// NO.21-059 | 07/2021

DISCUSSION PAPER

// LEONARDO GIUFFRIDA AND EMILIO RAITERI

Buyers' Workload and R&D Procurement Outcomes: Evidence From the US Air Force Research Lab

Leibniz Association

ZEW

Buyers' Workload and R&D Procurement Outcomes: Evidence from the US Air Force Research Lab*

Leonardo M. Giuffrida and Emilio Raiteri

July 29, 2021

Abstract

Does workload constitute a bottleneck to a public agency's mission, and if so, to what extent? We ask these questions in the context of the US government's procurement of R&D. We link tender, contract, patent, and office records to the identity of the officer responsible for the procurement process to estimate how workload in the federal acquisition unit affects the execution of R&D contracts. The identification comes from unanticipated retirement shifts among contracting officers, which we use to instrument workload. We find a large increase in patenting at the extensive margin when the same officer is exposed to a declining workload. In our sample, an additional contracting officer in the procurement unit, holding fixed the procurement budget and number of purchases, leads to a two percentage point increase in the probability for an R&D contract to generate patents. We provide suggestive evidence that backlogged contracting officers are unable to devote enough time to tender and contract specifications.

JEL codes: D23; H57; O31, O32.
Keywords: Workload, Public Procurement, Contracting Officer, R&D, Patents.

*Leonardo M. Giuffrida, ZEW Mannheim and MaCCI, email: leonardo.giuffrida@zew.de; Emilio Raiteri, School of Innovation Sciences, Eindhoven University of Technology, email: e.raiteri@tue.nl. For their helpful comments and suggestions, the authors would like to thank Adriano De Leverano, Francesco Decarolis, colleagues at their institutions as well as seminar participants at departmental seminars (ZEW) and conferences (KU Leuven, MaCCI) where earlier drafts of this study were presented. Giuffrida is thankful to the Leibniz Association for its support to the project Market Design by Public Authorities. The usual disclaimers apply.

1 Introduction

Governments frequently use their extensive purchasing power to promote an array of policy objectives (Simcoe and Toffel 2014). In particular, technical and scientific progress often depends on public spending, as shown by recent studies. Among the most influential, Fleming et al. (2019) show that in the US, federal government research increasingly drives the inventive activities that ultimately lead to jobs, industrial competitiveness, and entrepreneurial success. Moretti et al. (2019) provide evidence that government-funded R&D spurs private investment in R&D. The authors also suggest that defense-related R&D spending is the most important *de facto* industrial policy by which the US government influences the pace and direction of innovation in the country. Indeed, federal R&D spending reached nearly $130 billion in 2018 and accounted for about twenty percent of all R&D conducted in the US that year.[1]

One of the main channels through which the government funds innovative activities is through the procurement of R&D from business and higher education institutes. In de Rassenfosse et al. (2019), public procurement is reported to have accounted for about one-third of the US federally funded R&D in 2015, with the remainder distributed among research grants, cooperative agreements, and internal research at government laboratories (Bruce and de Figueiredo 2020). Even though, according to the Federal Acquisition Regulation (FAR) 35.003, the primary objective of R&D procurement is to acquire supplies or services for the direct benefit or use of federal agencies, policymakers and scholars increasingly view this type of public spending as a critical innovation policy tool with important spillover effects on the private sector and significant impacts on end-users. Anecdotal and historical evidence suggest that US federal procurement has been critical to the development of some of the most influential technologies invented in the 20th and 21st centuries, from computers and the Internet to GPS systems and robotics. Recent empirical contributions find support for the existence of such a link, indicating that R&D procurement can influence firms' innovative behavior and affect the direction of technological change (Raiteri 2018).

