An Empirical Model of R&D Procurement Contests: An Analysis of the DOD SBIR Program

VIVEK BHATTACHARYA

Department of Economics, Northwestern University and NBER

Feburary 11, 2021

ABSTRACT. Firms and governments often use R&D contests to incentivize suppliers to develop and deliver innovative products. The optimal design of such contests depends on empirical primitives: the cost of research, the uncertainty in outcomes, and the surplus participants capture. Can R&D contests in real-world settings be redesigned to increase social surplus? I ask this question in the context of the Department of Defense's Small Business Innovation Research program, a multistage R&D contest. I develop a structural model to estimate the primitives from data on R&D and procurement contracts. I find that the optimal design substantially increases social surplus, and simple design changes in isolation (e.g., inviting more contestants) can capture up to half these gains; however, these changes reduce the DOD's own welfare. These results suggest there is substantial scope for improving the design of real-world contests but that a designer must balance competing objectives.

KEYWORDS. R&D procurement, contests, Small Business Innovation Research program, holdup problem, intellectual property.

JEL CODES. C51, C78, O31.

CONTACT. vivek.bhattacharya@northwestern.edu. This is a revised version of Chapter 1 of my PhD dissertation at MIT. I would like to thank Nikhil Agarwal, Glenn Ellison, and Mike Whinston for invaluable advice and support throughout this process. Eric Auerbach, Panle Barwick, Arjun Bhattacharya, Gastón Illanes, Karam Kang, Harry Pei, Rob Porter, Mar Reguant, Jimmy Roberts, Ben Roth, five anonymous referees, and numerous seminar participants provided valuable feedback. Chris Snyder and Elena Krasnokutskaya provided helpful discussions. Conversations with Jonathan Libgober were instrumental in the early stages of the project. I had useful discussions with John Williams and G. Nagesh Rao at the Small Business Administration and Robert Smith at the Office of Naval Research about the institutions and data sources. Nandita Bhattacharya and Anumita Das provided valuable expositional comments. This material is based on work supported by the NSF Graduate Research Fellowship under Grant No. 1122374, the John Krob Castle 1963 Fellowship, and the George P. and Obie B. Shultz Fund. All errors are mine.

1. Introduction

Firms and government agencies frequently use R&D contests to procure products that are not commercially available. Potential suppliers conduct costly research to develop these products, and the procurer offers one of these firms a prize—often a delivery contract. While a sizable theoretical literature provides useful guidelines for the design of such R&D contests, optimal contest design depends on rich primitives of the setting and is fundamentally an empirical question. Despite the prevalence of these mechanisms and the debate in practice over how to structure them,[1] there is little work evaluating observed designs. In this paper, I develop a model of a real-world contest to estimate its primitives, ask by how much alternate designs improve efficiency, and explore why the observed design may deviate from the socially optimal one.

I study an example of a multistage R&D contest: the Small Business Innovation Research (SBIR) program in the Department of Defense (DOD). DOD SBIR provides a useful laboratory for studying contest design in an empirical setting. Data is readily available: the set of competitors is observed at each stage, the size of R&D and procurement contracts is known, and research can be tied directly to the product developed from it, a difficulty in studies of innovation. Competition between firms has the core elements of a contest: firms exert costly effort over multiple stages to improve surplus; research involves uncertainty and generates heterogeneity across contestants, influencing incentives to exert effort; and firms are motivated by the prospect of a prize, which in this case is a procurement contract. It is also an important setting in its own right: the DOD spends over $1 billion a year on R&D through SBIR, leading to almost $500 million in delivery contracts connected to all major defense acquisition programs. Moreover, government procurement of R&D broadly is a large market. In 2014, the US federal government accounted for almost 10% of all R&D expenditures worldwide, with over half attributed to the DOD (Schwartz et al., 2015).

I develop a model that captures the institutional features of DOD SBIR. As I discuss in Section 3, there are multiple sources of inefficiency embedded in the design. First, research can be underprovided due to a holdup effect: the firm only captures part of the surplus it generates through R&D, which gives it less than the socially efficient incentives to exert effort. Second, in this multistage contest, firms have an incentive to try to displace com-

[1] As an example, the Department of Defense has long used contests for the design and delivery of weapons, and there is still policy debate over when to mandate competitive prototyping (Drezner and Huang, 2009). Contests are also relevant outside traditional R&D. Cities often solicit costly proposals from multiple firms when procuring complex construction services. Auto companies ask suppliers to compete by designing custom components but differ from each other in how they structure such competition (McMillan, 1990).

petitors from future stages; this business-stealing effect can lead to overprovision of effort. Finally, the DOD provides R&D contracts that refund expenditures; this reimbursement effect, stemming from the fact that these contracts are socially neutral transfers but are treated as prizes by the firms, also is a force towards overprovision of effort. These sources of inefficiency manifest themselves in general models of contests—and, in the case of holdup and business stealing, in settings outside contests as well.

These inefficiencies, and the parameters underlying them, inform the tradeoffs in contest design that this paper explores. For instance, increasing the strength of incentives provided by the procurement contract ameliorates holdup by allowing the firm to capture more of the surplus; however, it exacerbates business stealing, as firms have a greater incentive to compete for the procurement contract. Which effect dominates depends on the marginal cost of effort in earlier stages as well as the current level of incentives provided to firms. As another example, increasing the number of competitors in a contest is directly beneficial if there is significant uncertainty in R&D outcomes; however, this leads to duplicative effort. Moreover, increasing competition may reduce the marginal benefit of effort, and firms would adjust their effort in response to competition as well; how they do so depends on the shape of the effort function. As a final example, mandating that firms share intermediate research can be beneficial, especially if there is significant heterogeneity in intermediate outcomes of research; however, if early-stage research is costly, firms may choose to cut their effort significantly upon realizing that information sharing reduces the share of the surplus they can capture from their own R&D. The relative magnitudes of these tradeoffs depend on the primitives of the setting—the stochastic map from research effort to components of surplus (values and costs), which captures both the shape of the effort function and the uncertainty in outcomes, and the share of the surplus the firms capture in delivery.

To connect the model to the data, I show in Section 4 that the primitives are identified from data on research spending and the delivery contract amounts. Identification relies on (i) a monotonicity property that firms more likely to generate high-surplus projects also exert more effort, (ii) a selection condition that the procurer never purchases a product that generates negative surplus, and (iii) a condition that research efforts are set optimally. I develop a computationally tractable estimation procedure that mirrors the constructive identification argument. Section 5 shows three robust observations: there is limited heterogeneity across competitors induced by the outcomes of initial-stage research; most of the variation in final outcomes comes from the outcomes of later-stage research; and firms are subject to high-powered incentives, capturing approximately half of the surplus.

Since SBIR is designed at least in part to support innovative small businesses, a natural objective for the DOD is the social surplus generated within each contest: its willingness to pay for the innovation, less total research and delivery costs. With these parameters, I quantify the inefficiencies and evaluate whether the design can be altered to improve social surplus.[2] Section 6 shows the surplus can be improved by about 40% by moving to the socially efficient design even if the number of competitors cannot be changed; the gain is substantially larger if the DOD can invite more contestants. I find this wedge is not merely due to difficulty in implementing the efficient design: arguably simple design changes—just providing competitors moderately higher incentives, increasing the number of competitors, combining the two phases of the contest, or requiring firms to share intermediate progress with competitors—can increase the social surplus up to halfway to the optimum.

These results could be interpreted as prescriptions for the DOD: like in the empirical auctions literature, estimates of the primitives allow me to suggest beneficial design changes. However, given the prospect of such substantial gains, why has the DOD not adjusted contest design to realize them? In Section 6.3, I consider a variety of explanations, including alternate specifications of the R&D process and the possibility of incentives not captured by the contracts I observe. The explanation with the most empirical support is that the DOD may balance social surplus with its own private welfare: on all dimensions, the observed design lies between the socially optimal one and the DOD-optimal one. Overall, this paper concludes that contest design can have a significant effect on outcomes in a real-world setting, but a procurer may be balancing competing objectives when determining the design.

Related Literature. The relationship between competition and innovation in markets is ambiguous.[3] Similarly, the theoretical literature on R&D contests also concludes that the optimal design of contests is an empirical question.[4] This paper builds off this theory

[2]Quantifying these inefficiencies is an incidental contribution of this paper. Berry and Waldfogel (1999) and subsequent work estimate the inefficiency generated by business stealing. Holdup is present in many settings and considerable work documents its effects on contracting (Joskow, 1987; Crocker and Masten, 1988); Bresnahan and Levin (2012) and Lafontaine and Slade (2012) provide surveys. However, I am aware of very few estimates quantifying the inefficiency it generates. (For exceptions, see Zahur (2020) in industrial organization, and Iyer and Schoar (2015) and Bubb et al. (2018) for field experiments in development.)

[3]Classic theoretical analyses provide competing views (Schumpeter, 1939; Arrow, 1962; Gilbert and Newbery, 1982). Vives (2008) highlights the importance of the demand parameterization and the notion of "competition" and provides a review. Empirical evidence is mixed: for instance, Blundell et al. (1999) show a positive relationship between competition and innovation, while Aghion et al. (2005) document a nonmonotone one.

[4]Researchers have studied the effects of the level of competition (Taylor, 1995; Fullerton and McAfee, 1999; Che and Gale, 2003; Koh, 2017), the reward structure (Moldovanu and Sela, 2001; Cohen et al., 2008), the number of stages (Moldovanu and Sela, 2006), and information sharing (Bhattacharya et al., 1990). Che et al. (2017) and Liu and Lu (2018) model the prize as a procurement contract and add asymmetric information;

literature to evaluate observed design. It extends the "Laffont program," coined by Hendricks and Porter (2015) in auctions, to contests: just as the empirical auctions literature exploits theory to estimate primitives (values) and compute the revenue and efficiency gains from changing observed auction design, this paper estimates primitives of the contest (e.g., the map from effort to surplus) to compute the gains from alternate contest designs.

The small handful of empirical papers taking contest models to data share this broad goal. To my knowledge, they have focused exclusively on the setting of online "ideation" contests, where a large set of individuals offer solutions to programming (Boudreau et al., 2016), prediction (Lemus and Marshall, 2020), or creative design problems (Kireyev, 2020; Gross, 2020). This paper considers a starkly different setting: a large contest assumption is not applicable, the contest has multiple stages that can improve surplus, and the prize comes in the form of a procurement contract. These features are realistic in many settings, and they require different modeling decisions to estimate incentives from the data at hand.

A small literature on SBIR broadly has studied its long-term effects on small businesses (Lerner, 2000; Wallsten, 2000; Howell, 2017). DOD SBIR has a somewhat different focus, as it places significant emphasis on acquiring specific technologies. Thus, I complement this literature by studying how the design of the program itself can be altered to increase the surplus generated within the program, but I would be hesitant to extend these results to other federal agencies. Rather, variations of this model may be suited for studying design-and-build contracts or other forms of procurement with a research stage.[5]

Defense procurement, with few exceptions (Carril and Duggan, 2020), has been largely ignored by empiricists. An advantage of studying SBIR relative to other DOD procurement is that projects are smaller in scope, the goals are well-specified, and asymmetric information about values and costs is arguably much less of an issue than in the procurement of major weapons systems. Yet, SBIR retains salient features of defense procurement (Rogerson, 1994, 1995; Lichtenberg, 1995). Defense procurement involves contracting for both R&D and delivery, and the DOD often considers multiple prototypes before narrowing the competition for delivery (Anton and Yao, 1989, 1992; Lyon, 2006). Contracts are structured so that firms earn economic profits, providing them incentives for investment in early stages (Rogerson, 1989). The DOD may contract with separate firms for innovation and delivery. The simple design changes I consider speak to all these methods for controlling incentives.

the latter focuses on the single-agent case. Cabral et al. (2006) and Williams (2012) provide surveys.

[5]Examples include the European Commission's EIC Accelerator Pilot, the EU's Pre-Commercial Procurement initiative, and the UK's Small Business Research Initiative. These programs have structures similar to DOD SBIR and place varying degrees of emphasis on procurement.

2. Empirical Setting and Data

The SBIR program provides small firms funding to commercialize early-stage research projects, either on the private market or to the government. Federal agencies with extramural R&D budgets of more than $100 million must allocate approximately 3% of it through this program. The DOD posts 150–250 solicitations for specific research projects each year for each main service—including a description of the required technology, technical requirements, goals for the three phases detailed below, and plans for delivery to the DOD. The Navy in particular keeps careful track of implementation and delivery contracts that result directly from R&D funded by SBIR, allowing me to track a technology from concept to acquisition. Further details and documentation are in Appendix E.

Unlike non-defense agencies, the Navy almost always uses SBIR to solicit research on technologies that it wishes to acquire.[6] Almost all solicitations are connected to specific acquisition programs (e.g., Virginia Class Submarine). Over 80% of the topics solicited by the Navy are developed by Program Execution Offices (PEOs), and the PEO and the prime contractor make most evaluation and contracting decisions themselves.[7] The solicited products are thus fairly specific to military applications and often are small components of major weapons systems; given this specificity, the private market for technologies produced through the DOD SBIR program is much more limited than with other agencies.

Firms interested in competing for a *Phase I* contract submit a technical proposal. Upon evaluating these proposals, the DOD awards Phase I contracts to some of the firms. The Navy SBIR Overview describes Phase I as "a feasibility study to determine the scientific or technical merit of an idea or technology that may provide a solution to the [Navy]'s need or requirement." It involves benchtop testing, computer simulations, and other low-cost preliminary research. The Navy awards approximately $80,000 for the base Phase I contract, with little variation across competitors and projects. After about six months, the firms submit reports detailing findings, a *Phase II* proposal that includes plans to implement or manufacture the product designed in Phase I, and a detailed cost proposal for Phase II research. The DOD evaluates the proposals primarily on technical merit and excludes consideration of the proposed cost of Phase II research. The targeted number of Phase II contestants is about 40% of the number of Phase I awards, although the DOD may award Phase II contracts to fewer firms.[8]

[6]Examples include a "Compact Auxiliary Power System for Amphibious Combat Vehicles" (Advanced Amphibious Assault) and "Air Cushion Vehicles Lift Fan Impeller Optimization" (Ship-to-Shore Connector).
[7]See http://www.navysbir.com/natconf14f/presentations/3-09-Navy-Comm-Williams.pdf.
[8]The Phase II desk reference (http://www.acq.osd.mil/osbp/sbir/sb/resources/deskreference/12_phas2.shtml)

The Navy SBIR Overview describes Phase II as "typically a demonstration phase in which prototypes are built and tested," and Phase II awardees conduct intensive research to assess commercial viability. Contracts are larger and vary considerably both across projects and across competitors within a project: the Navy guidelines note that Phase II is structured in a way "that allows for increased funding levels based on the project's transition potential." The firm submits progress reports and a final report after about two years. Finally, unlike many other federal agencies, the DOD SBIR process includes a formal Phase III, which is the final goal of most firms involved in these contests. Phase III is essentially a delivery phase in which the firm either implements or produces the technology developed in Phases I and II for the DOD or prime contractors through a DOD contract. Few contests result in a Phase III contract. While SBIR requirements do not stipulate that only one firm can be awarded a Phase III contract, this is almost always the case in practice: technologies developed by Phase II competitors are sufficiently substitutable that the DOD has value for at most one. This provides the fundamental source of competition in each contest.

