that allow us to extend the result in Claim B.5 to a general $\overline{T}$. This boils down to providing the conditions such that a solution to the system of two equations in two unknowns $p = p(b)$ and $b = b(p)$ exists and is unique. We do this in the following claim.

Claim B.6. *Let $\overline{T} > 2$ and $G(\varepsilon) = \frac{1}{1 + \exp(-\varepsilon)}$. First, if a solution (b^*, p^*), with $b^* \geq \underline{b}$ and $p^* b^* \in (S_1, 1 - A_1)$, exists, it is unique if $p^* > h(A_1, \overline{T})$, where $h(\cdot, \overline{T})$ is the lower bound for p in period $\overline{T}$. Let $b^{min} \equiv \arg\min_b g(b)$ and $b^{max} \equiv \max\{\underline{b}, \arg\max_b g(b)\}$, with $g(\cdot) \equiv b^{-1}(\cdot)$. The couple (b^*, p^*) exists if and only if the following conditions are satisfied: if $g(b^{max}) < p(b^{max})$(resp. $g(b^{max}) > p(b^{max})$) then $g(b^{min}) > p(b^{min})$ (resp. $g(b^{min}) < p(b^{min})$).*

Proof. To establish uniqueness conditional on existence, it is sufficient to prove the monotonicity of $p(b)$ and $b(p)$.

We begin with $b(p)$. Taking the total derivative of Equation (B.12), we find that

$$\frac{db}{dp} = b'(p) \equiv -\frac{\frac{\partial \alpha(p, b)}{\partial p}}{\frac{\partial \alpha(p, b)}{\partial b}} = -\frac{b}{p} \frac{\frac{1-b}{(1-bp)^2} + \frac{1 - A_1 - S_1}{(1 - bp - A_1)(bp - S_1)\hat{\pi}(2)}}{\frac{1 - b(2 - bp)}{(1-bp)^2} + \frac{1 - A_1 - S_1}{(1 - bp - A_1)(bp - S_1)\hat{\pi}(2)}} < 0$$

for all $1 - b(2 - bp) > 0$ (which is implied by $p > 1 - A_1^2$).

As far as $p(b)$ is concerned, we prove in Claim B.4 that, when $\overline{T} = 2$, $p'(b) < 0$ if $p > \frac{2b-1}{b^2}$ (which holds true when $p > 1 - A_1^2$). If $\overline{T} = 3$, similar calculations show that $p'(b) < 0$ for all $p > \frac{(3-b)b-2}{b^2(3-2b)} > \frac{2b-1}{b^2}$. Because $\frac{(3-b)b-2}{b^2(3-2b)}$ increases in b, plugging $b = (1 - A_1)/p$ yields $p > h(A_1, 3) \equiv \frac{1}{4}(1 - A_1)(3 + 3A_1 + \sqrt{1 + A_1(2 + 9A_1)})$, with

$\underline{p} = 1 - A_1^2 < h(A_1, 3) < 1$. Proceeding analogously for $\overline{T} = 4$, we find that $p'(b) < 0$ if $p > h(A_1, 4) > h(A_1, 3)$. Thus, we can conclude that, when $\overline{T} > 2$, there exists $h(A_1, \overline{T})$ such that $b(p)$ is strictly monotone for all $p > h(A_1, \overline{T})$.

We now develop the conditions for existence in the second part of the claim. First we note that, since $b(p)$ is monotone, there exists $g(\cdot) \equiv b^{-1}(\cdot)$. Then, we use the following facts:

- Function $p(b)$ is such that $p(\underline{b}) = 1$ and $p(1) = \sum_{t=1}^{\overline{T}} S_t$, with $t = 1, ..., \overline{T}$.

- The support of function $g(b)$ is given by $(S_1/p, (1 - A_1)/p)$. Moreover, $1 > g(b) > \underline{p}$.

Given these properties, existence is guaranteed if $p(b)$ and $b(p)$ cross over their domain, which holds true under the conditions in the claim. Q.E.D.

The claim does two things:

1. It establishes the condition ensuring the monotonicity of functions $p(b)$ and $b(p)$. Specifically, we define a lower bound for p, denoted by $h(A_1, \overline{T})$, with $h(A_1, \overline{T}) > h(A_1, \overline{T} - 1) > ... > h(A_1, 3) > 1 - A_1^2 = \underline{p}$, such that $b(p)$ and $p(b)$ are both strictly decreasing. This implies that, if a solution (b^*, p^*) exists, it is also unique.

2. Monotonicity alone only gives us that $p(b)$ and $b(p)$ cross at most once; that is, it is not sufficient to prove existence. Then, in the second part, the claim gives the conditions under which the functions cross over their domain.

C   Likelihood Function with Censored Projects

We define a censored project in T as a project that has not been either abandoned or published by version T. This means, the project has not experienced any high cost shock, which would have implied the termination of the process. Moreover, a censored project might have experienced a breakthrough in any $\tau < T$ (implying the post-breakthrough phase in T). We denote the status of a censored project by $\sigma = \varnothing$, and the probability that a censored project reaches version T by $\Pr(T, \sigma_T = \varnothing)$. This probability is equal to:

$$\Pr(T, \sigma_T = \varnothing) = \frac{\Pr(T, \sigma_T = 1)}{1 - G^1(T)} + \frac{\Pr(T, \sigma_T = 0)}{1 - G^0(T)}$$

where $\Pr(T, \sigma_T = 1)$ in Equation (14) is the probability that a project is published in T (with status $\sigma_T = 1$), and $\Pr(T, \sigma_T = 0)$ in Equation (16) is the probability that a project is abandoned in T (with status $\sigma_T = 0$). Moreover, we have

$$LL(T, \sigma_T = \varnothing | b, p, \boldsymbol{F}) = \log(\Pr(T, \sigma_T = \varnothing)).$$

We can then rewrite the log-likelihood of the data in Equation (18) as

$$LL(b, p, \boldsymbol{F}) = \sum_{i=1}^{N} LL(T_i, \sigma_{T_i} | b, p, \boldsymbol{F}) \tag{C.1}$$

with index i for a given project $i = 1, \ldots, N$ that reaches version T_i with status $\sigma_{T_i} \in \{0, 1, \varnothing\}$ in T_i. This expression for the log-likelihood accounts for all censored projects.

D   Heterogeneity Results

In this section, we estimate the rate of learning b as a function of additional project and author-team characteristics: communication, experience (two versions), commerciality, and team size. For each of these, we estimate four different values for the rate of learning, b_i with $i = 1, 2, 3, 4$, for four different categories of the explanatory variable. For our results in this section, we use the three-step estimator and, for the estimation of $\hat{\pi}(t)$ in the first step, rely on the full sample of RFCs, that means, including those initiated in or after the year 2010.[47]

Communication: We measure communication (or attention) using the average number of email messages sent during the revision process (i.e., average number of emails per version). We estimate the rate of learning for four categories: b_1 for projects with average number of emails, x, at or below the 25th percentile; b_2 for projects with x above the 25th percentile and at or below the median; b_3 for projects with x above the median and at or below the 75th percentile; b_4 for projects with x above the 75th percentile.