Given the relevance of R&D procurement to innovation, understanding the factors determining the success or failure of this type of government contracting has implications for policy. Innovation scholars emphasize the role of public agencies (i.e., the buyers) in clearly identifying their needs and specifying functional requirements for the procured activity that enable innovative solutions (Edquist 2015). An extensive strand of theoretical and empirical work on standard procurement suggests that the decisions made by the official responsible for the procurement process—commonly referred to as the contracting officer (CO)—may have a strong impact on the performance of the contract (Decarolis et al. 2020; Best et al. 2017; Bucciol et al. 2020; Baltrunaite et al. 2021). This thread of research is motivated by the CO being the leading party in the contract planning, solicitation, and award stages of the procurement process (Rendon et al. 2012), in addition to being actively involved in the needs assessment and contract execution phases. This means that COs are responsible for preparing pre-award solicitations and requests for proposals, selecting the procedure for contract award (e.g., sealed bidding or negotiations), determining contract terms, influencing the level of competition, identifying vendors, and issuing the contract, among the other key pre-award decisions. COs have wide discretionary power in performing these tasks, and their

[1] Source: National Patterns of R&D Resources: 2017-18 Data Update, National Science Foundation.

decisions may arguably affect contract performance outcome.

The demand-side relevance in the context of publicly funded R&D is highlighted by Bruce et al. (2019), who emphasize that federally funded R&D grants in the US have poorer innovation output than cooperative agreements, for which the buyer has greater discretion over the project.[2] Accordingly, the buyer's role should be even more central to procurement contracts, as they imply the highest degree of control over the R&D process compared to cooperative agreements and grants. Following this reasoning, Decarolis et al. (2021) provide the first empirical quantification of the importance of public buyers for R&D procurement success. In particular, the authors find that a disruptive event in the purchasing office—i.e., the death of relevant contracting employees—reduces the probability of a contract to deliver patented inventions. The effect they find appears to be driven by the pre-award phase, on which the CO has, in fact, the greatest impact. However, although Decarolis et al. (2021) show that contracting authorities matter for R&D procurement, the data and methodology used do not provide clear evidence on which agency characteristics affect the success of a federal R&D contract.

Existing literature indicates that workload can be a severe source of capacity constraints in contracting. As the workload increases, the COs can spend less time on each task they are responsible for, leading to suboptimal contract specifications and poorer performance. Contributing to the discussion of how institutions shape new knowledge (Furman and Stern 2011), this paper quantifies the impact of bureaucratic workload on the innovativeness of procurement of R&D in the context of a developed economy such as the US and, to do so, combines several different data sources for the first time. First, we use the Federal Procurement Data System (FPDS), which contains information on every contract awarded by US federal agencies. Second, we use the 3PFL Database of Federally Funded Patents (3PFL) collected by de Rassenfosse et al. (2019)—and employed in Decarolis et al. (2021)—to trace patented inventions associated with a federal contract and thus measure R&D contract performance, in line with the literature (Bruce et al. 2019). Third, we collect extensive information about the tender stage of the awards from the Federal Business Opportunities (FedBizOpp) website, including information about the identity of the CO. The latter information is more readily available for contracts awarded by the US Department of the Air Force, one of the major subdivisions of the US DoD. As nearly half of all R&D contracts awarded by the Air Force are from the Air Force Research Laboratory (AFRL), and because of the rich information on COs in that subsample necessary for our empirical strategy, we focus on contracts awarded by the contracting offices that compose the AFRL. We count the number of COs actively involved in procuring R&D projects in a given year: Following Warren (2014) and controlling for the number of R&D awards and the office's budget, we use this measure as an inverse indicator of the office's workload. After this data selection and merging process, we end up with a sample of 1,970 R&D contracts representing the universe of AFRL's R&D procurement processes featured with COs' records for 2005 through 2012.

Identifying the effect of the buyer's workload on R&D contract performance presents multiple challenges. First, the quality of the CO assigned to a particular contract is key in our setting as it is correlated with both the outcome and the workload. In case of a sudden and unpredictable