I collect information about all SBIR contracts awarded by the Navy from the Navy SBIR Program Office via www.navysbirsearch.com. This data includes firm information, including name and location; the topic number, which maps contracts to solicitations and contests; the systems command ("SYSCOM") in charge of the contract; the phase of the contracts; and the dates of execution. It also includes the title and keywords of the proposal from the firm and an abstract of the project. I match this data using the contract number to the Federal Procurement Data System (via www.usaspending.gov) and extract information for each contract. In particular, the FPDS contains information about all options exercised and all modifications for each contract, which allows me to compute the total amount awarded to the firm through the contract. I restrict the analysis to contests between 2000 and 2012. To control for heterogeneity across projects, I collect the full text of all Navy SBIR solicitations from www.sbir.gov. This text in these solicitations allows me to construct detailed project-level covariates to control for the topic of the contest via an unsupervised Latent Dirichlet Allocation (LDA) algorithm for topic modeling (Blei et al., 2003). Further details about data collection and cleaning are in Appendix A.1.

These contests are small, and participation is limited. Table 1 shows that about 75% of contests in the dataset have 2 or 3 Phase I competitors, and fewer than 4% have more than

notes, "[DOD] anticipates that at least 40% of its Phase I awards will result in Phase II projects. This is merely an advisory estimate and [DOD] reserves the right and discretion not to award to any or to award less than this percentage of Phase II projects."

	0	1	2	3	4	≥ 5	# In Phase	Pr(Advancing)
# Phase I Comp	–	12.9%	41.8%	32.8%	8.9%	3.6%	2875	83.1%
# Phase II Comp	16.9%	61.1%	19.0%	2.3%	0.6%	0.2%	2390	10.5%
# Phase III Comp	91.2%	8.8%					252	

Table 1: Distribution of the number of competitors in each phase, with the number of contests in each phase and the probability of a contest advancing to the next phase conditional on entering the current one. I restrict to solicitations posted between 2000 and 2012 and only consider ones with at most one Phase III contract.

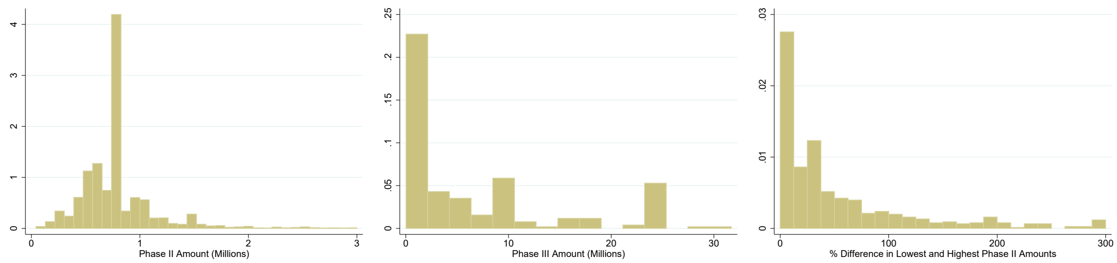

Figure 1: Distribution of (a) Phase II award amounts and (b) Phase III award amounts. The histogram in (a) includes a datapoint for each contract and can thus include multiple contracts for a particular contest. Panel (c) shows the percent difference between the highest and lowest Phase II award amounts within contests, restricting to contests with at least two Phase II competitors.

4. Over 80% of contests proceed to Phase II, but about three-fourths of contests that enter Phase II have only one competitor. Firms are awarded intermediate R&D contracts in these phases, but Table 1 highlights that the main source of revenues—a Phase III contract—is especially rare: only about 11% of contests that enter Phase II lead to a Phase III contract.

There is also substantial variation in the size of R&D and delivery contracts. Phase II contracts generally lie between \$250,000 and \$1.5 million (Figure 1(a)).[9] Figure 1(b) shows that Phase III contracts have a long right tail and can exceeded \$15 million. This variation is not just due to cross-contest heterogeneity in the types and costs of technologies but also to variation in outcomes within-contest. To highlight this variation, Figure 1(c) restricts the sample to contests with at least two Phase II competitors and plots the percent difference between the contract amounts for the firms with the largest and smallest contracts, within-contest. This comparison, which perfectly controls for contest heterogeneity, shows that differences in outcomes within-contest can be substantial: the best-funded competitor often receives more than 50% more funding than the worst-funded competitor.

Appendix A presents more comprehensive summary statistics and descriptive correla-

[9] There is a peak around \$750,000 for Phase II contracts, the standard amount for Phase II SBIR contracts in other agencies that is sometimes used as a baseline by the Navy. I have not found any systematic explanation for these contracts. The distribution of residuals when controlling for project-level effects shows no masses.

tions. The model I introduce in Section 3 allows me to move beyond such a descriptive analysis in three ways. First, it provides a framework for interpreting the transition rates and contract amounts presented above in welfare-relevant terms. For instance, it interprets the rarity of a Phase III contract as suggesting that the probability of generating a positive-surplus project is low, and it interprets the within-contest variation in research contracts as indicative of the value to the Navy of each competitor's project.[10] More importantly, a larger Phase III contract by itself is not necessarily indicative of larger surplus (as it could be due to the DOD acquiring a higher-value product or due to the DOD having to cover the delivery costs of a costlier product), and the model provides a structure for estimating the surplus generated by the contests from this data. Second, contests in the data with different numbers of competitors could be different, as the DOD chooses how many contestants to let into Phase I. As I show in Section 4.1, the model controls for such differences, which allows me to estimate the effect of changing competition on outcomes without relying on exogenous variation in competition. Finally, the model of course allows for simulating alternate counterfactual contest designs, including an optimal one.

3. Model

3.1. Model Setup

Each SBIR contest consists of three phases. The primitives are the number of contestants in Phase I (N_1), the maximum number allowed to enter Phase II ($\bar{N}_2$), the distribution from which firms draw values (V), the cost functions ($\psi(\cdot)$ and $H(\cdot;\cdot)$), the share of the surplus the firm captures in Phase III (η), and the Phase II contract amount $t_{N_2}^{DOD}(v)$ the DOD offers given the value and the number of Phase II competitors N_2.

Phase I. Phase I is a prototyping phase in which firms exert effort to determine both the feasibility and the potential value of the innovation, i.e., the DOD's willingness to pay if the innovation is delivered. Given the technologies are specific to the acquisition program, I maintain in the baseline that this value is the sole source of surplus generated in the contest (with delivery and research costs, discussed below, detracting from surplus).[11]

[10]This interpretation is consistent with the DOD's claim that it gives more funding to projects that have increased transition potential and with the evidence in Table A.3 in Appendix A that these projects are indeed more likely to lead to Phase III contracts. An alternate interpretation that attributes this variation to heterogeneity in research cost would not immediately be able to explain this correlation.

[11]These products are generally improvements to existing capabilities of acquisition programs rather than crucial components; the DOD's outside option is usually to simply not have this improved capability. Thus, the value should be interpreted as the added value to an existing weapons system from this capability.

The DOD invites N_1 identical firms to Phase I. If firm i spends the dollar amount $\psi(p_i)$ ($\psi(\cdot) > 0$, $\psi'(\cdot) > 0$, and $\psi''(\cdot) > 0$) on Phase I, it generates a successful innovation with probability p_i, independently of all competitors. The N_S firms that succeed each independently draw a value $v_i \sim V$ with cdf F. This structure parsimoniously models solicitations as having narrowly specified guidelines, and the difficulty is in meeting guidelines rather than exerting effort to exceed them. Foreshadowing results, I estimate a narrow distribution of values, consistent with limited variation in values conditional on meeting guidelines.[12]

At most $\bar{N}_2$ of the N_S firms that succeed are allowed to proceed to Phase II. That is, if $N_S \leq \bar{N}_2$, then all firms that succeed enter Phase II. If $N_S > \bar{N}_2$, then the $\bar{N}_2$ firms with the highest draws of v proceed. A contest fails in Phase I if none of the participants succeed. Firms are aware of the number of Phase I competitors N_1, the limit $\bar{N}_2$ on the number of Phase II competitors, and the primitives of the contest (F, η, $\psi(\cdot)$, and $H(\cdot;\cdot)$).

Phase II. The goal of Phase II is to develop a commercially viable production plan; that is, firms conduct research to reduce the delivery cost (e.g., manufacturing cost for physical products or implementation cost for software) of their innovation. In Phase II, each firm spends some dollar amount t: exerting effort t results in a draw of the delivery cost c from a distribution with cdf $H(\cdot; t)$ and density $h(\cdot; t)$. This distribution is first-order stochastically decreasing in t: more effort corresponds to drawing lower delivery costs. A contest fails in Phase II if all participants draw costs that exceed their values.

The DOD also provides R&D contracts; since the DOD claims that it provides higher funding to projects with greater transition potential, I model this as a function $t^{DOD}_{N_2}(v)$ increasing in v. A firm knows its own value v_i and the number N_2 firms that entered Phase II. However, it knows neither the number of successes N_S nor its opponents' values.[13] It forms beliefs (with cdf $F(\cdot; v_i, N_2, p)$) about its opponents' values, where p is its belief of the Phase I effort its competitors exerted. (I restrict to a symmetric equilibrium, so I only consider the case where firms believe that all opponents exerted the same effort p.) If $N_2 < \bar{N}_2$ or $N_2 = N_1$, then every firm that succeeded was granted entry into Phase II. Thus, all firms know that the values of their opponents are drawn from V. If $1 < N_2 = \bar{N}_2 < N_1$, then there is selection into Phase II, and a firm with value v believes that the other $\bar{N}_2 - 1$

[12] Another justification for this structure is practical: identifying a more flexible dependence $F(\cdot; p)$ of values on effort would require variation in equilibrium Phase I effort. Such sources of variation are limited: the main source consistent within the model would be variation in N_1 and $\bar{N}_2$. Since in estimation I allow F to depend on N_1, we would be limited to variation in $\bar{N}_2$ fixing N_1, which is minimal in this dataset.

[13] Firms know their technical score from Phase I. They can interact with the DOD Technical Point of Contact about their proposal and develop a coherent understanding of the DOD's evaluation. The information structure captures this relatively open dialogue and provides a simple benchmark for analysis.

players have values $\mathbf{v}_{-i}$ given by $f_v(\mathbf{v}_{-i}; v, \bar{N}_2, p)$, detailed in (B.5) in Appendix B.3.

Phase III. The DOD contracts with at most one of the firms to deliver the product. The DOD sees the realization (v_i, c_i) for all firms and contracts with the firm that generates the highest surplus (value of $v - c$), as long as it is positive. The DOD offers the winner a dollar transfer $T_3 = c + \eta(v - c - s)$, where s is the second-highest value of $(v - c)^+$, so that the winner's profit is a share η of the incremental surplus it generates. One could view η as a reduced-form way to capture the strength of the incentives from the procurement contract. Alternatively, one could view this as the outcome of a Nash bargaining process between the DOD and the winning firm, where the DOD's outside option is to go to the firm with the second-highest surplus and extract all its surplus.[14]

Equilibrium. Focus on Phase II with N_2 entrants. Consider a firm with value v and beliefs with pdf $f_v(\cdot; v, N_2, p^*)$ about its opponents' values. Suppose its opponents follow an effort function $t^*_{N_2}(v)$. The firm's problem for $t^*_{N_2}(v)$ is

$$\arg\max_t \left\{ \eta \int_{\underline{c}}^{v} \int_{-\infty}^{v-c} (v - c - \max\{s, 0\})\, dG(s; v, t^*_{N_2}(\cdot), p^*)\, dH(c; t) - t + t^{DOD}_{N_2}(v) \right\}, \qquad (1)$$

where $G(s; v, t^*_{N_2}(\cdot), p^*)$ is the cdf of a type v firm's beliefs about the highest surplus of its competitors. The cdf of the surplus that a type v' firm generates is $S(s; v', t^*_{N_2}(\cdot)) = 1 - H(v' - s; t^*_{N_2}(v'))$, and the cdf of the maximum surplus of a type-v firm's opponents is

$$G(s; v, t^*_{N_2}(\cdot), p^*) \equiv \iint_{\mathbf{v}_{-i}} \left(\prod_{v_{-i} \in \mathbf{v}_{-i}} S(s; v_{-i}, t^*_{N_2}(\cdot)) \right) f_v(\mathbf{v}_{-i}; v, N_2, p^*)\, d\mathbf{v}_{-i}. \qquad (2)$$

In Phase I, each firm chooses p to maximize the expected profits from Phase II, less the cost of Phase I effort. Since the expected profits from Phase II can be expressed as p times the profits conditional on success, we can write the firm's problem in Phase I as

$$p^* = \arg\max_{p \in [0,1]} \left\{ p \cdot \left[\sum_{N_S=0}^{N_1-1} \binom{N_1 - 1}{N_S} (p^*)^{N_S} (1 - p^*)^{N_1 - N_S - 1} \int_0^{\bar{v}} \lambda(v, N_S, \bar{N}_2) \pi(v, N_S, p^*)\, dF(v) \right] - \psi(p) \right\}, \qquad (3)$$

where $\lambda(v, N_S, \bar{N}_2)$ ((B.6) in Appendix B.3) is the probability that a successful firm with value v enters Phase II if N_S other firms succeed, and $\pi(\cdot)$ is the Phase II profits this firm

[14] About 75% of contests that enter Phase II have only one firm, and in general the low success rate suggests it is unlikely that multiple firms develop successful innovations. Thus, the precise extension of Nash bargaining to multiple parties is not especially relevant empirically. One could consider alternate models (Shaked and Sutton, 1984; Bolton and Whinston, 1993), or a bargaining procedure in which the DOD negotiates with the highest-value party instead of the highest-surplus party first. Many of these models still respect the monotonicity property derived in Section 3.3, but they do change incentives (by a small amount).

expects. A type-symmetric Bayesian Nash equilibrium of the R&D contest is a p^* and a set of effort functions $\{t^*_{N_2}(\cdot)\}_{N_2 \leq \bar{N}_2}$ that simultaneously satisfy (1) and (3).