Experience: For the experience of an author-team, we construct two different measures: a narrower one and a broader one). For the narrower, we count, for each author in an author-team, the number of successfully published RFCs at the time of the initial draft of a given project (max RFCs). The experience of the author-team is then the experience of the most *successful* author. For the broader, we count, for each author in an author-team, the number of completed projects (successful or failed) at the time of the initial draft of a given project (max Projects). The experience of the author-team is then the experience of the most *prolific* author. We estimate the rate of learning for four categories:

---

[47]For all results in Table 2 in the main text, we use a restricted sample to estimate patent citations $\hat{\pi}(t)$, excluding RFCs initiated in year 2010 or later. This is a consequence of our bootstrapping approach, where we estimate the patent citations $\hat{\pi}(t)$ for each of the 1,000 random samples. Because we restrict our ML estimation sample to projects initiated until 2009, later RFCs are excluded for the estimation of patent citations. For the results in this appendix, we do not face this restriction (we do not bootstrap standard errors) and therefore use an extended sample to estimate patent citations.

b_1 for projects with $x = 0$ and b_2 through b_4 for projects by the terciles of the remaining projects $(x > 0)$;

Commerciality: We measure commerciality as the share of corporate email addresses from which emails in response to different versions of a project draft are sent. This "suit-to-beard" ratio is meant to proxy the commercial interest in the project. We estimate the rate of learning for four categories: b_1 for $x = 0$, b_4 for $x = 1$, and b_2 and b_3 for projects with a suit-to-beard ratio below and above the median of the remaining projects $(x \in (0, 1))$.

Team Size: We use the number of authors of the initial draft of a project capture the size of an author-team. We estimate the rate of learning for four different author-team sizes: b_1 for author-teams with a single author, b_2 for projects with 2 or 3 authors, b_3 for projects with 4 or 5 authors, and b_4 for projects with 6 or more authors.

In Table D.1, we provide summary statistics of our explanatory variables. We also give the count of projects in each of the four categories. The proxies for experience and team size are pre-determined, that is, exogenous to the revision process of a given project. The proxies for communication and commerciality, on the other hand, are endogenous to the revision process. They are both based on the number of emails sent in response to a version of the project *during* the revision process. For the results below, we first use these contemporaneous measures. We then also use an alternative measure that is less plagued by endogeneity concerns, but available only for the working group sample: the average number of emails sent in response to a version of all *other* projects submitted to the same working group of a given project during that project's lifetime.

The results allow us to analyze how communication, authors' experience, commerciality of a project, and the team size affect the process of learning and the prospects for a project to experience a breakthrough. We show that, while heterogeneity gives rise to differential patterns in terms of authors' rate of learning, the qualitative features of our results remain similar to the ones discussed in the baseline analysis of Section 4.

D.1 Communication

For our first set of heterogeneity results, we ask how project-related communication drives learning. We let b vary with the amount of attention and feedback a project receives. We hypothesize that more attention (via more project-related communication) is associated with a higher learning rate (i.e., higher values for b).

The results for communication in Column (1) of Table D.2 show that more communication (and attention) increases the estimated value for b. The resulting differential

Table D.1: Summary Statistics: Explanatory Variables

						Category			
Explanatory variable x	N	Mean	SD	Min	Max	1	2	3	4
Communication (Emails)									
Full Sample	14,444	4.99	9.65	0	292	3,571	3,647	3,614	3,612
WG Sample	3,168	5.20	8.08	0	188	717	865	794	792
WG Sample (Exogenous)	3,086	4.66	3.35	0	26.6	772	771	771	772
Experience (max RFCs)	13,922	3.82	8.13	0	68	7,081	2,645	1,984	2,212
Experience (max Projects)	13,922	14.80	26.45	0	254	2,885	3,720	3,567	3,750
Commerciality (Suit-to-Beard)	10,710	0.78	0.27	0.00	1	626	4,751	851	4,482
Team Size (Author Count)	13,922	2.25	1.85	5,935	72	5,935	5,750	1,708	529

Notes: This table provides summary statistics for the explanatory variables x used for the heterogeneity results in Tables D.2, D.3, and D.4. The numbers are for the sample of standards-track projects only. The table reports the number of observations (N), mean, standard deviation (SD), minimum and maximum value, as well as the number of observations (i.e., project proposals) for each of the four categories. For communication, categories are by quartile; for experience, category 1 is for $x = 0$, and categories 2 through 4 by tercile of remaining projects; for commerciality, category 1 is for $x = 0$, category 4 for $x = 1$, and categories 2 and 3 are projects below and above the median of remaining projects; for team size, categories are for $x = \{\{1\}, \{2, 3\}, \{4, 5\}, \{6, \ldots, 72\}\}$. The difference in the number of observations is the result of missing author information for 522 standards-track projects and no emails for 3,734 projects.

effects of communication on the rate of learning are as follows: the conditional probability of experiencing a breakthrough in any period t increases from $b = 0.12$ in the lowest category to $b = 0.34$ in the highest category. This means that, compared to projects with few emails, communication and attention increase the breakthrough arrival rate by close to 200%, leading teams to more quickly abandon those projects without a breakthrough. The reason for this is that a higher value of b induces faster updating of beliefs, and players become pessimistic more rapidly. At the same time, a breakthrough, if experienced, arrives faster, which results in higher continuation value in the pre-breakthrough phase. We plot the ex-ante value of a project (before the initial draft is observed) in Figure D.1 below.

Sound internal communication is said to be important for the effective functioning of organizations (Arrow, 1974). Empirical evidence on the link between communication and productivity in organizations, however, is scant. Among the few exceptions are Palacios-Huerta and Prat (2012), who use email exchange to study the relationship between communication and the importance of the members of an organization; Bloom et al. (2014), who show that the introduction of intranet changes the extent of delegation within an organization; and Battiston et al. (2017), who study the impact of face-to-face communication on productivity. We contribute to this literature by showing that communication is directly linked to the rate of learning in a research organization.

Our measure of communication (the emails sent *during* the development of the project) is likely to be plagued by endogeneity concerns. We argue that communication and attention affect the learning process and the development of the project. In return, we would expect an evolving project to influence communication and attention. As an alternative measure for communication for a given project i, we use

Table D.2: Non-Parametric Heterogeneity Results

Specification	Communication	Experience		Commerciality	Team Size
(Explanatory variable x)	(Emails)	(RFCs)	(Projects)	(Suit-to-Beard)	(Authors)
	(1)	(2)	(3)	(4)	(5)
Step 3: Parameter Estimates b and p (Standard-Track Projects)					
b (Category 1)	0.12	0.23	0.21	0.27	0.26
	(0.003)	(0.004)	(0.005)	(0.007)	(0.005)
b (Category 2)	0.27	0.29	0.25	0.36	0.29
	(0.006)	(0.007)	(0.005)	(0.007)	(0.005)
b (Category 3)	0.33	0.34	0.30	0.38	0.33
	(0.007)	(0.009)	(0.007)	(0.010)	(0.010)
b (Category 4)	0.34	0.37	0.35	0.19	0.33
	(0.007)	(0.010)	(0.007)	(0.005)	(0.017)
p	0.41	0.40	0.40	0.43	0.40
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
log-likelihood/Project	-2.184	-2.239	-2.239	-2.524	-2.248
Projects (standards-track)	14,444	13,922	13,922	10,710	13,922