[2] Grants or cooperative agreements should be used when the primary purpose of the transaction is to stimulate or support R&D for another public purpose (FAR 35.003)

jump in workload, the office manager (also referred to as the program manager) could still assign the more complex projects that require extra effort and expertise to the highest quality (or most experienced) COs who have above-average productivity. In such cases, we would then underestimate the potential effect of a shock in workload for offices staffed with more high-quality COs and overestimate the shock for offices with fewer high-quality COs. Thanks to the wealth of information made available by FedBizOpps, we solve this issue by including CO fixed effects. A second unobserved factor that may introduce bias into our estimates involves the average complexity of the yearly procurement activity of the office, which is unobservable to the econometrician but likely anticipated by the program manager who plans the budget and employment in advance. This omitted information is also correlated both with the patentability odds of a project and workload. To handle this additional source of bias, we use an instrumental variable (IV) method that exploits exogenous changes in contracting employment based on unexpected retirement-postponing decisions. For this purpose, we use a fourth data source (i.e., FedScope) which contains detailed characteristics of the public workforce. In particular, we construct an instrument that builds on the fact that retirement decisions among federal employees are strongly influenced by i) the attainment of the threshold years of service that qualify workers for immediate pension benefits and ii) idiosyncratic motives. Although eligibility for retirement is anticipated by the program manager who makes hiring decisions in advance, actual retirement could be postponed, and according to the relevant literature, managers have little influence on individual retirement decisions (Lewis and Pitts 2018). Therefore, we consider the difference between the number of COs eligible for retirement and actual retirees to be a good potential indicator of an unanticipated workload shock. Specifically, the larger the difference between the number of COs eligible to retire and the actual retirement counts in the same office, the larger the number of COs who will decide to postpone retirement in divergence from the expectations of the office manager. As managers' hiring decisions are based on the scale of expected retirement, the larger this difference, the larger the short-run positive shock to the number of COs active in an office.

Our IV estimation strategy allows us to estimate a causal effect of a purchasing unit's workload on R&D contract outcomes, which is more than one order of magnitude larger than the corresponding endogenous model estimates. Our results stress that the same CO exposed to an increase in the workload—proxied by a decrease in the number of other COs employed, holding fixed the number of R&D purchases by the office and their total obligated amount—determines a decrease in the average probability of awards delivering a patented invention. Specifically, one additional CO in the procurement unit (corresponding to 3 percent of the average number of COs in the office-year pair in our sample) leads to a two percentage point increase in the probability for an R&D contract to generate patents. To provide a more transparent economic interpretation of the estimates, we consider what would happen if we used them to infer the effect of raising the workload of all AFRL's office-year combinations to the level of the office-year with the largest backlog in our sample. If we brought all offices up to its level, this would imply a reduction in the number of patents by 13 percentage points on average per contract (i.e., about 50 percent less likely).

Consistent with our findings, Warren (2014) shows that when procuring supplies and services, contracting offices that experience workload spikes are more likely to choose contract terms that ultimately result in poorer contract performance. Specifically, contracts awarded by busier offices undergo more modifications of the original contract and obligate more dollars on average. In this paper, to the best of our knowledge, we empirically investigate for the first time the relationship between the workload of the CO and the performance of R&D contracts awarded by the same bureaucrat. Given the high complexity of the tasks, the discretionary power of the CO, and the relevance of the pre-award phase in R&D procurement, our hypothesis that workload is particularly disruptive in this setting is validated by empirical analysis.

Our results are robust to several modifications in the empirical specification, emphasize the variation of workload as a driver of post-award procurement outcomes—in this vein, contributing to the related literature[3]— and have relevant policy implications. First, they confirm the evidence provided in the literature on the importance of the buyer and show that the government should pay particular attention to the workload of contracting employees during the tender stage and ensure adequate staffing. Second, our results show that capacity constraints at the contracting office level could limit the ability of the government to translate increases in the budget allocated to R&D work, intended to produce additional technical knowledge, into valuable knowledge and innovation. This, in turn, could dampen the positive spillover effects that publicly funded R&D has been shown to have on the economy (Fleming et al. 2019; Moretti et al. 2019). On a final note, we acknowledge the functioning of public bureaucracies as a key driver of government effectiveness (Weber 1921) and, in turn, the ability of the state to govern effectively as a crucial determinant of economic activity (Acemoglu 2005). Our micro-quantification of the impact of workload on R&D procurement outcomes underscores the disruptive role of a backlogged contracting staff within public organizations in terms of innovative spillovers to the economy and how unblocking such a bottleneck may be relevant for the efficient delivery of public goods.