3.2. Discussion of Modeling Assumptions

In this section, I comment on the assumptions of the model. Section 4.4 discusses extensions and robustness checks for the assumptions highlighted here.

First, the model can be specified without imposing any structure on $t^{DOD}(v)$, treating it as a primitive. However, an important assumption in the baseline estimation is that the research contract the DOD offers is firm-optimal: $t^{DOD}(v) = t^*(v)$. Since $t^*(\cdot)$ is increasing, as I show in Section 3.3, this assumption is consistent with the DOD's assertion that it gives more funding to projects with greater transition potential. Moreover, the DOD has extensive experience contracting with similar small firms, so it has a sense of how much the firm would be willing to spend. The DOD would be hesitant to offer the firm more funding than the optimal amount: given the DOD can only imperfectly monitor the firms, the firms can redirect excess resources to other ongoing projects. Finally, the DOD encourages firms to participate in the defense industrial base through this program, and as such, it may prefer to limit ex-post losses. Phase III is sufficiently rare that even though firms have positive expected profits, they would often incur losses if the DOD awarded less than the firm-optimal amount. Thus, if the DOD has limited monitoring capacity and does not wish for firms to incur ex-post losses, it must reimburse the firm-optimal cost.[15]

Second, all incentives in the model are governed by the relationship between effort and surplus: whether the surplus is generated by higher value or lower cost has no import. Nevertheless, the separation between values and costs is important because while all incentives depend on the difference between these quantities, the observed procurement contract depends on both of them positively, and the surplus is not a sufficient statistic to map to the data. The timing of the baseline model—that values are drawn early in the process and costs are drawn later—mirrors the structure of the SBIR program: the Phase I technical report is a validation of design specifications for the product, and planning for how to deliver the product (which would correspond to drawing delivery costs c) usually does not even begin until Phase II.

[15]Institutionally, Phase II is a cost-plus contract in which the DOD reimburses all reasonable R&D costs. The firm submits a detailed cost proposal to the DOD for Phase II research, and the DOD can approve the amount or propose modifications. The setup, together with $t^{DOD}(v) = t^*(v)$, captures a contract that reimburses reasonable costs without explicitly detailing how the DOD disciplines excess spending.

Third, the model considers each contest in isolation. As DOD SBIR is centered around procurement of specific technologies rather than funding R&D in a general topic area (like other SBIR programs), the technologies are tailored to the needs of specific acquisition programs. Moreover, some survey evidence discussed in Appendix E suggests that commercialization outside the DOD is somewhat limited. Thus, the focus on surplus generated within a contest is a reasonable baseline approximation to payoffs.

Fourth, I mention in Appendix E that the numbers of firms competing for entry into a SBIR contest on average exceeds N_1. That is, the DOD generally restricts entry into contests itself and in most situations could have let more contestants enter. While this has no bearing on the formulation and estimation of the model, it supports my assumption in the baseline counterfactuals that the DOD has the ability to find (up to 4) similar contestants.

3.3. Properties of the Model

The model has a monotonicity property (Proposition B.1). If each firm's beliefs about its opponents' values are independent of its own value, then $t^*_{N_2}(\cdot)$ is increasing: firms with higher values in Phase II exert more effort. This result traces back to the fact that higher-value firms have a higher probability of winning and a higher surplus conditional on winning, and the marginal winner earns zero profits.[16] While this property is not a feature of all contest models, it is desirable in this setting since it is consistent with the DOD's statement that it gives higher Phase II funding to projects with greater transition potential—which I interpret as higher v. It also is a key element of identification.

I outline the sources of the inefficiencies embedded in the baseline design. In Phase II, the only inefficiency is the holdup problem: when choosing R&D investment, the firm knows that during Phase III contracting, the DOD will only let the firm capture part of the surplus. As a result, effort in Phase II is underprovided relative to the social optimum. This result is clear for contests with $N_2 = 1$. Proposition B.4 shows that even when $N_2 > 1$, the social planner's optimum is supportable by the firms in equilibrium if $\eta = 1$: in Phase II, the winner's profit is a fraction η of the incremental surplus it generates, and this is exactly the winner's marginal contribution to social surplus when $\eta = 1$. Overall, holdup is a force that inefficiently reduces effort, and η governs how strong it is.

[16] Firms' beliefs about opponents are independent of values if there is no selection into Phase II ($N_2 < \bar{N}_2$ or $N_2 = N_1$). If firms' beliefs vary with values, then firms with weaker values tend to believe their opponents are weaker as well. This could encourage them to exert more effort than firms with higher values. However, I have not been able to find any example where the computed equilibrium is nonmonotone.

Holdup feeds through to Phase I and is a force towards inefficiently low effort in this phase as well, but two other forces counterbalance it. The first is what I term the reimbursement effect: the DOD reimburses costs in Phase II. The R&D contract $t^{DOD}(\cdot)$ is a socially neutral transfer, but firms treat it as a prize when choosing effort in Phase I, leading to overprovision of Phase I effort. This is a deliberate feature of the design, and the model reflects that: Rogerson (1994) notes that R&D funding is a method the DOD uses in other procurement settings to counteract the holdup problem. The second is a business-stealing effect from Mankiw and Whinston (1986): when setting research efforts in Phase I (with $\bar{N}_2 < N_1$), a firm exerts effort knowing it captures the full surplus from displacing a rival from Phase II rather than just the marginal, leading to overprovision of R&D relative to the social optimum. It is thus ambiguous whether effort is underprovided in Phase I.

4. Identification and Estimation

4.1. Identification

We observe the numbers of players N_1 and $\bar{N}_2$; the realized number of Phase II players N_2; the distribution of Phase II research contracts t^{DOD} (which I assume in the baseline coincides with $t^*_{N_2}$) for $N_2 \leq \bar{N}_2$; and the Phase III contract amount, if any. The primitives to identify are the cost function $\psi(\cdot)$, the value distribution V, the cost distribution $H(\cdot; t)$ as a function of Phase II research efforts, and the bargaining parameter η. We identify the Phase II and III primitives (everything except ψ) using (i) a selection equation stipulating delivery in Phase III occurs if and only if the winner's value exceeds its cost, (ii) monotonicity of the Phase II effort in the value to recover values from effort, and (iii) a first-order condition ensuring that Phase II research effort is set optimally, with knowledge of η.

The argument I provide is constructive. Fix N_1 and restrict attention to contests with $N_2 = 1$. Figure 2 illustrates the joint distribution of Phase II and Phase III contracts t_2 and T_3. The dots are sample points, and the shaded area (discussed below) is the support. Walking through this figure clarifies identification, and Appendix B.2 formalizes this discussion.

Consider the distribution of the Phase III transfers (vertical axis) conditional on a particular value t_2 of Phase II research (a point on the horizontal axis). Given monotonicity, this amounts to conditioning on some (yet unknown) value $v(t_2)$, given by the inverse of the effort function $t^*(v)$. The Phase III transfer is $\eta v(t_2) + (1-\eta)c$, where $c \sim H(\cdot; t_2)$, but it is only observed if $c \leq v(t_2)$ so that delivery of the product would generate positive surplus. This means that all observed Phase III transfers would be in the shaded region in the figure.

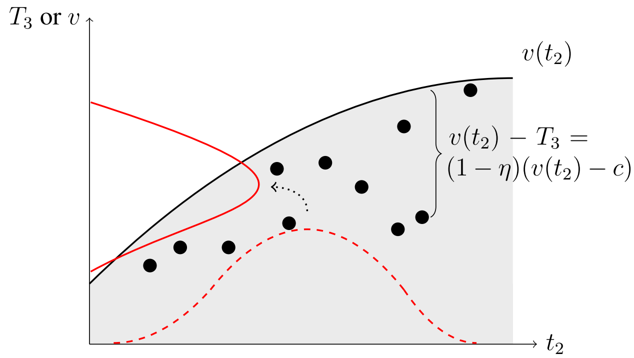

Figure 2: Identification of Phase II parameters using the joint distribution of Phase II and Phase III contracts. The axes plot Phase II R&D contracts t_2 and Phase III procurement contracts T_3.

The largest observed value of the Phase III transfer for a particular value of t_2 occurs when $c = v(t_2)$. Thus, the maximum of the support of the Phase III transfer identifies $v(t_2)$.

Repeating this argument for different values of t_2 identifies the entire function $v(\cdot)$, illustrated by the solid black line. The distribution of t_2 is observed (illustrated by the dashed red line), and thus the distribution of the values of competitors who enter Phase II (solid red line) is identified nonparametrically by mapping it through the now-identified function $v(\cdot)$. If there is no selection into Phase II (i.e., if $\bar{N}_2 > 1$ or $N_1 = 1$), then this is simply the distribution of V. Otherwise, we can correct for selection to recover the distribution of V.

If η is known, the distribution of costs $H(\cdot; t)$ is identified as a function of η. Consider a specific observed Phase III transfer, illustrated by a black dot. Since we observe the Phase II contract t_2, we know the value $v(t_2)$. With this knowledge and that of η, we can identify the cost c for this contract. We do this for all contracts in the gray region and determine the cost distribution as a function of η for $c \leq v(t_2)$. Denote its pdf by $h(\cdot; t_2, \eta)$. In short, the residual variation in the observed Phase III transfers, conditional on the Phase II contract, gives the distribution of the cost draws and thus the distribution of the generated surplus.

Thus far, we have not used optimality of the transfer, but the model imposes that the firm sets t_2 in response to its first-order condition, i.e.,

$$\eta \int_{\underline{c}}^{v(t_2)} (v(t_2) - c) \frac{dh}{dt}(c; t_2, \eta)\, dc = 1. \tag{4}$$

We use (4) to identify η: from monotonicity and selection, we have identified the marginal benefit of a dollar of research (conditional on η), and any wedge between this and the marginal cost (i.e., a dollar) must be due to the firm not capturing the full surplus through η.

I consider identification of the Phase I cost of research $\psi(\cdot)$ without assuming that

Phase I research efforts are observed; since Phase I contracts exhibit no variation and are institutionally set across all SBIR programs, I avoid using them in the baseline. From $H(\cdot;\cdot)$, V, and η, we can compute $\pi(v, N_S, p)$ for all values v, realizations of N_S, and p. These quantities allow us to compute the expected profit conditional on success for any p; denote this $\pi(p)$. Since the distribution of N_2 is truncated binomial with parameters N_1 and success probability p^* (truncated at $\bar{N}_2$), p^* is directly identified. The firm's first-order condition associated with (3) in Phase I, $\psi'(p^*) = \pi(p^*)$, lets us identify the marginal cost of Phase I research at one point. Variation in $\bar{N}_2$ fixing N_1 identifies more points.

I make four points about identification. First, identification of the Phase II parameters does not leverage variation across N_1 and $\bar{N}_2$. This is an important strength. I cannot make comparisons of contests with different levels of competition since such contests may simply be unobservably different. Because the model has empirical content even fixing $(N_1, \bar{N}_2)$, in estimation I let some parameters vary flexibly on this dimension, and the structure of the model lets me evaluate the effect of competition on outcomes.

Second, the model interprets larger variation in Phase III contracts as indicative of larger variation in surplus.[17] This holds since the residual uncertainty in Phase II is only in costs; it would also hold in an alternate (but less realistic) model in which the residual uncertainty is only in values. However, this observation clarifies the limits of the argument. If Phase II research affects both values and costs, larger variation in Phase III contracts might not map to larger variation in surplus: a low transfer indicate a drop in values (low surplus) or a drop in costs (high surplus). Here, the joint distribution in Figure 2, together with monotonicity and selection conditions, is not sufficient to identify the variation in surplus generated in Phase II, and we would need more structure to identify this more general model.

Third, the optimality condition is an important source of the model's empirical content. To my knowledge, exploiting an ex-ante investment is a novel source of identification for a bargaining parameter that could be applicable to other settings with R&D: this strategy leverages the holdup problem, since if the firm is "underinvesting" by a large margin, we would expect that it is unable to recover much of the generated surplus. Moreover, this

[17] Variation in outcomes at the end of Phase I yields differential incentives for research, which allows for identification of the response of surplus to research efforts. This observation highlights that the fact that $H(\cdot;\cdot)$ does not depend directly on v is important for identification. This assumption is reasonable in this empirical setting, in which projects are similar and there likely are not drastically different approaches to research. Otherwise, we would need access to a source of variation that changes either t while holding v constant, or vice versa. An instrument for research costs would be an example, but variation in N_2 is in principle another example. In fact, since costs and value do not change based on the realization of N_2, this source of variation is also implicitly used in estimation.

optimality condition can inform more than just the bargaining power in more general models. In the model sketched in the previous paragraph, where both values and costs could change with Phase II research, a shock to values in Phase II (resulting in either a lower or a less certain distribution of surplus) would affect the marginal benefit of exerting effort. Firms are setting efforts so that this marginal benefit equals marginal cost, so any departure can help inform the prevalence of these shocks to value.[18]

Finally, this discussion is deliberately simple, to clarify the map between the model and data. The elements of Figure 2 are not the only source of information about parameters. In practice, the pattern of the failure rate as a function of t_2 (not illustrated) also informs estimates, as does variation in the realized number of Phase II contestants for otherwise similar contests. Moreover, overidentifying restrictions in (4) and knowledge of the failure rate allow for identification of extensions—including multiplicative unobserved heterogeneity, which I use in the empirical implementation. (See the end of Appendix B.2 for a discussion.)

4.2. Empirical Model

For each contest, I observe the number N_1 and N_2 of contestants in Phase I and II. I restrict to contests with $N_1 \leq 4$ and infer $\bar{N}_2$ from the 40% rule of thumb provided by the DOD SBIR program: I assume $\bar{N}_2 = 1$ if N_1 is 1 or 2, and $\bar{N}_2 = 2$ if N_1 is 3 or 4. If N_2 exceeds the candidate $\bar{N}_2$, I set $\bar{N}_2 = N_1$. The Phase III contract amount, if any, is observed and maps to the bargaining transfer of $c + \eta(v - c - s)$. The Phase II contract amount maps to $t^*_{N_2}(v)$. I do not use Phase I research contracts in the baseline estimation.