Notes: This table presents ML estimates for the three-step model in Column (2) in Table A.2 with the rate of learning b varying in x. The top panel presents estimates for b and p from the third step of our estimation procedure (using data on standards-track projects), given the cost estimates from the second step (using data on nonstandards-track projects): 2.58 (0.053) for $t = 1$, 0.63 (0.008) for $t = 10$, and 0.59 (0.031) for $t = 20$. The rate of learning b is estimated for four different categories of projects: in (1), a value b_i for each quartile; in (2) and (3), b_1 for projects $x = 0$, and b_2 through b_4 for each tercile of remaining projects; in (4), b_1 for $x = 0$, b_4 for $x = 1$, and b_2 and b_3 for projects below and above the median of remaining projects; in (5), b_i for $x = \{\{1\}, \{2,3\}, \{4,5\}, \{6,\ldots,72\}\}$. Estimates for project values $\hat{\pi}(t)$ are based on the extended sample, including RFCs initiated in year 2010 or after. Standard errors for the parameter estimates are reported in parentheses. Standard errors for the cost estimates are calculated using the delta method.

the number of emails that were sent within the working group, during the lifetime of project i, but exclude emails that pertain to the specific project i. Assuming that there are no communication spillovers (or little) between projects, this approach allows us to treat communication as exogenous to project i itself. This measure, however, can be constructed only for the working group sample. We provide the parameter estimates of this alternative approach in Table D.3. We find that the basic pattern observed in Column (1) of Table D.2 survive.

D.2 Experience

Table D.1 provides a glimpse at the distribution of the measures of author experience. One half of the author-teams of standards-track projects have authors that, at the time they submitted the initial draft of a given project, had not successfully completed a project (Category 1). The fact that such a large fraction of author-teams is inexperienced is important for our estimation. It implies that for a large number of projects the learning process is not tainted by repeat authors who have previously figured out how to properly write to experience a breakthrough. This feature of the

Table D.3: Non-Parametric Heterogeneity Results – WG Emails

Specification (Explanatory variable x)	Communication (Full) (1)	Communication (WG) (6)	Communication (within WG) (7)
Step 3: Parameter Estimates b and p (Standards-Track Projects)			
b (Category 1)	0.12	0.15	0.20
	(0.003)	(0.007)	(0.010)
b (Category 2)	0.27	0.24	0.38
	(0.006)	(0.015)	(0.023)
b (Category 3)	0.33	0.37	0.35
	(0.007)	(0.015)	(0.018)
b (Category 4)	0.34	0.39	0.36
	(0.007)	(0.020)	(0.018)
p	0.41	0.46	0.45
	(0.002)	(0.004)	(0.004)
log-likelihood/Project	-2.184	-3.496	-3.513
Projects (standards-track)	14,444	3,168	3,086

Notes: This table presents ML estimates for the three-step model in Column (2) in Table A.2 with the rate of learning $b(x)$ varying in x. We reproduce Column (1) from Table D.2. Column (6) is the baseline model with varying $b(x)$ for the working group sample. For Column (7), we use emails sent within a working group (but without those sent in response to the specific project) as communication measure. The difference in observations stems from projects that are the only projects in a working group (at that time); these projects are dropped from the sample for Column (7). The top panel presents estimates for $b(x)$ and p from the third step of our estimation procedure (using data on standards-track projects), given the cost estimates from the second step (using data on nonstandards-track projects): 2.58 (0.053) for $t = 1$, 0.63 (0.008) for $t = 10$, and 0.59 (0.031) for $t = 20$ in the full sample (with 1,647 observations) and 2.12 (0.081) for $t = 1$, 0.70 (0.015) for $t = 10$, and 0.51 (0.043) for $t = 20$ in the WG sample (with 764 observations). The rate of learning b is estimated for four different categories of projects, with b_i the value for the i's quartile of emails. Estimates for project values $\hat{\pi}(t)$ are based on the extended sample, including RFCs initiated in year 2010 or after. Standard errors for the parameter estimates are reported in parentheses. Standard errors for the cost estimates are calculated using the delta method.

data lends support to our model assumption that projects of these authors are not correlated over time.

Columns (2) and (3) in Table D.2 summarize the results. We find that more experienced author-teams (both successful and prolific) exhibit a higher rate of learning (with an increase of roughly 50% from the lowest to the highest category). This effect might be the result of more prominent authors receiving more attention (Simcoe and Waguespack, 2011) or simply accumulated knowledge about how to identify useful comments and suggestions and incorporate them into a revision.

D.3   Commerciality

The results in Column (4) of Table D.2 consider heterogeneity in the rate of learning that varies with the commerciality (or "suit-to-beard" ratio) of a proposal. We find a non-monotonic relationship between commerciality and the learning parameter b.[48]

---

[48]This pattern is robust to a variety of different versions of the suit-to-beard ratio.

Figure D.1: Ex-Ante Project Value (Varying with Rate of Learning b)
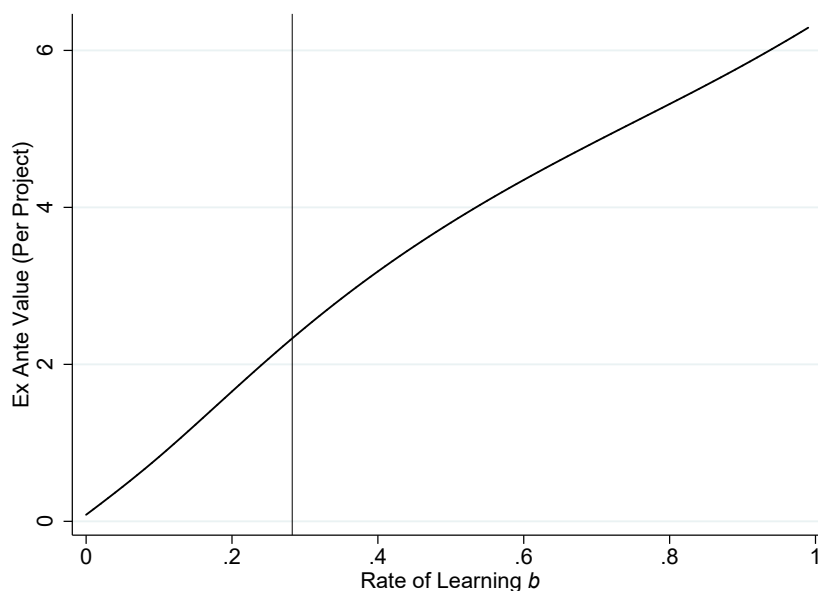
Notes: This figure depicts the ex-ante project value (in $t = 0$ before authors are endowed with their initial version) as a function of the rate of learning b. We plot the ex-ante value for the parameter estimates from the three-step model in Column (2) of Table A.2. The vertical line depicts $b^* = 0.283$.

The decline in b that we observe when moving from Category 2/3 to Category 4 is consistent with the findings in Simcoe (2012), where commerciality is linked to slower standards production.[49] For these results, we can interpret b as a reduced-form parameter that captures slower compromise when commercial participants have stronger vested interests in the outcome of the standardization process. The low value of b for proposals at the minimum of commerciality is likely due to confounding with communication. Specifically, projects that generate very few emails are more likely to exhibit extreme values of the "suit-to-beard" measure, and these have lower values of b as seen in Column (1).

D.4   Team Size

We find that larger author-teams have a higher rate of learning b. To interpret these results we offer the following possible explanations. The first is that more authors write better drafts because of more combined knowledge and skills (especially in the presence of complementarities in skills), and thus experience a breakthrough faster. This argument relies on insights that are similar to those offered to interpret the results in columns (2) and (3), and is consistent with the evidence in Hamilton et al.