Our data do not allow us to unambiguously identify the channel through which the additional workload worsens contracting output. In particular, we can neither observe nor proxy the actual effort the CO put in for specific procurement activities and, therefore, we cannot determine if the increases in workload lead to a reduction in the effort provided for a single contract. That said, the heterogeneity of the R&D activities in our sample nevertheless allows us to provide indirect evidence of the importance of the CO in drafting solicitations and contractual agreements in a clear work statement. Such relevance reaches its maximum for contracts awarded for intermediate stages of development, especially for applied research, where contracts for the procurement of R&D work are relatively definable. In such cases, and workload permitting, the CO has room to draft a request for proposal and the resulting contract clearly and completely. To test this hypothesis, we split our sample into three groups based on the R&D stages, Basic Research, Applied Research, and Advanced Development, and run the same model as in the focal analysis. We demonstrate that the overall negative baseline effect of workload on contract performance seems to be largely driven by applied research awards. Although we are aware this result does not provide conclusive evidence, it strongly suggests that backlogged COs cannot devote enough time to tender and

[3]This work relates to the growing economic analysis of the effects of different designs and institutions on procurement outcomes. Examples include awarding design (Decarolis 2014, 2018), wasteful year-end spending (Liebman and Mahoney 2017), external audits (Gerardino et al. 2017), industry consolidation (Carril and Duggan 2020), performance-based insurance schemes (Giuffrida and Rovigatti 2018), and the impact of centralized purchase agreements (Bandiera et al. 2009).

contract specification, with a reduction in the performance of the contracts they award.

The remainder of the paper is as follows. In Section 2, we outline the institutional setup, identification problem, and our research design. Data sources and sample selection are described in Section 3. Section 4 presents our baseline results and the robustness analysis. In Section 5, we discuss a potential mechanisms underlying our findings. Section 6 draws conclusions.

2 Institutional and Empirical Setting

2.1 Trends in the US contracting staff's workload

In the previous section, we briefly discussed the role of COs in the public procurement process and argued that they are particularly sensitive to an increase in workload, especially in the complex realm of R&D contracting, due to the multitude of tasks involved. But problems arising from backlogged contracting personnel would be a second-order problem if spikes in agency workloads were sporadic and temporary events. However, over the past twenty years, federal institutions and scholars have expressed concern about an increasing trend in procurement spending that has not been accompanied by adequate growth in contract personnel.

In 2007, the Acquisition Advisory Panel reported to the Office of Federal Procurement Policy and the US Congress that between 2000 and 2005, total government purchasing volume had increased by nearly 75 percent, from \$219 billion to more than \$380 billion (AAP 2007), while the federal procurement workforce had remained stable during the same period and shrank significantly compared to the 1990s. The panel reported a significant mismatch between the demands on the acquisition workforce and the personnel available to meet them and recommended that an improved human capital planning process be implemented. In 2010, procurement volume reached \$534 billion, and although it declined to \$430 billion in 2015, the upward trend has continued over the past five years, reaching \$579 billion in 2019.[4] Federal spending on R&D followed a similar trend, peaking at \$57 billion in 2010 with a subsequent decline to \$38 billion in 2015, followed by another increase to \$52 billion in 2020. As a result, several agencies still lament an acquisition workforce shortage. A recent Government Accountability Office (GAO) report highlighted that, although the DoD made important changes to its workforce planning to address the AAP 2007's recommendation and increased its procurement workforce by 24 percent between 2006 and 2016, in 2017, it still fell well short of its workforce growth goal, particularly in areas such as contracting and audit (GAO 2017). In 2017, the GAO highlighted persistent problems with the acquisition workforce in its High-Risk list and emphasized that DoD agencies were still facing challenges maintaining sufficient staffing levels and overseeing their acquisition workforce.[5]

The increase in the federal contracting budget and institutional concerns are not the only indications that contracting personnel are struggling with problems stemming from excessive workloads. For example, studies based on survey data confirm that federal procurement personnel cite understaffing as one of the primary problems within their work unit (Rau and Stambersky 2009). Specifically, Rendon et al. (2012) show that in two of the DoD's largest subagencies, the

[4]Source: `www.usaspending.gov`.