I add two components to the model in Section 3 to take it to the data: (i) observed covariates that affect values, costs, and the costs of research and (ii) heterogeneity unobserved to the econometrician that affects all these quantities. In particular, each contest j is characterized by a set of covariates X_j and an unobserved shifter $\theta_j \sim \Theta$, where $\log \Theta$ is normalized to have mean zero. A particular firm i in a contest j has value v_{ij}, cost

[18] An argument exploting optimality is different from ones used in other empirical bargaining papers. Grennan (2013) identifies the bargaining parameter roughly by comparing distributions of transfers that are generated by different value distributions ("added value" in his paper) but similar cost distributions: if the transfer distributions change dramatically, then the effect of the value on the transfer—governed by η—would be high. Crawford and Yurukoglu (2012) match the model-implied outcomes to estimated outcomes with auxiliary knowledge about one of the components of the transfer (i..e, that channel input costs are zero). I do not have similar knowledge, because delivery costs are nonzero and unobserved in my setting, but unlike Crawford and Yurukoglu (2012), I can leverage the optimality of the investment that I do observe. Relatedly, note that in an extension discussed in Section 4.4, optimality informs both η and the distribution of shocks to value in Phase II; I leverage the overidentifying restrictions in the FOC to estimate them parametrically.

of Phase I research $\psi_j(p)$, and delivery cost c_{ij} given by $v_{ij} \equiv \tilde{v}_i \cdot \theta_j \cdot \exp(X_j\beta)$, where $\tilde{v}_i \sim \tilde{V}$; $\psi_j(p) \equiv \theta_j \cdot \exp(X_j\beta) \cdot \tilde{\psi}(p)$; and $c_{ij} \equiv \theta_j \cdot \exp(X_j\beta) \cdot \tilde{c}_i$, where $\tilde{c}_i$ has cdf $\tilde{H}(\cdot; t/(\theta_j \cdot \exp(X_j\beta)))$. The primitives to be estimated are Θ, β, $\tilde{V}$, $\tilde{\psi}(\cdot)$, $\tilde{H}(\cdot;\cdot)$, and η.

This specification induces a correlation between values, implementation costs, and costs of research: certain projects are more valuable to the DOD but also more costly to implement and research. Thus, (θ_j, X_j) controls for "vertical" heterogeneity across projects. The residual heterogeneity $\tilde{v}$ comes from heterogeneous match quality with the DOD, and it is orthogonal to research and implementation costs. Adding unobserved heterogeneity also softens the hard constraint induced by the fact that Phase III does not happen if $v \leq c$: an especially large transfer need not signify that values are high but rather that the particular contest in question had a large value of θ_j. Finally, the multiplicative specification has computational benefits: substituting into the equilibrium conditions (1) and (3) yields

$$t^*_{N_2}(v_{ij}; X_j, \theta_j) = \theta_j \cdot \exp(X_j\beta) \cdot \tilde{t}_{N_2}(\tilde{v}_i) \tag{5}$$

for some effort function $\tilde{t}_{N_2}(\cdot)$, so I can control for heterogeneity tractably via regression.

I place parametric restrictions to assist in estimation. I assume that (i) $\tilde{V}$ is lognormal with location parameter μ_{N_1} and scale parameter σ_{N_1}, (ii) $\tilde{H}(\cdot; t)$ is lognormal with mean parameter $\mu(t)$ with $\mu'(t) < 0$ (the parameterization is in Appendix C.2) and scale parameter σ_C, and (iii) $\tilde{\psi}(p) = \alpha_{N_1} p^2/2$. I place no restrictions on the distribution of θ. Allowing some parameters ($\tilde{V}$ and α) to depend on N_1 allows for a second way of controlling for unobserved heterogeneity, capturing potential endogeneity in N_1: the DOD may choose a different number of Phase I competitors for projects that have different primitives.

4.3. Estimation Procedure

Since solving the model is computationally intensive, a full-solution approach to estimation is unwieldy. However, the identification argument in Section 4.1 is constructive and lends itself to a transparent estimation procedure: the argument highlights the upper bound of Phase III transfers as a function of Phase II research efforts as an object that can be directly parameterized. I embed this intuition in a maximum likelihood procedure. I estimate (i) the dependence on X_j in a first-stage regression, (ii) the distribution of Θ nonparametrically using the residual correlation in Phase II contracts within-contest, (iii) the cost and value distribution using MLE, and (iv) the bargaining parameter by minimizing the distance between the contracts implied by the estimated parameters and the solution of the model. I restrict the sample to settings in which there is guaranteed to be no selection (i.e., I drop all contests with $(N_1, N_2) = (3, 2)$ or $(N_1, N_2) = (4, 2)$ when estimating Phase II parameters) so that searching for a monotone effort function is guaranteed to be internally consistent. Details are in Appendix C.2.

Step 1 (Partialling out Covariates). Taking logs of (5) gives $\log t^*_{N_{2j}}(v_{ij}; X_j, \theta_j) = X_j\beta + \log\theta_j + \log\tilde{t}_{N_{2j}}(\tilde{v}_i)$. Thus, a regression of the log of Phase II effort on contest-level covariates returns "normalized efforts" plus the unobserved heterogeneity, i.e., $\nu_{ij} \equiv \log\tilde{t}_{N_{2j}}(\tilde{v}_i) + \log\theta_j \equiv \log\tilde{t}_i + \log\theta_j$, along with an estimate $\hat{\beta}$ of the impact of the covariates. I then residualize the Phase III transfer by dividing by $\exp(X_j\hat{\beta})$, where X_j includes year fixed effects, SYSCOM fixed effects, and the topics from the text analysis via LDA.

Step 2 (Unobserved Heterogeneity). I use a deconvolution argument from auctions (Li and Vuong (1998), Krasnokutskaya (2011)) to estimate the distributions of Θ and the normalized efforts $\tilde{t}$ for each (N_1, N_2) combination. In particular, consider pairs (ν_{i_1j}, ν_{i_2j}) from the same contest j. Since $\nu_{ij} = \log\tilde{t}_i + \log\theta_j$, with $\tilde{t}_{i_1}$, $\tilde{t}_{i_2}$, and θ_j mutually independent (since we are restricting to contests without selection) and the distribution of $\log\theta_j$ normalized to mean zero, the distributions of θ_j and $\tilde{t}_i$ are identified from the joint distribution of (ν_{i_1j}, ν_{i_2j}) (Kotlarski, 1967). I follow Krasnokutskaya (2011) for estimation.

Step 3 (Phase II Parameters). I maximize the likelihood of observing the Phase II and III data, integrating out over the distribution of unobserved heterogeneity estimated in Step 2. In particular, I maximize over the distributions $\tilde{V}$ and $\tilde{H}(\cdot;\cdot)$, fixing the bargaining parameter η. However, rather than solving the model explicitly, I leverage monotonicity to recover an implied effort function. Since efforts are one-to-one with values fixing (N_1, N_2), a firm with a value in the q^{th} quantile of a candidate distribution of $\tilde{V}_{N_1}$ will exert effort in the q^{th} quantile of the distribution of $\tilde{t}_{(N_1,N_2)}$, estimated in Step 2. Thus, for a candidate $\tilde{V}$, I can compute the inverse effort function $\tilde{v}(\cdot)$ for $\theta_j = 1$ without solving the model directly, and I can then compute the likelihood efficiently.

Step 4 (Bargaining Parameter). Information on η comes from the firm's FOC, which I impose by solving the Phase II model. I do so at each value of η on a fine grid, at the estimated parameters from Step 3. I then use a simulated method-of-moments procedure to match the failure rate and the Phase III transfers of the observed data with simulated values from each of the solved models for the various values of η.

Step 5 (Phase I Cost Function). For each $(N_1, \bar{N}_2)$, I estimate the probability $p^*_{(N_1,\bar{N}_2)}$ of a particular contestant succeeding (when there are N_1 contestants in Phase I and a limit of $\bar{N}_2$ on Phase II) via maximum likelihood of a censored binomial model. The unobserved number of successes N_{Sj} in contest j is $N_{Sj} \sim$ Binomial$(N_1, p^*_{(N_1,\bar{N}_2)})$, but the observed quantity is $N_{2j} = \min\{N_{Sj}, \bar{N}_{2j}\}$. Upon estimating $p^*_{(N_1,\bar{N}_2)}$, I compute the profits from Phase II by solving the model using the estimated parameters from Step 4 for all values of N_1 at the estimated $p^*_{(N_1,\bar{N}_2)}$. I then use the FOC for (3) as the estimating equation for α_{N_1}.

4.4. Robustness Checks

In this section, I describe the robustness checks associated with the various modeling components discussed in Section 3.2. The modularity of the estimation procedure allows to to adapt it easily for these robustness checks.

I first consider the possibility that the DOD only partially reimburses firm costs. One could be worried that the firm spends internal funds to supplement the observed Phase II R&D contract. I assume that $t^*(\cdot) = \gamma \cdot t^{DOD}(\cdot)$ for a known $\gamma > 1$. In this case, Steps 1–3 of estimation proceed identically, and I rescale $\tilde{H}(\cdot; t)$ by γ before moving to Steps 4 and 5. I take $\gamma = 1.75$, which corresponds to actual R&D costs being 75% larger than the Phase II contract. Appendix E provides some evidence that this number is relatively large.

I next consider the possibility that Phase II changes values as well: values might not be learned perfectly at the end of Phase I, or prototyping may highlight necessary modifications to the underlying technology which alter value. In this extension, I let the draw of v at the end of Phase I represent an "initial" value; I let the cost in Phase II be drawn from $H(\cdot; t)$; and the initial value be shocked to $\zeta \cdot v$ where $\zeta \sim F_\zeta(\cdot)$. I parameterize $F_\zeta(\cdot)$ as a beta distribution with parameters $(\beta_\zeta, 1)$, which has a clear economic interpretation: Phase II research may reveal that certain aspects of the Phase I proposal are commercially infeasible, causing value to drop. The beta distribution, even with the second parameter restricted to 1, allows for a variety of means and variances. Having a parsimonious parameterization is also useful since I expand Step 3 by optimizing $\tilde{V}$ and $\tilde{H}(\cdot; \cdot)$ over a grid of both η and β_ζ.[19]

I also consider the possibility of external incentives to exert effort. One may be concerned that SBIR participation produces social surplus outside the specific contest. A firm may gain experience that helps it develop higher-quality products for government or private contractors. Knowledge gained during research can improve the human capital of its scientists. Some of these benefits may be difficult to measure in any dataset, much less tie back to a specific contest. Thus, I consider a data-driven robustness check to capture sources

[19]This model addresses the concern at the end of Section 4.1 that variation in Phase III transfers conditional on Phase II transfers may not be one-to-one with variation in surplus. As discussed in that section, the optimality condition provides information about this $F_\zeta(\cdot)$.

Percentile	2.5%	10%	25%	50%	75%	90%	97.5%
θ	0.387	0.728	0.876	1.012	1.165	1.346	1.938
	(0.095)	(0.063)	(0.044)	(0.022)	(0.054)	(0.176)	(0.459)

Table 2: Quantiles of the distribution of unobserved heterogeneity Θ. The estimation sample includes all 240 contests with $N_2 \geq 2$ but without selection.

of surplus other than Phase III: I allow for success in Phase I to lead to benefits B that are internalized by the firm and are also socially valuable. To estimate B, I use the value of the Phase I contract. I keep Steps 1–4 as is and modify Step 5 to match both the Phase I contract and the FOC associated with (3), with B added to the term in square brackets.

5. Structural Estimates

This section reports parameter estimates for the baseline and the robustness checks. I bootstrap standard errors, sampling with replacement while fixing the distribution of $(N_1, \bar{N}_2)$.

5.1. Baseline Model

Table 2 shows the distribution of unobserved heterogeneity θ. A contest in the 10^{th} percentile has values and costs that are about 70% of the median contest, and one in the 90^{th} percentile has values that are about 35% larger than median. There is a somewhat large range: moving from the 2.5^{th} percentile to the 97.5^{th} percentile increases values and costs by a factor of 5.

Column (1) of Table 3 shows estimates pertaining to Phase II—values, costs, and the bargaining parameter. Panel (A) shows means of the value distribution as a function of N_1. I scale estimates of $\tilde{v}$ by the estimated mean of $\theta_j \cdot \exp(X_j \beta)$ from Steps 1 and 2. Typical projects are worth \$21–\$28 million to the DOD, and there is some variation in these values across N_1. The variance informs how much heterogeneity there exists at the end of Phase I across competitors in their potential to generate surplus. Panel (B) shows that these value distributions are rather narrow, so such heterogeneity is limited: the "95% range" (the difference between the 2.5^{th} and 97.5^{th} percentiles) is between \$2.6 and \$3.5 million.

The identification argument in Section 4.1 sheds light on the moments in the data that influence these estimates. Most of the observed Phase III transfers lie below the 95^{th} percentile of the estimated values (see Figure 1(b)), and in this sense, the values serve as an upper bound for the transfer distribution: points beyond this upper bound are explained by the heterogeneity encapsulated by X and θ. The slope of this "soft" upper bound (as a

	(1)	(2)	(3)	(4)	(5)
(A) Mean Values ($M)					
$N_1 = 1$	28.20	11.95	27.29	28.52	10.24
	(7.36)	(6.35)	(6.37)	(5.31)	(2.75)
$N_1 = 2$	22.43	13.00	23.61	22.13	11.28
	(4.45)	(3.56)	(4.13)	(2.74)	(2.27)
$N_1 = 3$	21.40	13.62	23.74	26.36	10.12
	(4.47)	(3.29)	(4.30)	(4.18)	(4.54)
$N_1 = 4$	21.99	16.51	25.90	23.22	17.09
	(4.52)	(3.28)	(5.45)	(5.61)	(3.41)
(B) 95% Range of Values ($M)					
$N_1 = 1$	3.55	1.41	3.46	3.58	1.16
	(0.91)	(0.75)	(0.79)	(0.65)	(0.41)
$N_1 = 2$	2.89	1.60	3.04	2.84	1.39
	(1.07)	(0.64)	(0.79)	(0.85)	(0.37)
$N_1 = 3$	2.64	1.61	2.91	3.28	1.19
	(0.63)	(0.51)	(0.51)	(0.54)	(0.56)
$N_1 = 4$	2.65	2.03	3.23	2.93	2.09
	(0.53)	(0.39)	(0.66)	(1.00)	(0.61)
(C) Properties of the Cost Distribution					
$\Pr(c < v)$	0.079	0.056	0.061	0.070	0.062
	(0.009)	(0.009)	(0.009)	(0.012)	(0.015)
$\mathbb{E}[c \mid c < v]$ ($M)	12.43	7.49	12.43	13.39	6.26
	(2.37)	(1.69)	(2.02)	(1.19)	(1.00)
1^{st} Percentile ($M)	4.77	3.76	5.83	5.69	3.03
	(1.04)	(0.85)	(0.91)	(0.75)	(0.52)
5^{th} Percentile ($M)	15.53	12.23	18.98	18.53	9.87
	(3.38)	(2.77)	(2.96)	(2.46)	(1.71)
10^{th} Percentile ($M)	29.14	22.94	35.62	34.76	18.51
	(6.35)	(5.20)	(5.56)	(4.66)	(3.23)
Elasticity	-0.221	-0.140	-0.100	-0.248	-0.402
	(0.087)	(0.92)	(0.070)	(0.053)	(0.048)
(D) Other Information					
η	0.51	0.58	0.51	0.25	0.75
	(0.06)	(0.05)	(0.12)		
β_ζ			11.0		
			(4.8)		

Table 3: Phase II parameter estimates for the (1) baseline, (2) model with partial reimbursement with $\gamma = 1.75$, (3) model with a shock to v in Phase II, and (4)/(5) model with fixed η. The estimation sample uses $N =$ 1,965 contests and excludes all contests with potential selection into Phase II.

function of Phase II effort) informs the variance in the value distribution: the fact that even projects with low levels of Phase II funding tend to occasionally have reasonably high Phase III contract amounts suggests that these projects have high values as well. Of course, due to the parametric assumptions and heterogeneity, the estimates of values are influenced by matching the failure rate as well, which depends on the cost estimates below.