---

[49]In that study, commercial interest was measured at the WG level, although supplemental results included estimates based on a proposal-level measure of commerciality similar to the one used here.

([2003](#)) that more heterogeneous teams are more productive than other teams with the same ability.

The second explanation is more mechanical, and relies on the idea that more authors cover a larger spectrum/make up a larger part of the community and thus experience a breakthrough easier (take 6 members: 1 author has to convince 5 others; 5 authors have to convince only 1 other member). A third explanation is that authors write better drafts because of combined effort. Accordingly, for a given draft quality, each author exerts less effort and incurs less cost; more authors can then exert more effort at still lower cost and write a better draft. However, this argument relies on agents' effort choice, which is outside of our model.

D.5 Heterogeneity and Projects' Ex-Ante Quality

For the results in Table D.2, we keep the value for the quality prior p fixed. In other words, while allowing the rate of learning b to vary with communication, experience, commerciality, and team size, we assume that the ex-ante quality of a project is a constant p regardless of the explanatory variable x. We extend this approach by allowing $p(x)$ to vary in x alongside $b(x)$ and report the results in Table D.4. We find that the variation in $b(x)$ is not the consequence of a constant value of p. In fact, letting $p(x)$ vary in x reveals that the point estimates for the rate of learning show a stronger response to our explanatory variables than the estimates for the quality prior. For author-team experience, for example, this result can be interpreted as follows: while more experienced authors initiate better projects (with a higher ex-ante quality), the effect of their experience on receiving consensus is stronger than their skills in picking good projects.

Table D.4: Non-Parametric Heterogeneity Results (Varying b and p)

Specification (Variable x)	Communication	Experience		Commerciality	Team Size
	(Emails)	(RFCs)	(Projects)	(Suit-to-Beard)	(Authors)
	(1)	(2)	(3)	(4)	(5)
Step 3: Parameter Estimates b and p (Standards-Track Projects)					
b (Category 1)	0.16	0.25	0.23	0.22	0.27
	(0.009)	(0.005)	(0.007)	(0.018)	(0.006)
b (Category 2)	0.29	0.29	0.26	0.34	0.29
	(0.006)	(0.008)	(0.007)	(0.005)	(0.006)
b (Category 3)	0.33	0.32	0.30	0.38	0.32
	(0.007)	(0.010)	(0.007)	(0.012)	(0.010)
b (Category 4)	0.34	0.36	0.34	0.28	0.32
	(0.008)	(0.010)	(0.007)	(0.007)	(0.018)
p (Category 1)	0.34	0.38	0.38	0.36	0.39
	(0.010)	(0.003)	(0.005)	(0.013)	(0.003)
p (Category 2)	0.40	0.40	0.39	0.46	0.40
	(0.003)	(0.004)	(0.004)	(0.002)	(0.003)
p (Category 3)	0.42	0.41	0.40	0.52	0.40
	(0.003)	(0.004)	(0.003)	(0.005)	(0.004)
p (Category 4)	0.42	0.42	0.41	0.36	0.41
	(0.003)	(0.004)	(0.003)	(0.004)	(0.008)
log-likelihood/Project	-2.181	-2.237	-2.237	-2.483	-2.248
Projects (standards-track)	14,444	13,922	13,922	10,710	13,922

Notes: This table presents ML estimates for the three-step model in Column (2) in Table A.2 with both the rate of learning $b(x)$ and the quality prior $p(x)$ varying in x. The top panel presents estimates for $b(x)$ and $p(x)$ from the third step of our estimation procedure (using data on standards-track projects), given the cost estimates from the second step (using data on nonstandards-track projects): 2.58 (0.053) for t = 1, 0.63 (0.008) for t = 10, and 0.59 (0.032) for t = 20. Parameters b and p are estimated for four different categories of projects: in (1), a value b_i, p_i for each quartile; in (2) and (3), b_1, p_1 for projects x = 0, and b_2, p_2 through b_4, p_4 for each tercile of remaining projects; in (4), b_1, p_1 for x = 0, b_4, p_4 for x = 1, and b_2, p_2 and b_3, p_3 for projects below and above the median of remaining projects; in (5), b_i, p_i for x = $\{\{1\}, \{2,3\}, \{4,5\}, \{6, \ldots, 72\}\}$. Estimates for project values $\hat{\pi}(t)$ are based on the extended sample, including RFCs initiated in year 2010 or after. Standard errors for the parameter estimates are reported in parentheses. Standard errors for the cost estimates are calculated using the delta method.

E Data Appendix

E.1 Construction of ID Sample

Our primary data source is the online archives of the IETF, which contain the full text for each version of every Internet Draft (ID) submitted after July 1990, along with various pieces of bibliographic information.

We begin by obtaining a list of all IETF IDs from the Internet Draft Status Summary file at

https://tools.ietf.org/id/all_id.txt.[50]

IDs in this Status Summary file are identified by a string of hyphenated alpha-numeric characters. For example, the ID name for RFC 7368 ("IPv6 Home Networking Architecture Principles") is

draft-ietf-homenet-arch.

All IDs begin with `draft`; the string `ietf` indicates that this particular ID is a working group ID (see below); the remaining pieces identify a specific proposal. Different versions of an ID are indexed by a suffix `-00` for the initial version and `-NN` for the NN's revision. The Status Summary file contains a list of the ID-version strings for the last version of an ID. For example, for RFC 7368, this last ID was `draft-ietf-homenet-arch-17` (where `-17` indicates the 17th revision of the initial proposal or the overall 18th version).

The total number of IDs on the list is 28,627. For each, we download the text document of all versions from

http://tools.ietf.org/id/[ID-version].txt

where `[ID-version]` is the respective ID-version combination. For example, for RFC 7368, we know the first version is `draft-ietf-homenet-arch-00`, and the last version is `draft-ietf-homenet-arch-17`. If available, we download the files from the first to the last version. Moreover, because the Status Summary file contains the date only for the last ID-version, we parse the history information for a given ID at http://datatracker.ietf.org/doc/[ID]/history (with `[ID]` the ID name without the version numbers) to obtain the online publication dates for each version. These are the dates a given version is posted to the online repository. Especially for older IDs, this date is not always the date the version was finished and circulated. We manually clean the date information in two steps. First, assuming that a version $t+1$ is not published before a version t, we find the set of versions for which the time difference between $t+1$ and t is negative (and version $t+1$ has a date before version t). We then manually inspect the draft document to find the correct dates of publication.

---

[50]Downloaded on September 10, 2015. This date marks our data cut-off.

If a complete publication date is not provided, we use the statutory expiration date (six months after the ID version is posted; if listed in the document) to backtrack the publication date; if the day of the month of the ID version is not given, we use the first of the month (or the date of previous ID version, whichever comes later); if no date is provided in the draft document, we use the date of the previous ID version.

IDs go through various stages of development and change their ID name along the way. For instance, an ID may be initiated by an individual IETF contributor and is later adopted by a working group (WG). Our example ID (RFC 7368) is such a case. Its original ID name was `draft-arkko-townsley-homenet-arch` (with one version). It was replaced (i.e., superseded) by ID `draft-chown-homenet-arch` (with two versions), before being adopted by the `homenet` WG as ID `draft-ietf-homenet-arch` (with 18 versions).

```
draft-arkko-townsley-homenet-arch

                ↓

       draft-chown-homenet-arch

                ↓

        draft-ietf-homenet-arch

                ↓

              RFC 7368
```
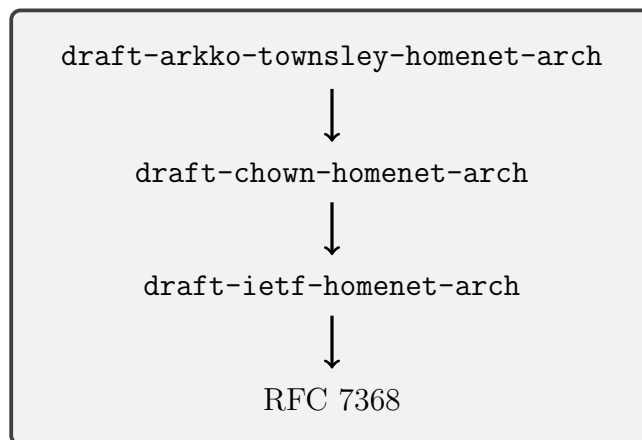
The Status Summary file contains the status of each ID, including information on the ID name that replaced a given ID. We use this information to link such chains of IDs to create a single *project*. In many cases, a subsequent ID supersedes an "expired" ID. As a statutory rule of the IETF, an unpublished ID expires after six months of inactivity (i.e., no submission of a new version), leading to its removal from "active" status. Authors, however, can reactivate and resume an expired ID. In our data, we see numerous cases where more than six months have passed between two versions of a given ID (see below for how we handle such cases). Likewise, we see numerous cases where more than six months have passed between two subsequent IDs in a project chain. For the construction of our project chains, we implement the following rule (referred to as the "24-months rule"). If less than 24 months have passed between the last version of an ID i and the first version of its successor (the ID $i + 1$ that replaces the previous i in the chain), then we link the two IDs to obtain a single project chain. Instead, if more than 24 months have passed between the last version of an ID i and the first version of its successor $i + 1$, we delink the two IDs. The first chain ends with the last version of the predecessor ID i, the second chain begins with the first version

of the successor ID $i + 1$. We thus delink 32 project chains.

After linking IDs, our sample contains 25,532 project chains with 96,770 versions. The longest chain links five IDs; 10 chains link four IDs; 205 chains link three IDs; 2,098 chains link two IDs, and 23,218 chains comprise only one ID.

We have applied our 24-months rule to delink chains if the gap between two presumably related IDs—recall, we obtain the linking information from the IETF's Status Summary file—is too big. Moreover, for many projects, we see considerable gaps between two versions of the same ID. Gaps longer than six months are in violation of the IETF's statutory rule for expiration, but is fairly common. We therefore take a conservative approach when determining the de-facto status of a project chain. First, if there is a gap of more than 24 months within a project chain (i.e., within an ID of a given project chain), we drop all versions after the gap (a total of 1,780 versions in 451 chains). Second, project chains that have a version that was published less than 24 months before our data cut-off date (September 10, 2015) and that are not published as an RFC are considered "active." We thus get three different outcomes for our project chains:

Published: A project chain is "published" if it is published as an RFC or in the final stages (i.e., in the "Queue" but not yet assigned an RFC number); the full sample contains 5,834 published project chains.

Abandoned: A project chain is "abandoned" if it is not published as an RFC and no new version has been posted after September 10, 2013; the full sample contains 16,038 abandoned project chains.

Active: A project chain is "active" if it is not published as an RFC and a new version has been posted after September 10, 2013; the full sample contains 3,660 active project chains.

For the first version in each project chain, we take the first version for which we have a document. Thus, by construction, we have all initial versions for each project chain. We further miss 432 documents for later version numbers. For version-level information that we obtain from the ID documents (authors, text distance), we interpolate to account for these missing values.

After constructing our raw sample, we take the following steps to construct our estimation sample:

Step 1: Drop 456 project chains from six special technology areas. We provide more details in Section E.2.3 below.

Step 2: Drop 325 non-IETF projects: These projects are 19 IESG projects ("Internet Engineering Steering Committee"), 3 IANA projects ("Internet Assigned Numbers Authority"), 193 IRTF projects ("Internet Research Task Force"), and 110 IAB projects ("Internet Architecture Board").

Step 3: Drop 444 project chains with specialized RFCs. We provide more details in Section E.2.2 below.

Step 4: Drop 529 project chains initiated before 1996.

Step 5: Drop 3,526 active project chains (that means, active projects following the 24-months rule).

Step 6: Drop 4,161 completed (published or abandoned) project chains initiated in 2010 or later. We thus minimize selection on outcomes by dropping completed projects initiated in 2010 or later. Out of the 3,526 projects active at the cut-off date, and 3,479 (98.7%) projects were initiated in 2010 or later.

Table E.1: Sampling

Step	Description	Projects	Versions
	Full Sample (after applying the 24-months rule)	25,532	94,990
Step 1	Drop six areas	25,076	92,249
Step 2	Drop special non-IETF series	24,751	90,845
Step 3	Drop special tracks	24,307	88,187
Step 4	Drop projects initiated before 1996	23,778	86,757
Step 5	Drop active projects	20,252	73,654
Step 6	Drop completed projects initiated in or after 2010	16,091	57,179
	Final Estimation Sample	16,091	57,179

We summarize these steps in Table E.1. Our final sample consists of 16,091 project chains (3,857 published; 12,234 abandoned) with 57,179 versions. This is the sample used for all results but those presented in Column (4) in Table A.2. For this latter set of results, we use the information from right-censored observations. In other words, we do not drop active project chains (**Step 5**). Moreover, because by including active project chains we are not selecting on outcomes when keeping all projects initiated in 2010 or later, we do not drop this set of recent projects (**Step 6**). This extended sample consists of 23,778 project chains (4,781 published; 15,471 abandoned; 3,526 active) with 86,757 versions.

E.2 Further Information on IDs

E.2.1 Working Group

The file-naming convention of the IETF allows us to identify a project chain's relevant working group. IDs with names that begin with `draft-ietf` are IDs from a working group, and the working group name is the third part of the name. For our example

project chain (RFC 7368), the third (and last) ID is from the `homenet` working group. We construct a subsample of all project chains that are initiated by a working group and refer to it as the working group sample (or WG sample). A project chain is in the WG sample if its first ID (in the chain) is from a working group. The WG sample consists of 3,932 project chains (2,206 published; 1,726 abandoned) with 22,025 versions. The project chains in this sample are initiated by 285 different working groups (both active and closed) and (at some point in time) part of 315 different working groups.

E.2.2 Standards-Track and Nonstandards-Track

For the distinction of standards-track and nonstandards-track project chains, we use information on the "status" of published RFCs. We first access the "RFC Index" at http://www.rfc-editor.org/rfc-index.html to obtain information on the title, the publication date, the status, area, and the working group of published RFCs. An RFC's status holds the information we need to identify a project chain's *track*. An RFC can be of one of a number of statuses: "Best Current Practice," "Draft Standard," "Experimental," "Historic," "Informational," "Internet Standard," "Proposed Standard," and "Unknown." We obtain three track categories for RFCs:

> **No Track:** An RFC with status equal to "Best Current Practice" (184 projects) "Draft Standard" (104 projects) "Historic" (111 projects), "Internet Standard" (45 projects) or "Unknown." In **Step 3** above (see Table E.1), we drop these projects from our sample.