Panel (C) shows the estimates related to the delivery cost distributions, aggregating across all data points.[20] Just as values are positively selected conditional on success, cost draws are negatively selected; I thus discuss moments of the distribution that help interpret the relevant range. The probability that costs are less than values is 0.08, slightly lower than the observed success rate.[21] The mean of these costs draws is \$12.4 million. Given the low success rate, low quantiles of the unconditional distribution are the relevant range: I find the 1st percentile is \$4.8 million and the 5th percentile \$15.5 million. Thus, there is substantial variation in costs, and consequently generated surplus, even conditional on success. The elasticity of the cost quantiles with respect to effort is moderately low: if research efforts increase by 1%, the quantiles of the delivery cost distribution decrease by 0.2%.[22]

Conditional on η, the cost distributions are estimated from two main patterns (see Table A.3). First, the failure rate decreases with research effort; the rate of this decrease (after accounting for increases in value) and the failure rate itself inform the distribution and the elasticity. Second, the observed transfers increase with the Phase II amount, which must be due to the increase in the values. A decrease in the cost would counteract this effect, so the estimated elasticity cannot be so high as to cause the observed transfers to drop.

Panel (D) reports that the estimated bargaining parameter is 0.51: the DOD gives the winning firm about half the (incremental) surplus generated from the process. As discussed in Step 4, this estimate directly uses information about the firm choosing research efforts optimally. It is determined by fitting the equilibrium transfers and failure rates. Roughly, a larger value of η would overpredict the transfers (by bringing them closer to the value of the project) and reduce the failure rate by increasing the incentives to conduct research.

Column (1) of Table 4 shows the estimate of α, averaged across all auctions in the dataset. This estimate suggests that a marginal increase in the probability of success by 1 pp costs \$3,210. This estimate comes from equating the marginal cost of Phase I effort with the marginal benefit computed from the Phase II estimates in Table 3. This amounts to an average expenditure of \$46,000 in Phase I across all contests. The estimated expenditure on

[20] I fix a value of unobserved heterogeneity θ and compute moments of the cost distribution at the implied value of $\tilde{t}_{2ij} = t_{2ij}/\theta$ for *each* contestant i in each contest j. I then average across all these data points and integrate out over θ and scale the estimates to millions of dollars.

[21] Note that $\Pr(c < v)$ is not exactly a failure rate (although it is quite close), since I compare the cost draw to a generic draw from the value distribution. I view this moment as a descriptive feature of the cost distributions.

[22] Since research effort only parameterizes the mean of the lognormal, this elasticity is uniform across quantiles.

	(1)	(2)	(3)	(4)
Average Marginal Cost α ($M)	0.321	0.231	0.290	0.457
	(0.072)	(0.078)	(0.076)	(0.076)
Average Phase I Expenditure ($M)	0.046	0.027	0.037	0.063
	(0.009)	(0.010)	(0.010)	(0.010)
External Benefits ($M)				0.051
				(0.012)

Table 4: Phase I parameter estimates for the (1) baseline, (2) model with partial reimbursement with $\gamma = 1.75$, (3) model with a shock to v in Phase II, and (4) model with external benefits of research. $N = 2{,}773$ contests.

Phase I is slightly lower than the DOD-specified amount of $80,000, but it is nevertheless in the right ballpark. Given this value was not used at any point in estimation, the similarity between these numbers provides one external check of model fit. Appendix D.3 shows other checks, including for contests held out in Phase II estimation.

These estimates inform the economic incentives that feed into the analysis of contest design. The somewhat narrow value distribution suggests that (conditional on success in Phase I), there is limited variation across competitors within contests. This is sensible within the empirical context since projects are well-specified ex-ante: there is not much room for innovation on the dimension of the quality of the proposal at the end of Phase I. Thus, firms are not strongly differentiated in their ability to generate surplus at the start of Phase II. Economically, this means that the benefit of multiple draws from Phase I (especially given success rates are high in Phase I) are limited. Delivery costs, however, are substantially different across firms. This suggests that there is considerably more heterogeneity in the surplus that can be generated by different competitors during Phase II. The map from efforts to costs is moderately flat, which affects how efforts respond to incentives generated by the contest design. Finally, the DOD lets the firms capture a fairly large portion of the surplus they generate, giving them high-powered incentives to conduct research. Section 5.2 shows that these qualitative observations hold in all robustness checks.

5.2. Estimates for Alternate Models

I first discuss estimates without imposing any optimality condition, fixing η and stopping the estimation procedure after Step 3. Column (4) of Table 3 shows estimates assuming firms capture an especially low share of the surplus ($\eta = 0.25$). Here, estimated values are only slightly larger than baseline, consistent with the idea from Figure 2 that the upper bound of Phase III transfers informs these values. The estimated cost distribution is significantly larger than in the baseline: with $\eta = 0.25$, Phase III contracts are closer to costs than in the baseline, so costs must increase to rationalize the observed Phase III contracts. With $\eta = 0.75$ (Column (5)), values are somewhat smaller: rationalizing especially low Phase III contracts becomes difficult if η is high, so the values are closer to the Phase III contracts. The cost distribution is smaller than in the baseline—both mechanically due to the smaller values and due to the aforementioned effect of η. These estimates provide intuition connecting the identification argument to the parameter estimates to clarify the mechanics of the procedure. More importantly, they highlight that the salient features of the estimates—limited heterogeneity in potential surplus after Phase I but considerably more heterogeneity after Phase II—are robust to even relaxing the equilibrium assumption.

Column (2) of Table 3 shows estimates for a model where the DOD only partially reimburses research efforts: the firm-optimal effort is 75% larger than the Phase II contract. The economic implication of this assumption is that the benefit to effort is lower than in the baseline (as all efforts are scaled up). The lower estimate of elasticity of costs with respect to effort reflects this effect. Moreover, this also means that to justify the observed effort, the model estimates that the firm expects to capture a larger share of the surplus (58%) than in the baseline. Partial reimbursement of costs implies that expected profits from Phase II are lower, which the model justifies through lower estimates of the marginal cost α of Phase I effort (Column (2) of Table 4). This estimate implies that the average Phase I expenditure is $27,000; while this is lower than the standard contract amount, the assumption that the DOD is reimbursing only about half the costs is stark.

Column (3) of Table 3 considers the model with shocks to values in Phase II. The estimates suggest a limited role for this shock. Coincidentally, the estimated share of the surplus is also 0.51, and the shock ζ has mean 0.92 (s.e. 0.13) and standard deviation 0.077 (s.e. 0.023). Accordingly, the estimated mean values of v are slightly larger since the true value at the end of Phase II is scaled down by ζ. The estimate for the marginal cost of Phase I effort (Column (3) of Table 4) is also slightly lower: in the baseline, the entire departure from the "upper bound" in the data is attributable to increases in surplus, but this variation is partly muted by ζ. However, the relative heterogeneity in values and costs remains robust.

Finally, Column (4) of Table 4 considers the possibility of an external benefit that firms earn from Phase II. Since this benefit only affects Phase I effort, I use the same Phase II estimates as in the baseline. As mentioned at in Section 4.4, this extension is the only one in which I use the dollar value of the Phase I contract: the baseline Phase I estimates underestimate the observed Phase I contract amount, and adding an extra benefit from Phase II can explain the departure since it increases the marginal benefit and thus the estimated marginal cost. With this new benefit, the average Phase I expenditure is estimated to be \$63,000—significantly closer to the observed value but admittedly still somewhat smaller. The final row of the table shows that the average benefit from entering Phase II not attributable directly to the SBIR contract is \$51,000. This number is sizable, but it still suggests that Phase III is the main source of surplus in the contest: counterfactuals below show that total external benefits are around half of the surplus from procurement.

6. The Efficiency of R&D Procurement Contests

This section studies the efficiency of the design of R&D procurement contests in this real-world setting. In Section 6.1, I estimate how much surplus can be improved by moving to the optimal design and quantify the sources of inefficiency. Section 6.2 investigates how much of these improvements can be captured by (plausibly) easily implementable design changes. I find substantial gains from changing contest design and that simple changes can capture a large share of these gains. This motivates the analysis in Section 6.3, where I explore explanations for the departure between the observed and the socially optimal design.

6.1. Quantifying the Inefficiency of the Current Design

Consider a planner who can choose the optimal design from a large, albeit still restricted, space of mechanisms. I allow the planner to adjust the incentives of the contest: the share η of the surplus offered to the firms in Phase III and potentially a prize or fee for entering Phase II (which one can conceptualize as modifying $t^{DOD}(\cdot)$). In some cases, I also allow the planner to change the number of contestants. Firms then choose their efforts in response to these incentives. How much can such a planner increase social surplus?

To understand the answer, I quantify the sources of inefficiency outlined in Section 3.3: holdup, reimbursement, and business stealing. Since holdup manifests in lower Phase II effort, I quantify it by scaling all Phase II effort functions by $\lambda > 1$ without changing Phase I effort. That is, I compute the surplus of a contest with the equilibrium effort p^* in the first stage and effort $\lambda \cdot t^*_{N_2}(\cdot)$ in the second stage. I find the value λ^{Opt} that maximizes social surplus. Table 5 shows $\lambda^{\text{Opt}} \approx 2$: doubling Phase II effort increases surplus by 40–50% for parameters with $N_1 \geq 2$ and about 20% for the parameters with $N_1 = 1$. Holdup, governed primarily by the fact that the firm captures only half the surplus generated, is a significant inefficiency. Stepping aside briefly from the specific setting, this result highlights the importance of evaluating holdup in other settings (Joskow, 1987; Crocker and Masten,

N_1	$\bar{N}_2$	Baseline		Adjusting Phase II		Adjusting Phase I	
		p^*	Δ Surplus ($M)	λ^{Opt}	Δ Surplus ($M)	p^{Opt}	Δ Surplus ($M)
1	1	0.62	0.250	1.89	0.047 / 18.6%	0.90	0.028 / 10.9%
		(0.00)	(0.067)	(0.47)	(0.027)	(0.03)	(0.018)
2	1	0.54	0.104	2.22	0.044 / 42.9%	0.49	0.001 / 1.0%
		(0.01)	(0.041)	(0.48)	(0.027)	(0.04)	(0.001)
3	2	0.44	0.117	2.35	0.063 / 54.0%	0.47	0.001 / 0.5%
		(0.00)	(0.085)	(0.50)	(0.036)	(0.06)	(0.011)
4	2	0.42	0.189	2.19	0.077 / 41.0%	0.39	0.001 / 0.5%
		(0.00)	(0.154)	(0.43)	(0.043)	(0.04)	(0.004)

Table 5: Social surplus and first-stage effort for various values of $(N_1, \bar{N}_2)$. λ^{Opt} is factor by which Phase II efforts must be multiplied, keeping Phase I efforts fixed, to maximize surplus. p^{Opt} is the optimal Phase I effort, fixing Phase II efforts.

N_1	$\bar{N}_2$	Baseline	Δ Social Surplus			
			Fixing $(N_1, \bar{N}_2)$	$\bar{N}_2 \to N_1$	$N_1 \to 4$	$(N_1, \bar{N}_2) \to (4, 4)$
1	1	0.250	0.097 / 38.7%	0.097 / 38.7%	0.172 / 68.6%	0.862 / 344.4%
		(0.067)	(0.049)	(0.049)	(0.061)	(0.323)
2	1	0.104	0.046 / 43.9%	0.176 / 169.8%	0.070 / 67.5%	0.394 / 380.2%
		(0.041)	(0.031)	(0.053)	(0.031)	(0.136)
3	2	0.117	0.071 / 61.0%	0.109 / 93.0%	0.093 / 80.0%	0.170 / 145.9%
		(0.085)	(0.052)	(0.065)	(0.050)	(0.085)
4	2	0.189	0.081 / 42.9%	0.185 / 98.0%	0.081 / 42.9%	0.185 / 98.0%
		(0.154)	(0.062)	(0.111)	(0.062)	(0.111)

Table 6: Baseline surplus ($M) and gains from socially efficient designs

1988). Estimates of the inefficiency it generates remain limited (Zahur, 2020).

If holdup were the only inefficiency, Phase I effort would also be underprovided. Reimbursement and business-stealing counteract this underprovision. I quantify the net inefficiency in Phase I by computing the gains from the planner choosing Phase I effort (p^{Opt}) while keeping Phase II effort fixed at $t^*_{N_2}(\cdot)$. Table 5 shows that Phase I effort is generally close to optimal, since $p^* \approx p^{\text{Opt}}$. We see slight overprovision for $N_1 = 2$ or 4 ($p^{\text{Opt}} < p^*$) and slight underprovision for $N_1 = 3$; the gains from adjusting effort are less than 1%. Overall, the reimbursement and business stealing effects are also large and counteract the magnitude of holdup.[23] Appendix D.1 separately quantifies these two effects.

Putting these together, I estimate the gains from choosing the optimal design by adjusting incentives in both phases together. This involves setting $\eta = 1$ to alleviate holdup in Phase

[23]One exception to this conclusion are contests with $N_1 = 1$, where Phase I effort is significantly underprovided. This conclusion is partly driven by different parameters for Phase I contests but is also based on a more fundamental economic difference for these contests: since there is only one competitor, there is no business stealing in these contests, and the reimbursement effect is not strong enough by itself to balance out holdup.

II and setting fees for entry into Phase II to alleviate the net effect of reimbursement and business stealing in Phase I. Table 6 shows that fixing the number of competitors, the optimal design increases surplus substantially, by about 40–60%.

The number of competitors could also be a design variable. Introducing another competitor into either phase has the direct benefit of another draw—of costs in Phase II and values in Phase I (together with success). The tradeoff is the additional cost of R&D. In the last three columns of Table 6, I allow the planner to increase the number of competitors in each phase to up to 4 and then incentivize effort optimally. First, I find that removing the restriction on Phase II competitors (setting $\bar{N}_2$ to N_1) increases surplus considerably. For intuition, note that an additional competitor is socially harmful if it is an "ex-ante" substitute, e.g., if it is common for both competitors to succeed simultaneously and produce similar products. Given cost draws are variable and success rates are low in Phase II, competitors are effectively not substitutes, and the planner prefers to increase competition in Phase II. The planner also gains from increasing competition in Phase I even without relaxing the limit on Phase II competition: the main benefit of this change is to increase the number of competitors in Phase II, as the variation in quality conditional on success at the end of Phase I is limited. The final column shows that under all parameter values, the planner prefers to increase N_1 to 4 and place no restriction on entry into Phase II, assuming that it can control incentives after choosing the number of competitors.

Overall, I find the surplus of the current design can be improved substantially without even changing the set of participants. The planner also prefers competition: allowing for introducing more participants yields significantly larger increases in the surplus.

6.2. The Effectiveness of Simple Changes to Contest Design

The gains from the optimal design are substantial, but implementing it can be difficult. The DOD may have trouble committing to giving firms the entire surplus in procurement. The optimal design involves a fee to enter Phase II (at $\eta = 1$, Phase I effort is overprovided), engendering ex-post losses that may harm participation. Here, I consider a set of "simple" design changes motivated both by theoretical considerations and what may be feasible in this setting. These changes may still come with implementation issues, which in I discuss in Section 6.3. Nevertheless, the takeaway is that simple changes can capture a large fraction of the surplus gains: implementation does not explain the surplus left on the table in this setting. More broadly, procurers should not overlook simple tools at their disposal.

Table 7 shows the results of all such changes. I consider various combinations of

$(N_1, \bar{N}_2)$	(1, 1)	(2, 1)	(3, 2)	(4, 2)
Baseline Surplus ($M)	0.250	0.104	0.117	0.189
	(0.067)	(0.041)	(0.085)	(0.154)
Strength of Incentives				
η^*	0.95	0.59	0.63	0.62
	(0.08)	(0.11)	(0.10)	(0.10)
Δ Surplus from η^*	0.096	0.010	0.026	0.021
	(0.056)	(0.017)	(0.040)	(0.032)
Number of Competitors				
Δ Surplus from $\bar{N}_2^* = N_1$	—	0.070	0.013	0.048
		(0.050)	(0.012)	(0.052)
Δ Surplus from $N_1^* = 4$	0.105	0.018	0.018	—
	(0.029)	(0.004)	(0.013)	
Δ Surplus from $(N_1^*, \bar{N}_2^*)$	0.603	0.212	0.050	0.048
	(0.195)	(0.141)	(0.043)	(0.052)
Contest Structure				
Δ Surplus from Combining Stages	—	0.064	0.012	0.043
		(0.057)	(0.019)	(0.055)
Δ Surplus from Mandatory IP Sharing	—	—	0.070	0.088
			(0.027)	(0.057)
Δ Surplus from IC IP Sharing	—	—	-0.092	-0.133
			(0.041)	(0.079)

Table 7: Changes in social surplus from simple design changes

$(N_1, \bar{N}_2)$, with the parameters associated with the corresponding value of N_1. The Baseline in the first row is the surplus associated with $(N_1, \bar{N}_2)$ in the observed design.

Strength of Incentives. The DOD could provide higher-powered incentives to the firms by promising a larger share η of the surplus. It may require some time for firms to believe a promise of a higher Phase III profit margin, but this design change is possible even if finding new competitors is difficult. Increasing η ameliorates holdup by giving firms a greater claim to the surplus, and doing so is unambiguously beneficial for surplus in Phase II. However, larger η increases both business-stealing and reimbursement, assuming the DOD continues to reimburse all Phase II R&D costs. The optimal η is a balance between these effects.

The first panel of Table 7 shows that promising firms a larger share of the surplus improves overall social surplus by about 10–20%. Moreover, while the change in the incentive is modest (the optimal η is around 0.6, not much larger than 0.51), the effect on surplus is substantial: about one-third to one-half of the gains from setting the optimal design without changing the number of competitors.[24]

[24]For the case with $N_1 = 1$, I find η^* to be much larger: as there is no business stealing, the cost of increasing η is only to exacerbate the reimbursement effect.

Number of Competitors. A theme in contest theory is that limiting contestants can be socially and privately beneficial since excess competition alters the incentives to conduct research (Taylor, 1995; Fullerton and McAfee, 1999; Che and Gale, 2003). With stochastic outcomes, the main tradeoffs were outlined in Section 6.1: another contestant allows for an additional chance at successful research but induces duplicative R&D costs. In Section 6.1, I allowed the planner to adjust incentives to exert effort as the number of contestants changed. Now I consider whether the planner would want to change the number of contestants in each phase once it takes into account the incentive effect it would induce on the research effort they exert. Here, I show the total effects of changing competition; Appendix D.1 provides a decomposition and shows that the incentive effect is small.

The second panel of Table 7 considers three ways to increase the level of competition. The first alters the number of competitors who can enter Phase II by optimizing $\bar{N}_2$ while fixing N_1; in all cases, the optimal $\bar{N}_2 = N_1$. This adjusts effort down in Phase I (which could be beneficial since effort is sometimes overprovided) but allows for more chances for success in Phase II. As mentioned in Section 6.1, firms in Phase II are not substitutable ex-ante, and allowing more to enter increases surplus. Of course, once in Phase II the firms still exert less than the efficient level of effort, given holdup. On net, the surplus gains from simply changing the number of competitors is sizable: changing $\bar{N}_2$ by itself lets the planner capture up to one-third of the gain from the optimal design (changing $\bar{N}_2$ and incentives).

The second counterfactual fixes $\bar{N}_2$ but optimizes over the number of Phase I contestants N_1; the optimum is always $N_1 = 4$. This counterfactual requires the DOD to be able to find equally competent firms to expand competition; I discuss sensitivity to this assumption in Section 6.3. Without changes in the number of firms allowed to compete in Phase II, the benefit of encouraging competition in Phase I comes from (i) a greater number of viable competitors in Phase II and (ii) more chances at especially high value draws in Phase I. Since heterogeneity after Phase I is low, the former channel is more important. On net, gains are also substantial: simply changing Phase I competition can capture one-quarter to one-half of the gains from changing Phase I competition and incentives (from Table 6).

If the planner can change both N_1 and $\bar{N}_2$, it selects $(4, 4)$, and increases in surplus are again substantial. In sum, these counterfactuals suggest that even if the only lever available to the DOD were to change the number of contestants, it should increase competition beyond levels we see. This result poses a counterpoint to wisdom from contest theory: here, the uncertainty in outcomes is sufficiently large that it overpowers the possibly adverse effect competition has on incentives. In real-world settings, entry restrictions may be too severe.

Contest Structure. I finally consider contest structure itself. Appendix D.2 provides details.

Should the DOD combine the two phases? Suppose any firm can conduct "Phase II" research to draw a cost after finishing "Phase I" research to draw a value. The DOD still reimburses the firm-optimal effort. Holdup still exists, and the reimbursement effect in "Phase I" still (partially) balances this effect, but there is no business-stealing effect since spots in "Phase II" are not limited. Differently from the baseline, since phases are combined and firms do not see realizations of competitors' intermediate outcomes, firms choose "Phase II" effort without knowing the number of competitors also conducting such research.

Table 7 shows that the surplus from combining phases is similar to that from increasing $\bar{N}_2$ to N_1. By combining the two phases, the DOD loses the ability to screen out lower-quality outcomes from Phase I. Moreover, if R&D were such that multiple similar successes in Phase II were likely (projects were ex-ante substitutes), then the DOD could avoid duplicative research by restricting entry. However, in this setting, firms who do not develop successful initial-stage research would have no scope of generating surplus in later stages and thus exert no later-stage effort even when phases are combined. Effectively, the only difference between combining phases and setting $\bar{N}_2 = N_1$ amounts to small incentive changes due to disclosing the exact number of competitors.

A second feature is that firms do not share intellectual property. In the current setup, certain firms work during Phase II to develop ideas with strictly worse value than their opponents'. There is a direct social benefit for firms with higher draws of v to share these plans with competitors; moreover, the DOD could share successful plans with firms who would otherwise not have been invited to Phase II, leading to another direct benefit. However, there are countervailing incentive effects: while an otherwise weaker firm given access to a higher-quality idea may have more incentive to exert effort, the firm with the high-quality idea would shade its effort below its level in the equilibrium without information sharing. The net effect on surplus is ambiguous. Moreover, the net effect on firms' profits in Phase II is also ambiguous, and this in turn affects equilibrium Phase I research efforts. Columns (3) and (4) show the effects of mandatory IP sharing for cases where $\bar{N}_2 > 1$, which is necessary to implement it. The gains are large (46–60%), suggesting that the countervailing incentive effects of IP sharing do not outweigh the direct benefits.

Unlike other changes, IP sharing is less "simple": it may be difficult to mandate that firms share their IP. I consider a modification: the DOD gives the firms the option of sharing their IP in return for a prize $K(v)$, and I compute a schedule $K(v)$ that would make sharing (barely) incentive-compatible for all v. While this prize overcomes the mandate, it increases incentives in Phase I, leading to overprovision of effort. On net, this excess effort leads to decreases in surplus from the incentive compatible IP sharing design. This may be one reason the DOD does not incorporate IP sharing into its design.

6.3. Why Does the DOD Avoid the Optimal Design?

The results so far suggest substantial surplus gains from even seemingly simple changes to contest design. Given the surplus left on the table, why does the DOD not adjust the design?

Unmodeled practical concerns do not obviously explain the departure. There may be institutional reasons the DOD is unable to provide large profit margins to firms in delivery, but the optimal η is generally close to the estimated one. Finding new competitors may be difficult; while I explore this below, DOD already restricts entry into Phase I, so it could relax this constraint. Furthermore, increasing the limit on Phase II competition (or combining phases) does not require finding new competitors. Sharing IP is also within the DOD's experience: in larger procurement settings, the DOD requests Technical Data Packages from contractors in case future acquisition contracts are awarded to other firms.

Here, I ask whether concerns other than implementation can explain the departure. I consider forms of model misspecification analyzed in Section 5: incomplete reimbursement of research efforts and changes to value in Phase II. I consider the possibility of non-SBIR sources of surplus. I formalize the concern that competitors may be scarce by modeling selection into Phase I. Finally, I consider alternate objectives for the DOD. Among these explanations, I conclude that the latter is most plausible: the DOD's design may indicate that it is balancing social surplus and its own private welfare.

Incomplete Reimbursement. I first consider partial reimbursement of Phase II research, assuming true efforts are 75% more than the Phase II contract. This changes parameter estimates, as Section 5.2 shows, although the main patterns are similar. A larger η means holdup is less severe in Phase II. However, this does not lead to significant overprovision of Phase I effort: the reimbursement effect is weaker since R&D costs are not fully reimbursed.

The economics—the magnitude of heterogeneity in values and costs, and the relative magnitudes of the inefficiencies—do not change much. Accordingly, the patterns in the counterfactuals do not change either. Column (2) of Table 8 shows that the socially optimal η is slightly larger than the estimated one of 0.58, although the gain is smaller.[25] The planner still prefers to invite more contestants into both phases. Finally, combining stages or IP

[25]It is also larger than the optimal one in the baseline: increasing η exacerbates the business stealing and the reimbursement effects, and the reimbursement effect was already lower here relative to baseline. This difference, together with changes in parameter estimates, explains the higher optimal incentive.

	Social Surplus						DOD Welfare	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Baseline ($M)	0.189	0.409	0.353	0.277	0.220	0.226	0.237	-0.053
	(0.154)	(0.102)	(0.129)	(0.154)	(0.149)	(0.177)	(0.083)	(0.065)
Incentives								
η^*	0.62	0.64	0.61	0.60	0.64	0.65	0.61	0.34
	(0.10)	(0.15)	(0.06)	(0.07)	(0.06)	(0.05)	(0.09)	(0.06)
Δ Outcome	0.021	0.003	0.018	0.019	0.026	0.022	0.025	0.081
	(0.032)	(0.221)	(0.037)	(0.022)	(0.026)	(0.026)	(0.026)	(0.066)
# of Competitors								
$\bar{N}_2^*$ Only	4	4	4	4	4	4	4	1
Δ Outcome	0.048	0.130	0.091	0.086	0.037	0.076	0.061	0.021
	(0.052)	(0.042)	(0.041)	(0.063)	(0.045)	(0.080)	(0.030)	(0.044)
N_1^* Only	4	4	4	4	4	4	4	2
Δ Outcome	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.016
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.028)
$(N_1^*, \bar{N}_2^*)$	(4, 4)	(4, 4)	(4, 4)	(4, 4)	(4, 4)	(4, 4)	(4, 4)	(2, 1)
Δ Outcome	0.048	0.130	0.091	0.086	0.037	0.076	0.061	0.027
	(0.052)	(0.042)	(0.041)	(0.063)	(0.045)	(0.080)	(0.030)	(0.051)
Contest Structure (Δ Outcome)								
Combining Stages	0.043	0.125	0.090	0.085	0.030	0.073	-0.005	-0.009
	(0.055)	(0.043)	(0.040)	(0.064)	(0.047)	(0.081)	(0.030)	(0.030)
IP Sharing	0.088	0.115	0.123	0.053	0.094	0.030	0.064	0.035
	(0.057)	(0.053)	(0.059)	(0.065)	(0.053)	(0.021)	(0.027)	(0.027)
IC-IP Sharing	-0.133	-0.121	-0.143	-0.142	-0.221	-0.098	0.121	-0.724
	(0.079)	(0.067)	(0.102)	(0.068)	(0.090)	(0.061)	(0.051)	(0.407)

Table 8: Effects of simple design changes under various modeling assumptions: (1) baseline; (2) partial reimbursement of Phase II R&D costs, with $\gamma = 1.75$; (3) allowing for a shock to values in Phase II; (4) external benefits to Phase II without changes in Phase I costs; (5) external benefits to Phase II with changes in Phase II costs; (6) selection into Phase I; (7) baseline parameters with Phase III DOD welfare; (8) baseline estimates with overall DOD welfare.

sharing would increase surplus, although incentivizing IP sharing would not. Given that incomplete reimbursement by 75% is already large, we cannot explain the DOD's reluctance to adopt these changes through incomplete reimbursement of firms' efforts.