> **Nonstandards Track:** An RFC with status equal to "Experimental" or "Informational."

> **Standards Track:** All other RFCs.

To link RFC-level information to our sample of project chains, we access metadata for each RFC through

$$\text{http://datatracker.ietf.org/doc/rfc[NNNN]}$$

where `[NNNN]` is the four-digit RFC number. We parse the website to pull the last ID name that lead to the published RFC. We then match our RFC-level information with our project-chain information. For each published project chain, we thus obtain track information through its RFC status. We further assume that all other project chains (i.e., those that are abandoned in our final estimation sample) are standards-track project chains.

Our final estimation sample consists of 14,444 project chains on the standards track (2,210 published; 12,234 abandoned) and 1,647 project chains on the nonstandards track. Out of these nonstandards-track projects, 1,385 are "Informational" and 262 are "Experimental."

E.2.3 Areas

We obtain area information via two separate avenues. First, working groups are assigned to specific technology areas. We download information for area-working group assignments for active working groups from https://datatracker.ietf.org/wg and https://tools.ietf.org/area (these lists are overlapping) and for concluded working groups from https://datatracker.ietf.org/group/concluded and https://tools.ietf.org/wg/concluded (these lists are overlapping). We can thus assign an area to each project chain that has at some point been adopted by a working group. If a project chain has multiple areas, we choose the last area the project chain was assigned to.

For the non-working group sample (i.e., individual projects), we obtain some area information from http://datatracker.ietf.org/doc/[ID] (with [ID] the ID name): for 1,020 ID names, the information on Document Type contains the line "individual in [A] area" where [A] is an area identifier. This means, an individual ID (outside a working group) has been assigned to an area. In our estimation sample, we thus obtain area information for 712 additional project chains.

In **Step 1** above, we have dropped projects from six areas from the sample: 246 projects (1,256 versions) from the *gen* area ("General Area"), 160 projects (1,308 versions) from the *art* area ("Applications and Real-Time Area"), 35 projects (134 version) from the *usr* area, 3 projects (11 versions) from the *irtf* area ("Internet Research Task Force"), 1 project (6 versions) from the *adm* area, and 11 projects (26 versions) from the *ora* area. Moreover, we combine 60 projects from the *sub* area with the *rtg* area ("Routing Area") and the 53 projects from the *mgt* area with the *ops* area ("Operations and Management Area"). After further applying **Steps 2** through **6**, we arrive at the following area count in Table E.2:

Table E.2: Areas

	Area Description	Projects		Versions
		Published	Abandoned	
app	Applications Area	492	411	5,173
rai	Real-Time Applications and Infrastructure Area	496	204	5,583
tsv	Transport Area	419	190	4,272
int	Internet Area	599	388	6,121
rtg/sub	Routing Area	498	305	5,713
ops/mgt	Operations and Management Area	423	287	4,447
sec	Security Area	447	324	4,403
	Projects with area information	3,374	2,109	35,667
	Projects without area information	483	10,125	21,512

The IETF technology areas correspond roughly to the various layers in the engineering "protocol stack" as described in Simcoe (2012). From top to bottom, those layers/areas are: Applications (app), Realtime Applications and Infrastructure (rai), Transport (tsv), Internet (int), and Routing (rtg/sub). The IETF also recognizes two areas that cut across the various layers: Operations (ops/mgt) and Security (sec).

For the full sample, we have area information for 7,948 project chains. For our estimation sample, we have area information for 5,483 project chains.

E.2.4 Author Information

We use Jari Arkko's `authorstats` tool, available at https://www.arkko.com/tools/authorstats.html, to parse ID documents and extract author names, affiliations, and email addresses. The goal is to identify authors and their project history at the IETF.

After running the `authorstats` tool on all documents of our full sample, we conduct some simple cleaning procedures. Using Python's name-entity-recognition libraries, we drop entries (for authors) in the output file that do not represent names. We further extract, where possible, author names from email addresses to confirm author names.

We then use two auxiliary data sets to further clean and complete the author information. First, we use the author names and organizations of authors on published RFCs. The information is manually collected (Simcoe, 2012) and serves as a reliable source. We merge the RFC author data with our project chain data to correct and complete the name information for project chains associated with an RFC. Second, we use the attendance lists of IETF meetings (meeting 29 in April 1994 through meeting 93 in July 2015) to construct a list of IETF contributors/members that serve as the list of potential authors. For most of the meetings, we obtain author names, affiliations, and email addresses. Email address information allows us to fill gaps in the author information for project chains. In a last step, we conduct basic author-name disambiguation to obtain unique author identifiers.

Using unique author-name identifiers for project chains, we obtain an author's project history at the IETF. For our measure of team size for a project chain, we count the number of authors on the first version of the project. We further construct two measures of author experience. First, for a given project chain, we count an author's number of successfully completed project chains (i.e., published as RFCs) before that project chain is initiated. At the project-chain level, we then use the experience of the most *successful* author of the author team (if multiple authors are listed) of the initial draft. Second, for a given project chain, we count an author's number of completed project chains (i.e., published or abandoned) before that project chain is initiated. At the project-chain level, we then use the experience of the most *prolific* author of the author team (if multiple authors are listed) of the initial draft. Because we have, by construction, all first version documents, author information

is missing for project chains only if the `authorstats` tool is not able to parse the document.

E.2.5   Patent Citations

We use the number of citations from U.S. patents as a proxy for commercial impact. In U.S. patents, IETF RFCs or IDs are cited in the non-patent literature (NPL) references. We obtain the list of non-patent literature references for all U.S. utility patents granted between 1976 and February 2016 from http://www.patentsview.org/download/ (file: otherreference.zip). We count the number of citations of a given project chain from patents in three steps:

> **Step 1:** Using regular expressions that help us take typos and formatting errors into account, we search the NPL references for entries in the form of ID names (e.g., `draft-ietf-homenet-arch`). We pre-sample and use for the search only NPL references that include the terms "IETF," "internet," or "draft" (we use regular expressions to account for differences in capitalization). We use the full sample of ID names (from all project chains) as our search sample. When possible, we also extract the version number that is cited.

> **Step 2:** Using regular expressions that help us take typos and formatting errors into account, we search the NPL references for entries in the form of RFC numbers (e.g., "RFC 7368"). We pre-sample and use for the search only NPL references that include the terms "RFC," "IETF," "internet," or "draft" (we use regular expressions to account for differences in capitalization).

> **Step 3:** As part of the initial download of ID and RFC metadata, we also obtain the titles of all IDs and RFCs. We search the NPL references for these titles, using only titles that have at least four words and 10 characters. We pre-sample and use for the search only NPL references that include the terms "IETF," "internet," "RFC," "Request For Comment(s)," or "draft" (we use regular expressions to account for differences in capitalization).

These three approaches result in overlapping lists of citing patents for a given project chain. We therefore tally the number of unique patents that cite the given project chain in its NPL references. Our approach (counting not only citations to RFC numbers but also including ID identifiers) allows us to capture citations from U.S. patents that were made while the project chain was still active (before publication).