Shocks to Values in Phase II. Column (3) shows counterfactuals if values change in Phase II as well. Section 5.2 shows that this shock plays a limited role. As such, counterfactual designs are similar to the baseline. The optimal level of incentive is 0.61, close to the baseline. Surplus is maximized when more firms are invited to either phase, the phases are combined, or IP is mandatorily shared. Overall, the departure from the optimal design cannot be explained by an alternate specification of how values evolve through the process.

External Incentives. If success in Phase I generates surplus beyond the specific contest, the DOD would have a direct reason to invite participants into both phases. Column (4) uses the estimate of \$51,000 of external benefits (Column (4) of Table 4) while still using the other parameter estimates from the baseline. The optimal η is slightly lower than in the baseline. The DOD prefers to invite more contestants, as expected, and the gain from doing so is larger: expanding Phase II to 4 participants increases surplus by 31% rather than 25% in the baseline. Combining phases has similar effects, as expected.

However, as shown in Section 5.2, parameter estimates change with the possibility of an external benefit: if firms anticipate these benefits, the marginal cost of effort in Phase I must be higher to rationalize the effort we observe. Column (5) of Table 8 uses estimates from Column (4) of Table 4. Taking this higher cost into account, the effects of simple contest designs are still qualitatively similar to the baseline. Overall, the possibility of internalizing additional benefits does not explain the DOD's reluctance to implement these changes.[26]

Heterogeneous Participants. The DOD may be restricting competition due to difficulty finding competent competitors. To gauge this concern, I model selection into Phase I: if the DOD increases N_1, they pick competitors with higher research costs. I let firms be heterogeneous in Phase I: each firm i has $\alpha_i \sim F_{\alpha,N_1}$, with F_{α,N_1} stochastically increasing in N_1. A firm with costs α_i selects Phase I effort $p(\alpha_i)$, taking into account the equilibrium distribution of its competitors' efforts. Appendix C.1 outlines the solution method.

I calibrate the parameterization of F_{α,N_1}, modeling it as the mixture of the lowest N_1 of $\bar{N}$ draws from a distribution F_α. Here, I take F_α to be $\bar{\alpha}$ times a beta distribution with parameters $(1.5, 1.5)$, set $\bar{N} = 8$, and choose $\bar{\alpha}$ so that $\mathbb{E}[F_{\alpha,4}]$ equals the estimate of α_4. With this parameterization, $\mathbb{E}[F_{\alpha,1}]$ is 50% of $\mathbb{E}[F_{\alpha,4}]$. Thus, the DOD restricting to just one competitor would double the cost effectiveness of a typical Phase I participant, which is sizable. (Results are similar in a large neighborhood of these parameters.)

This parameterization reduces the losses to the DOD from reducing N_1. One can compute that moving from $N_1 = 4$ to $N_1 = 2$ reduces social surplus by only 5.6%, relative to 33.1% in the baseline. This is expected: reducing N_1 now comes with the added benefit of selecting lower-cost competitors. However, Column (4) of Table 8 shows that despite this benefit, heterogeneity of this magnitude cannot explain why the DOD restricts competition. The socially optimal level is still $N_1 = \bar{N}_2 = 4$. The effects of other counterfactuals are similar, so selection into Phase I is unlikely to explain the departure from the optimal design.

[26] I have run robustness checks where I scale B to up to twice the estimated value and adjust the Phase I cost parameter accordingly. The conclusion is similar, and results are available upon request.

DOD Welfare. The DOD claims it wants to support small businesses engaging in defense innovation, suggesting it partially internalizes their welfare. Many agencies outside the DOD run SBIR purely through grants, with no mechanism to directly capture any surplus. Thus, I have so far assumed the DOD wishes to maximize surplus. However, given the DOD has interwoven SBIR with acquisition, it may be closer to a traditional procurer, maximizing the value it extracts while respecting participation constraints. Would the socially optimal design changes still be desirable if the DOD were maximizing a different function?

The first objective I consider is the surplus the DOD earns specifically in procurement: value less the procurement contract ("Phase III welfare"). Since the DOD is mandated to spend a portion of its budget on Phases I and II, it may view R&D contracts as unavoidable, leading to this objectve. Column (7) shows that with this objective, the effects of simple changes would resemble the baseline, so this alternative does not justify the departure.

A second objective is the DOD's overall welfare: the value, less the Phase III contract, less all R&D contracts. Column (8) reports results under this objective, and a starkly different picture emerges. Since the DOD captures half the surplus and pays out all research costs, it loses \$53,000 per contest. Consequently, the observed level of incentives lies between the DOD-optimal and the socially-optimal values: the DOD prefers to give firms 34% of the incremental surplus. Doing so yields lower surplus in procurement but expends less research costs (which the DOD internalizes fully), letting it earn \$29,000 per contest. Similarly, it prefers to restrict entry into Phase II even further—to just one contestant—since the benefit of a second contestant is not worth the additional research costs. It prefers to restrict to just two contestants in Phase I (if it must let two into Phase II). Its optimal choice of the number of competitors restricts entry further, to $(N_1, \bar{N}_2) = (2, 1)$. Finally, the DOD increases its losses by combining stages. While IP sharing is beneficial as long as it can be mandated, paying prizes is costly: the prizes come directly out of the DOD's budget and are not neutral to the objective. Thus, none of these simple changes to the contest structure would be advisable for a DOD that maximizes its own welfare, and the observed design generally lies between the DOD-welfare-maximizing and the surplus-maximizing ones.

7. Conclusion

Theoretical work has established that all aspects of the optimal design of an R&D contest—the strength of incentives, the level of competition, and the structure of the contest itself—depend on the primitives of the setting. As such, it is a priori ambiguous whether the designs we observe in real-world settings are close to efficient. This paper develops a model of R&D

contests to evaluate the observed design of the DOD SBIR program, a multistage R&D contest incentivized by a procurement contract. I conclude that the surplus generated is significantly less than the social optimum. The model suggests that that arguably simple design changes—providing moderately stronger incentives, increasing competition, and making changes to contest structure—capture a large share of these gains. These conclusions hold even under a variety of modifications to the underlying assumptions.

Why has the DOD not adjusted the design? The model provides one explanation: these changes are often starkly at odds with the DOD's private welfare objective, which would cause it to want to provide lower-powered incentives to firms or reduce competition. Together, these results underscore the benefits of continuing to extend the "Laffont program" to contests—using the structure of contest theory to estimate primitives and suggest improvements to observed designs—but also highlight the importance of evaluating a variety of objectives when asking whether the design of real-world contests ought to be changed.

References

AGHION, P., N. BLOOM, R. BLUNDELL, R. GRIFFITH, AND P. HOWITT (2005): "Competition and Innovation: An Inverted-U Relationship," *Quarterly Journal of Economics*, 120, 701–728.

ANTON, J. J. AND D. A. YAO (1989): "Split Awards, Procurement, and Innovation," *RAND Journal of Economics*, 20, 538–552.

——— (1992): "Coordination in Split Award Auctions," *Quarterly Journal of Economics*, 107, 681–707.

ARROW, K. (1962): "Economics Welfare and the Allocation of Resources for Invention," in *The Rate and Direction of Inventive Activity*, ed. by R. Nelson, Princeton: Princeton University Press, 609–625.

BERRY, S. T. AND J. WALDFOGEL (1999): "Free Entry and Social Inefficiency in Radio Broadcasting," *RAND Journal of Economics*, 30, 397–420.

BHATTACHARYA, S., J. GLAZER, AND D. E. M. SAPPINGTON (1990): "Sharing Productive Knowledge in Internally Financed R&D Contests," *Journal of Industrial Economics*, 39, 187–208.

BLEI, D. M., A. Y. NG, AND M. I. JORDAN (2003): "Latent Dirichlet Allocation," *Journal of Machine Learning Research*, 3, 993–1022.

BLUNDELL, R., R. GRIFFITH, AND J. VAN REENEN (1999): "Market Share, Market Value, and Innovation in a Panel of British Manufacturing Firms," *Review of Economic Studies*, 66, 529–554.

BOLTON, P. AND M. D. WHINSTON (1993): "Incomplete Contracts, Vertical Integration, and Supply Assurance," *Review of Economic Studies*, 60, 121–148.

BOUDREAU, K. J., K. R. LAKHANI, AND M. E. MENIETTI (2016): "Performance Responses to Competition Across Skill-Levels in Rank Order Tournaments: Field Evidence and Implications for Tournament Design," *RAND Journal of Economics*, 47, 140–165.

BRESNAHAN, T. AND J. LEVIN (2012): "Vertical Integration and Market Structure," in *The Handbook of Organizational Economics*, ed. by R. Gibbons and J. Roberts, Princeton, NJ: Princeton University Press, chap. 21, 853–890.

BUBB, R., S. KAUR, AND S. MULLAINATHAN (2018): "The Limits of Neighborly Exchange," Working paper, UC Berkeley.

CABRAL, L., G. COZZI, V. DENICOLÓ, G. SPAGNOLO, AND M. ZANZA (2006): "Procuring Innovations," in *Handbook of Procurement*, ed. by N. Dimitri, G. Piga, and G. Spagnolo, Cambridge: Cambridge University Press, chap. 19, 483–528.

CARRIL, R. AND M. DUGGAN (2020): "The Impact of Industry Consolidation on Government Procurement: Evidence from Department of Defense Contracting," *Journal of Public Economics*, 184, 104141.

CHE, Y.-K. AND I. GALE (2003): "Optimal Design of Research Contests," *American Economic Review*, 93, 646–671.

CHE, Y.-K., E. IOSSA, AND P. REY (2017): "Prizes versus Contracts as Incentives for Innovation," Working paper.

COHEN, C., T. R. KAPLAN, AND A. SELA (2008): "Optimal Rewards in Contests," *RAND Journal of Economics*, 39, 434–451.

CRAWFORD, G. S. AND A. YURUKOGLU (2012): "The Welfare Effects of Bundling in Multichannel Television Markets," *American Economic Review*, 102, 643–685.

CROCKER, K. J. AND S. E. MASTEN (1988): "Mitigating Contractual Hazards: Unilateral Options and Contract Length," *RAND Journal of Economics*, 19, 327–343.

DREZNER, J. A. AND M. HUANG (2009): *On Prototyping: Lessons from RAND Research*, RAND Corporation.

FULLERTON, R. L. AND R. P. MCAFEE (1999): "Auctioning Entry into Tournaments," *Journal of Political Economy*, 107, 573–605.

GILBERT, R. J. AND D. M. G. NEWBERY (1982): "Preemptive Patenting and the Persistence of Monopoly," *American Economic Review*, 72, 514–526.

GRENNAN, M. (2013): "Price Discrimination and Bargaining: Empirical Evidence from Medical Devices," *American Economic Review*, 103, 145–177.

GROSS, D. P. (2020): "Creativity under Fire: The Effects of Competition on Creative Production," *Review of Economics and Statistics*, 102, 583–599.

HENDRICKS, K. AND R. H. PORTER (2015): "Empirical Analysis and Auction Design," *Econometric Society Presidential Address*.

HOWELL, S. T. (2017): "Financing Innovation: Evidence from R&D Grants," *American Economic Review*, 107, 1136–1164.

IYER, R. AND A. SCHOAR (2015): "Ex Post (In) Efficient Negotiation and Breakdown of Trade," *American Economic Review Papers and Proceedings*, 105, 291–94.

JOSKOW, P. L. (1987): "Contract Duration and Relationship-Specific Investments: Empirical Evidence from Coal Markets," *American Economic Review*, 77, 168–185.

KIREYEV, P. (2020): "Markets for Ideas: Prize Structure, Entry Limits, and the Design of Ideation Contests," *RAND Journal of Economics*, 51, 563–588.

KOH, Y. (2017): "Incentive and Sampling Effects in Procurement Auctions with Endogenous Number of Bidders," *International Journal of Industrial Organization*, 52, 393–426.

KOTLARSKI, I. (1967): "On Characterizing the Gamma and the Normal Distribution," *Pacific Journal of Mathematics*, 20, 69–76.

KRASNOKUTSKAYA, E. (2011): "Identification and Estimation of Auction Models with Unobserved Heterogeneity," *Review of Economic Studies*, 78, 293–327.

LAFONTAINE, F. AND M. E. SLADE (2012): "Inter-Firm Contracts: Evidence," in *The Handbook of Organizational Economics*, ed. by R. Gibbons and J. Roberts, Princeton, NJ: Princeton University Press, chap. 24, 958–1013.

LEMUS, J. AND G. MARSHALL (2020): "Dynamic Tournament Design: An Application to Prediction Contests," *Journal of Political Economy*, Forthcoming.

LERNER, J. (2000): "The Government as Venture Capitalist: The Long-Run Impact of the SBIR Program," *Journal of Private Equity*, 3, 55–78.

LI, T. AND Q. VUONG (1998): "Nonparametric Estimation of the Measurement Error Model using Multiple Indicators," *Journal of Multivariate Analysis*, 65, 139–165.

LICHTENBERG, F. R. (1995): "Economics of Defense R&D," in *Handbook of Defense Economics*, ed. by T. Sandler and K. Harley, Elsevier, vol. 1, chap. 15, 431–457.

LIU, B. AND J. LU (2018): "Pairing Provision Price and Default Remedy: Optimal Two-Stage Procurement with Private R&D Efficiency," *RAND Journal of Economics*, 49, 619–655.

LYON, T. P. (2006): "Does Dual Sourcing Lower Procurement Costs?" *Journal of Industrial Economics*, 54, 223–252.

MANKIW, N. G. AND M. D. WHINSTON (1986): "Free Entry and Social Inefficiency," *RAND Journal of Economics*, 17, 48–58.

MCCALLUM, A. K. (2002): "MALLET: A Machine Learning for Language Toolkit," Software package, `http://mallet.cs.umass.edu`.

MCMILLAN, J. (1990): "Managing Suppliers: Incentive Systems in Japanese and U.S. Industry," *California Management Review*, 32, 38–55.

MOLDOVANU, B. AND A. SELA (2001): "The Optimal Allocation of Prizes in Contests," *American Economic Review*, 91, 542–558.

——— (2006): "Contest Architecture," *Journal of Economic Theory*, 126, 70–96.

ROGERSON, W. P. (1989): "Profit Regulation of Defense Contractors and Prizes for Innovation," *Journal of Political Economy*, 96, 1284–1305.

——— (1994): "Economic Incentives and the Defense Procurement Process," *Journal of Economic Perspectives*, 8, 65–90.

——— (1995): "Incentive Models of the Defense Procurement Process," in *Handbook of Defense Economics*, ed. by T. Sandler and K. Harley, Elsevier, vol. 1, chap. 12, 309–346.

SCHUMPETER, J. (1939): *Business Cycles*, London: Allen Unwin.

SCHWARTZ, M., W. GINSBERG, AND J. F. SARGENT (2015): "Defense Acquisitions: How and Where DOD Spends its Contracting Dollars," CRS Report, Congressional Research Service.

SHAKED, A. AND J. SUTTON (1984): "Involuntary Unemployment as a Perfect Equilibrium in a Bargaining Model," *Econometrica*, 52, 1351–1364.

TAYLOR, C. R. (1995): "Digging for Golden Carrots: An Analysis of Research Tournaments," *American Economic Review*, 85, 872–890.

VIVES, X. (2008): "Innovation and Competitive Pressure," *Journal of Industrial Economics*, 56, 419–469.

WALLSTEN, S. J. (2000): "The Effects of Government-Industry R&D Programs on Private R&D: The Case of the Small Business Innovation Research Program," *RAND Journal of Economics*, 31, 82–100.

WILLIAMS, H. (2012): "Innovation Inducement Prizes: Connecting Research to Policy," *Journal of Policy Analysis and Management*, 31, 752–776.

ZAHUR, N. B. (2020): "Long-Term Contracts and Efficiency in the Liquefied Natural Gas Industry," Working paper, Queen's University.

A. Data and Descriptive Statistics

A.1. Details on Data Collection

I provide further details about the data collection and cleaning procedure and about how datasets from different sources are cross-checked and merged.

SBIR Data from the Office of Naval Research. The website www.navysbirsearch.com from the Office of Naval Research (ONR) has information about all SBIR and STTR (Small Business Technology Transfer) contracts let by the Navy, as described in Section 2. I scraped the data from the website and corrected obvious mistakes, including fixing typos in contract numbers and dropping duplicates.[27] I define a contest to consist of all Phase I, II, and III awards given under the same topic number, and I can track a firm through the three phases using its unique DUNS number. I include both SBIR and STTR contracts in the analysis.[28]

Federal Procurement Data System. From the Federal Procurement Data System (FPDS) via www.usaspending.gov, I downloaded contract data for all contracts from the DOD from 2000 onwards.[29] I use this dataset as the source for contract values: data from the ONR sometimes simply lists a standard SYSCOM-specific award amount. From this dataset, I extract all contracts where the contract number matches one from the ONR dataset. I then check that the DUNS number and the firm name match for the merged contracts. For cases that are not exact matches, I verify through online searches that the difference can be attributed to a name change or an acquisition.

For each contract in the FPDS, I compute the total funding provided though the contract by summing across the dollars obligated in the base contract and all contract modifications. For the majority of contracts, I use this measure as the contract value. This amount usually agrees with the ONR data to within $1, and the amounts differ by less than 5% for the majority of the remainder. Many of the remaining discrepancies can be explained by a single contract modification recorded in the FPDS data. I use the data from the ONR as the measure of contract values if (i) I am unable to verify whether the merge is correct, (ii) the FPDS yields a contract amount that is less than 25% of the ONR data, or (iii) the base

[27]The data in this analysis was scraped from www.navysbirsearch.com on February 18, 2015.

[28]In a small number of cases, two different Phase III SBIR awards (two different companies) were listed with the same contract number but under separate contests. I treat these joint awards as separate awards for each contest. There are also a few contests in which the number of Phase II competitors exceeds the number of Phase I competitors. Since every firm that wishes to compete in Phase II must also have competed in Phase I for the Navy in my sample period, I assume that the competitor who appears first in Phase II actually was awarded a Phase I contract too.

[29]The data in this paper was last downloaded from www.usaspending.gov on May 24, 2015.

Title	Keywords
modeling	modeling, simulation, analysis, software, prediction
aircraft	aircraft, control, unmanned vehicles, flight, operations
data	data, network, software, architecture, security
power	power, energy, heat, thermal, cooling
acoustics	acoustics, sonar, underwater, submarine, anti-submarine warfare
radio	radio, communications, rf, signal, interference, frequency
materials	composite, corrosion, coating, materials, structures
optics	optics, laser, fiber, infrared, wavelength
ballistics	armor, gun, shock, fire, blast
engines	engine, turbine, aircraft control, engines, propulsion
battery	fuel, battery, water, energy storage, cell

Table A.1: Representative topics generated by the LDA algorithm in MALLET. MALLET only returns a representative list of words corresponding to each topic; I assign the topic name myself.

contract is missing in the FPDS. Otherwise, I use the contract amount in the FPDS.[30]

SBIR Solicitations. I copied the full text of the Navy SBIR solicitations from the DOD archive of solicitations, from 1999 onward.[31] For each topic number, I created a document containing all abstracts from all winning firms for all phases related to the topic as well as the full solicitation. This set of documents comprises the "corpus" that I fed into the software package MALLET (McCallum, 2002) to generate the technology topics.[32]

I train topics using the entire set of contests, including those I exclude from the final sample. I treat each document as a sequence of word features, remove stopwords, and allow for hyperparameter optimization to allow some topics to be more prominent than others. I set the number of topics to 20, but I have done robustness checks on the descriptive regressions using between 10 and 100 topics. MALLET outputs a set of topics, each of which is described by a list of words that categorize these topics. Of the 20 topics, 19

[30] Some Phase III contracts are for exceptionally low values, usually $50,000. They are overwhelmingly near the beginning of my sample, when the Phase III contract in the data did not always correspond to the delivery contract, and I am unable to systematically identify the actual delivery contract in the FPDS. I treat values of Phase III contracts less than $1 million as "missing at random," saying that I do not observe the value of Phase III contract, although one is known to exist. I have checked that descriptive regressions (Column (6) in Table A.3) are robust to including these contracts.

[31] See http://www.acq.osd.mil/osbp/sbir/solicitations/index.shtml.

[32] Briefly, this LDA algorithm takes as inputs a set of documents, each of which it treats as a sequence of words, and a fixed number of latent topics. It places Dirichlet priors on the distribution of topics for each document and on the distribution of words for each topic. The data generating process the model specifies is roughly one in which each document is a mixture of topics and each topic is a mixture of words: a document can be generated by recursively (and multinomially) selecting a topic from this mixture and then selecting a word from this topic. Since the Dirichlet distribution is the conjugate prior for the multinomial, this model lends itself to a computationally attractive sampling procedure to generate topics and assign documents to mixtures of these topics. Further details can be found in Blei et al. (2003).

correspond to technology areas; the most popular one, however, consists of generic terms such as "system", "phase", "technology", "design", and "navy". I drop this topic from the list and use the remaining 19 topics. MALLET also assigns each document d a weight p_{dt} for each topic t such that $\sum_t p_{dt} = 1$ for all d. I renormalize these proportions after eliminating the generic topic and am left with a set of 19 "fixed effects" for each contest.[33] Table A.1 lists some representative topics in the dataset (when generating 20 topics), along with common words associated with each topic.

A.2. Descriptive Statistics and Correlations

Table A.2 presents summary statistics, reinforcing observations in Table 1 and Figure 1.[34] Most contests involve small numbers of competitors: there are on average 2.5 competitors in Phase I and 1.1 in Phase II. Phase II and III contracts average about \$800,000 and \$8.8 million, respectively, with large standard deviations. The Naval Air and Sea Systems Commands solicit about 60% of the contests. Table A.2 also lists the proportions of the six most common topics. No topic dominates the contests, as the means are not much larger than 0.05, which we would approximately expect if documents were randomly assigned. The median value for each topic in the dataset is tiny, suggesting that contests are typically not assigned to many topics and that the LDA algorithm does discriminate between topics.

Table A.3 reports linear probability models of the contest transitioning to a particular phase on funding and competition, controlling for contest-level heterogeneity using year and SYSCOM fixed effects and the topics. Column (1) indicates that increasing the number of competitors in Phase I by 1 is associated with an increase in the probability of at least one firm advancing to Phase II by 6.6 pp, compared to a mean of 83%. Column (2) shows that adding a Phase II competitor is associated with an increase in the probability of transitioning to Phase III by 7.5 pp, a large number compared to the mean success rate of 10.5%.

I then investigate the probability that an individual competitor generates successful research, using censoring models to estimate the probability $p(X_{ij})$ that a contestant i succeeds in contest j as a function of contest-level covariates and individual-level funding.[35]

[33] These variables are proportions rather than binary variables.

[34] Table A.2 restricts the sample to all contests let between 2000 and 2012, dropping 17 contests with more than 1 Phase III awardee. In Table A.3, I further restrict the sample to match the one used in structural estimation by only considering contests that have no more than 4 Phase I competitors.

[35] Success at the individual level is not observed directly, as the number of entrants into Phases II and III are capped, so not all successful innovations receive a contract. For the transition from Phase I to II, I estimate a censored binomial model in which for each contest j, the unobserved number of successes N_{Sj} is such that $N_{Sj} \sim \text{Binomial}(N_1, p(X_j))$, but the observed quantity is $N_{2j} = \min\{N_{Sj}, \bar{N}_{2j}\}$. I leverage the 40% rule

	N	Mean	Median	SD
A. Number of Competitors				
Phase I	2875	2.51	2	1.09
Phase II	2875	1.09	1	0.74
Phase III	2875	0.088	0	0.283
B. Contract Amount (Millions)				
Phase II	3143	0.803	0.749	0.453
Phase III	252	8.77	2.93	13.24
C. Fiscal Year ≤ 2006	2875	0.505	1	0.500
D. Systems Command				
NAVAIR	2875	0.329	0	0.470
NAVSEA	2875	0.273	0	0.445
E. Topics				
Information/Data	2875	0.080	0.00150	0.182
Materials/Composites	2875	0.074	0.00015	0.188
Algorithms/Sensing	2875	0.071	0.00044	0.163
Aircraft	2875	0.064	0.00226	0.147
Manufacturing	2875	0.063	0.00017	0.167
Power/Energy	2875	0.061	0.00019	0.162

Table A.2: Summary statistics for the dataset of all solicitations posted between 2000 and 2012, dropping ones in which multiple Phase III contracts were awarded.

	Contest Success		Individual Success		Log(Amount)	
	Phase I	Phase II	Phase I	Phase II	Phase II	Phase III
	(1)	(2)	(3)	(4)	(5)	(6)
# Phase I Comp	0.066	-0.017	-0.127	-0.024	0.016	0.219
	(0.009)	(0.008)	(0.008)	(0.017)	(0.012)	(0.109)
# Phase II Comp		0.075		0.027	-0.003	-0.393
		(0.016)		(0.012)	(0.016)	(0.173)
Log(Average Phase II Amount)		0.156				0.322
		(0.018)				(0.191)
Log(Individual Phase II Amount)				0.114		
				(0.010)		
R^2	0.082	0.128			0.131	0.417
N	2773	2292	2773	2292	2292	151

Table A.3: Descriptive regressions of whether the contest enters Phase II ((1) and (3)) or Phase III ((2) and (4)) on the number of competitors, restricting to contests with no more than 4 Phase I competitors. Columns (2) and (4) restrict to contests that enter Phase II. Columns (5) and (6) consider the contract amount in Phases II and III. All columns control for year FEs, SYSCOM FEs, and topic covariates.

described in Section 4.2 to determine the limit $\bar{N}_{2j}$ on Phase II competition. I estimate this model via MLE, controlling for the same contest-level covariates, and I report $p(\cdot)$. I model the transition from Phase II to III

Column (3) shows that adding one competitor to Phase I is associated with a decrease in the probability of an individual competitor generating a successful innovation by 12.7 pp. Column (4) indicates that contestants in contests with one additional Phase II competitor have a higher probability of success, by 2.7 pp.[36]

How does competition affect Phase II contracts? Column (5) shows that contests with one more Phase I competitor have on average 1.6% larger contracts. Thus, adding more Phase II competitors has little impact on average funding overall, although one can check there is a large and significant drop when moving from contests with 3 to contests with 4 Phase II competitors. Since the model interprets Phase II contracts as indicative of the DOD's value for the project, and since more funding should also directly increase success probabilities, we expect that funding correlates positively with success in Phase III. Indeed, Columns (2) and (4) show that increasing funding by 10% is associated with an increase in the contest-level success rate of 1.6 pp and in individual success by 1.1 pp.[37]

Column (6) regresses the Phase III contract amount against Phase II amounts and measures of competition.[38] If projects with more Phase I competitors have higher values (because higher-value projects survive into later rounds), we would expect Phase I competition to be correlated with an increase in the value of the procurement contract. Consistent with this prediction, I estimate that adding one Phase I competitor is associated with an increase in the Phase III contract by about 22%. Increasing Phase II competition would give the DOD more chances for a lower draw of the delivery cost and also let it leverage competition, thus leading to lower Phase III contracts. Adding a Phase II competitor is indeed associated with a reduction in the Phase III contract by about 39%. Finally, more Phase II funding is associated with both higher-value projects and likely better draws of cost (via more research), and the net effect on the Phase III contract is ambiguous. On net, I find that a 10% increase in average Phase II funding is associated with a 3.2% increase in the Phase III contract.

as follows: a contestant i generates a successful innovation in contest j with probability $p(X_j; t_{ij})$, where t_{ij} is the Phase II funding; if multiple contestants succeed, one contestant is awarded the Phase III contract uniformly at random, which is a simple benchmark.

[36] Columns (2) and (4) show that success rates from Phase II to III are slightly lower when there is one more Phase I competitor. The source of the negative coefficient on N_1 is primarily contests with $N_1 = 4$. If anything, this correlation highlights the endogeneity concern that contests with different numbers of Phase I competitors could be systematically different.

[37] I have checked that this observation holds within-contest as well. A nonparametric regression of Phase III success at the contestant level on the ratio between that contestant's Phase II funding and the lowest Phase II funding in that same contest, controlling for contest covariates, yields a (slightly noisy) positive relationship.

[38] I run this regression only on observed Phase III contracts, without directly accounting for the selection into having a Phase III contract. This provides a description of the data, and this correlation also helps understand the parameter estimates in Section 5.

Overall, these observations are consistent with the main features of the model: firms conduct research to learn values of their projects in Phase I, there is heterogeneity engendered by R&D, the strongest firms move on to Phase II and are awarded larger Phase II contracts, and Phase II research increases success rates by lowering delivery costs. However, because of the endogeneity in competition discussed in Section 2, I do not view these statistics and correlations as causal, and they must be interpreted with caution.