The project chain-level data we obtain from the above procedure is the basis for our patent citation estimates $\hat{\pi}(t)$. For our main results in Table 2, we use the patent citations for successful project chains (published RFCs) in the estimation sample to estimate $\hat{\pi}(t)$. Because for the results presented in all other tables, we do not bootstrap standard errors, we are not restricted to the estimation sample when estimating $\hat{\pi}(t)$. For all other results, we therefore use the extended sample of RFCs

(before taking **Step 6** in Table E.1) and the respective patent citations to estimate $\hat{\pi}(t)$.

E.2.6 RFC Citations

In addition to patent citations, we also count the number of RFC citations each project chain receives. Each RFC contains a list of references that typically contains several citations to previously published RFCs. We download the RFC documents via http://www.rfc-editor.org/rfc/rfc[NNNN].txt (where [NNNN] is the RFC number)[51] and scrape the documents to obtain a list of each citation. We construct a count of RFC citations that is equal to the number of times a given RFC is cited by some other RFC published at a later date. We then link the citations for a given RFC with our project-chain data. Unlike the data for patent citations, our numbers for RFC citations do not include citations to the unpublished RFC (while the project chain was still active).

As with patent citations, we think of RFC citations as an indication that the citing references build on the ideas contained in the cited references, so that citations serve as a measure of the cited RFC's technological importance in a cumulative innovation context. Simcoe (2012) shows that RFC and U.S. patent citations are highly correlated. For our updated data (for both RFC citations and U.S. patent citations), the correlation coefficient is 0.61.

E.2.7 Text Distances

The goal is to calculate the textual distance (or, dissimilarity) of an ID version t from the initial version. For project-level information, we use the distance of the last version, $t = T$, from the initial version, $t = 1$. We use R text-analysis libraries to load and preprocess the documents. Preprocessing steps are: drop punctuation, drop stop words, drop numbers, convert all text to lower case, and eliminate excessive white space. We then create document-term matrices. Each row i of such a matrix represents a document, and each column j represents a term in the corpus (the collection of all documents). The value in cell (i, j) is the number of times a term j is used in document i. A document-term matrix gives us vector-space representation of a document i, where a document is a vector $x_i = (x_{i,1}, \ldots, x_{i,n})$ of term frequencies, with n unique terms in the corpus.

As our measure of textual distance (or dissimilarity) $\text{dist}(T, 1)$ between document $t = T$ and $t = 1$ we use the *cosine dissimilarity* (we refer to it as *cosine distance*). More specifically, we calculate $1 - $ cosine similarity, where the latter is the cosine of the angle between two term-frequency vectors x_T and x_1, so that the cosine distance

---

[51]Our data sample is as of June 2017.

is

$$\text{dist}(T, 1) \equiv 1 - \frac{x_T \cdot x_1}{\|x_T\| \, \|x_1\|} \quad \text{with} \quad \|x_i\| = \sqrt{\sum_{l=1}^{n} x_{i,l}^2}. \tag{E.1}$$

E.3 Emails

We use emails exchanged among IETF contributors (authors and members of the community who provide feedback) to construct two variables:

Email Count: The number of emails sent in response to a given version t of a project chain. At the version level it is, for version t, the number of emails sent in response to the previous version $t - 1$. At the project chain-level level it is, for project chain i, the average number of emails sent in response to any version of i.

Commerciality: The suit-to-beard ratio as the share of corporate email addresses (from which emails are sent) over all email addresses.

We construct our email data in a number of steps.

Step 1: We first download the "IETF Mail Archive" and the "Concluded WG IETF Mail Archive" via ftp://ftp.ietf.org/.[52] These email archives contain all emails sent to various IETF email lists. We obtain the emails in monthly batch files. We use a Python script to first split the monthly batch files into individual emails.

Step 2: We extract the email address from which an email was sent and the date and time it was sent.

Step 3: In order to assign a given email to a given version of a project chain, we search for each ID name (from the complete list of all ID names) in the text of that email. Any email that mentions an ID name is considered to be sent in response to that ID. We also extract the version number of that listed ID name (if possible). The email archive provides us with emails from all email lists, including organizational lists and the `i-d-announce` lists (in its various forms: before 1998, between 1998 and 2004, and after 2004). A new version of an ID is announced to the community at large via this email list. It provides a very clean account of the date the version is posted (with an occasional delay of a day or two relative to the posted date listed on the draft). We do not use ID announcement (or other organizational emails) for our email counts, but use the date information (for each ID version) to fill missing version numbers in the next step.

---

[52] Our email data is as of October 17, 2015.

Step 4: If a version number (for the email assignment) is not available, we take two approaches to fill the gap. First, if an email j without an assigned ID version number was sent between two emails that do have the same version number, we use the ID version number of the two embracing emails. Second, if an email j without an assigned ID version number was sent between two consecutive ID announcement emails, we use the earlier email's ID version number as the assigned ID version number for email j.

Step 5: We delete all ID announcement emails and other organizational emails so as to not include them in email counts. These emails do not contain any information pertinent to the content of the ID.

Step 6: In a last step, we delete all duplicate entries (where the same email is assigned to the same version multiple times). This does not mean that we delete duplicate emails, because an email j can be assigned to two or more project chains. Such an email enters our email database multiple times. We also delete all emails sent before the first ID version was posted and all emails sent more than 24 months after the last ID version was posted.

We have at least one email for 20,532 project chains out of 25,532 project chains in the full sample (a total of 645,064 emails); and at least one email for 12,396 project chains out of 16,091 project chains in the estimation sample. This number amounts to 331,094 emails in the estimation sample.

The emails in the full sample are sent from 20,968 unique addresses (from 7,533 domains). The earliest email in the full sample is from July 6, 1990. The last email is from October 17, 2015. The emails in the estimation sample are sent from 15,404 email addresses (from 5,775 domains). The earliest email in the estimation sample is from January 4, 1996, the last email from October 4, 2015 (assigned to a project chain that was published as RFC).

To construct the suit-to-beard ratio, we categorize email addresses as corporate if their top-level domain (TLD) is `com`, `net`, or `biz`. If the TLD is a country domain, we use `.co.` as an identifier for corporate email addresses. In order to avoid counting email addresses from email provider services, we compile a list of more than 3,000 free (and for-cost) email providers. Many of these use, for instance, `com` as their TLD. We do not categorize such email addresses as corporate.

The suit-to-beard ratio for a given project chain is the number of corporate email addresses from which the project chain's assigned emails are sent, divided by all email addresses of that project chain. We have at least one email address (and are able to construct the suit-to-beard ratio) for 10,710 standards-track projects (out of 14,444 standards-track projects in the estimation sample).

References

AKCIGIT, U., D. HANLEY, AND N. SERRANO-VELARDE (2016): "Back to Basics: Basic Research Spillovers, Innovation Policy and Growth," CEPR Discussion Paper 11707, Centre for Economic Policy Research, London, UK.

AKCIGIT, U. AND Q. LIU (2016): "The Role of Information in Innovation and Competition," *Journal of the European Economic Association*, 14, 828–870.

ALLEN, T. J. (1966): "Studies of the Problem-Solving Process in Engineering Design," *IEEE Transactions on Engineering Management*, 13, 72–83.

ARROW, K. J. (1974): *The Limits of Organization*, New York, N.Y.: W.W. Norton.

AZOULAY, P., J. GRAFF ZIVIN, AND G. MANSO (2011): "Incentives and Creativity: Evidence from the Academic Life Sciences," *RAND Journal of Economics*, 42, 527–554.

BATTISTON, D., J. BLANES I VIDAL, AND T. KIRCHMAIER (2017): "Face-to-Face Communication in Organisations," London School of Economics, unpublished manuscript.

BENNETT, V. M. AND J. SNYDER (2017): "The Empirics of Learning from Failure," *Strategy Science*, 2, 1–12.

BHATTACHARYA, V. (2017): "An Empirical Model of R&D Procurement Contests: An Analysis of the DOD SBIR Program," Massachusetts Institute of Technology, unpublished manuscript.

BLOOM, N., L. GARICANO, R. SADUN, AND J. V. REENEN (2014): "The Distinct Effects of Information Technology and Communication Technology on Firm Organization," *Management Science*, 60, 2859–2885.

BOLTON, P. AND C. HARRIS (1999): "Strategic Experimentation," *Econometrica*, 67, 349–374.

BONATTI, A. AND J. HÖRNER (2011): "Collaborating," *American Economic Review*, 101, 632–663.

BRAY, R. L., D. COVIELLO, A. ICHINO, AND N. PERSICO (2016): "Multitasking, Multiarmed Bandits, and the Italian Judiciary," *Manufacturing & Service Operations Management*, 18, 545–558.

CHAN, T. Y. AND B. H. HAMILTON (2006): "Learning, Private Information, and the Economic Evaluation of Randomized Experiments," *Journal of Political Economy*, 114, 997–1040.

CHING, A. T., T. ERDEM, AND M. P. KEANE (2013): "Learning Models: An Assessment of Progress, Challenges, and New Developments," *Marketing Science*, 32, 913–938.

———— (2017): "Empirical Models of Learning Dynamics: A Survey of Recent Developments," in *Handbook of Marketing Decision Models*, ed. by B. Wierenga and R. van der Lans, New York, N.Y.: Springer.

CRAWFORD, G. AND M. SHUM (2005): "Uncertainty and Learning in Pharmaceutical Demand," *Econometrica*, 73, 1137–1173.

DICKSTEIN, M. J. (2014): "Efficient Provision of Experience Goods: Evidence from Antidepressant Choice," Stanford University, unpublished manuscript.

DORASZELSKI, U., G. LEWIS, AND A. PAKES (forthcoming): "Just Starting Out: Learning and Equilibrium in a New Market," *American Economic Review*.

ELLISON, G. (2002): "Evolving Standards for Academic Publishing: A qr Theory," *Journal of Political Economy*, 110, 994–1034.

ERDEM, T. AND M. P. KEANE (1996): "Decision-Making under Uncertainty: Capturing Dynamic Brand Choice Processes in Turbulent Consumer Goods Markets," *Marketing Science*, 15, 1–20.

ERICSON, R. AND A. PAKES (1995): "Markov-Perfect Industry Dynamics: A Framework for Empirical Work," *Review of Economic Studies*, 62, 53–82.

FERNANDEZ-VILLAVERDE, J. AND J. F. RUBIO-RAMIREZ (2007): "Estimating Macroeconomic Models: A Likelihood Approach," *Review of Economic Studies*, 74, 1059–1087.

GALASSO, A., M. MITCHELL, AND G. VIRAG (2016): "Market Outcomes and Dynamic Patent Buyouts," *International Journal of Industrial Organization*, 48, 207–243.

GALASSO, A. AND T. SIMCOE (2011): "CEO Overconfidence and Innovation," *Management Science*, 57, 1469–1484.

GANGLMAIR, B. AND E. TARANTINO (2014): "Conversation with Secrets," *RAND Journal of Economics*, 45, 273–302.

GILBERT, R. AND C. SHAPIRO (1990): "Optimal Patent Length and Breadth," *RAND Journal of Economics*, 21, 106–112.

GROSS, D. P. (2017): "Performance Feedback in Competitive Product Development," *RAND Journal of Economics*, 48, 438–466.

HAMILTON, B. H., J. A. NICKERSON, AND H. OWAN (2003): "Team Incentives and Worker Heterogeneity: An Empirical Analysis of the Impact of Teams on Productivity and Participation," *Journal of Political Economy*, 111, 465–497.

HEIDHUES, P., S. RADY, AND P. STRACK (2015): "Strategic Experimentation with Private Payoffs," *Journal of Economic Theory*, 159A, 531–551.

HIRSHLEIFER, D., A. LOW, AND S. HONG TEOH (2012): "Are Overconfident CEOs Better Innovators?" *Journal of Finance*, 67, 1457–1498.

HOFFMAN, P. (2012): "The Tao of IETF: A Novice's Guide to the Internet Engineering Task Force," available at https://www.ietf.org/tao.html.

HOPENHAYN, H., G. LLOBET, AND M. MITCHELL (2006): "Rewarding Sequential Innovators: Prizes, Patents, and Buyouts," *Journal of Political Economy*, 114, 1041–1068.

HOWELL, S. T. (2017): "Learning From Feedback: Evidence From New Ventures," NBER Working Paper 23874, National Bureau of Economic Research, Cambridge, Mass.

KELLER, G., S. RADY, AND M. CRIPPS (2005): "Strategic Experimentation with Exponential Bandits," *Econometrica*, 73, 39–68.

KERR, W. R., R. NANDA, AND M. RHODES-KROPF (2014): "Entrepreneurship as Experimentation," *Journal of Economic Perspectives*, 28, 25–48.

KRIEGER, J. L. (2017): "Trials and Terminations: Learning from Competitors' R&D Failures," Massachusetts Institute of Technology, unpublished manuscript.

MAGNAC, T. AND D. THESMAR (2002): "Identifying Dynamic Discrete Decision Processes," *Econometrica*, 70, 801–816.

MOSCARINI, G. AND L. SMITH (2001): "The Optimal Level of Experimentation," *Econometrica*, 69, 1629–1644.

MOSCARINI, G. AND F. SQUINTANI (2010): "Competitive Experimentation with Private Information: The Survivor's Curse," *Journal of Economic Theory*, 145, 639–660.

MURRAY, F., S. STERN, G. CAMPBELL, AND A. MACKORMACK (2012): "Grand Innovation Prizes: A Theoretical, Normative, and Empirical Evaluation," *Research Policy*, 41, 1779–1792.

PAKES, A. (1986): "Patents as Options: Some Estimates of the Value of Holding European Patent Stocks," *Econometrica*, 55, 911–922.

PALACIOS-HUERTA, I. AND A. PRAT (2012): "The Impact Factor of Managers," Columbia University, unpublished manuscript.

ROACH, M. AND W. COHEN (2013): "Lens or Prism? Patent Citations as a Measure of Knowledge Flows from Public Research," *Management Science*, 59, 504–525.

ROBERTS, K. AND M. L. WEITZMAN (1981): "Funding Criteria for Research, Development, and Exploration Projects," *Econometrica*, 49, 1261–1288.

SIMCOE, T. (2012): "Standard Setting Committees: Consensus Governance for Shared Technology Platforms," *American Economic Review*, 102, 305–336.

SIMCOE, T. AND D. M. WAGUESPACK (2011): "Status, Quality, and Attention: What's in a (Missing) Name?" *Management Science*, 57, 274–290.

STEPHAN, P. E. (1996): "The Economics of Science," *Journal of Economic Literature*, 34, 1199–1235.

WEITZMAN, M. L. (1979): "Optimal Search for the Best Alternative," *Econometrica*, 47, 641–654.