[5]In 1990, the GAO began a program to report on government operations identified as *high risk*. The list is used to identify and address serious vulnerabilities in areas where significant resources are expended and critical services are provided to the public. See `https://www.gao.gov/highrisk/overview`.

Department of the Army and the Department of the Air Force, the vast majority of procurement personnel responsible for service acquisition disagree that the size of the procurement workforce is adequate to meet objectives and also disagree that vacant positions are adequately filled.

2.2 Research design and prima facie evidence

As discussed above, peaks in workload strongly influence the discretionary decisions of COs, and previous research has consistently found that, in a non-R&D context, the performance of contracts worsens.[6] The primary objective of this paper is to assess whether an increase in workload within contracting offices procuring R&D negatively affects contract performance.[7] Properly identifying such an effect presents us with several empirical challenges. First, we need to define a satisfactory method to measure both the workload of the procurement office and the performance of a particular R&D project. Second, we need to consider potential factors that might challenge the causal interpretation of a negative relationship between our measures of workload and project performance.

To address the measurement issues, we rely largely on the recent literature. Warren (2014) discusses the complexities associated with constructing a robust measure of workload. In the paper, the author opts for a relatively agnostic approach and uses the size of contracting personnel (i.e., COs plus lower-level employees) in a federal agency while controlling for the number of contracts (*purchases*, from now on) as a proxy for office workload. We follow a similar approach and use the CO employment actively working in the R&D process for a contracting office in a given fiscal year as a proxy for the workload. As in Warren (2014), we control for the number of purchases; in addition, we use the total amount of obligations (*budget*, from now on) of the contracting agency during a fiscal year. The main difference is that our focus is exclusively on R&D contracting. Therefore, our workload measure counts only COs responsible for R&D procurement, and in the same vein, we control only for the number of contracts and dollars obligated by a given office for the procured R&D activities. In contrast to Warren (2014), our data and setting allow us to pinpoint the exact procurement unit of a federal agency; these and other details about the office and the definition of the CO are presented in Section 3.1.

Regarding the measurement of R&D contract performance, we follow the recent contribution of Decarolis et al. (2021) and Bruce et al. (2019). As reported in FAR Part 35, most R&D contracts are focused on goals for which work or methods cannot be precisely described in advance and for which it is not easy to assess the probabilities of success *ex-ante*. Because of the uncertainty that characterizes this process and the resulting high degree of incompleteness, it is not easy to estimate costs accurately. For this reason, FAR recommends the use of cost-reimbursement contracts for R&D procurements rather than fixed-price contracts, which are typically preferred for off-the-shelf procurements and more standardized services.[8] Instead, the primary goal of R&D

[6]FAR 1.602 details the role and responsibilities of the federal CO.

[7]A contracting office is an entity that executes procurement transactions—goods, services, constructions, and R&D—on behalf of the government. A contracting office belongs only to a subagency, that is, the bureau responsible for the transaction.

[8]Fixed-price and cost-plus contracts are two different types of contracts commonly used in procurement. In fixed-price contracts, the buyer offers the seller a predetermined price to complete the project. In a cost-plus contract, no price is set, but the contractor is reimbursed for the cost plus a markup. Cost-plus contracts are generally preferred for the procurement of R&D. Approximately 80 percent of R&D contracts over $1 million awarded yearly by federal

contracts is to advance scientific and technical knowledge and apply that knowledge to the extent necessary to achieve agency goals, not to deliver a product in the most cost- and time-effective manner. Given the unique characteristics associated with R&D contracting, standard measures of contract performance—such as unit price comparisons, cost overruns, time delays, and the number of contract renegotiations—are not well suited to assessing the performance of an R&D contract. As in Bruce et al. (2019) and Decarolis et al. (2021), we consider an R&D contract to be successful if associated with subsequently patented inventions. Although using patents as a proxy for innovation is not without drawbacks, the issuance of a patent application associated with an R&D contract ensures that the contract has generated new knowledge that can be used to solve a particular technical problem.[9] In addition, Peña et al. (2017) confirm that most DoD research offices themselves use metrics such as patent applications and grants to assess the success of their early-stage research and technology projects.

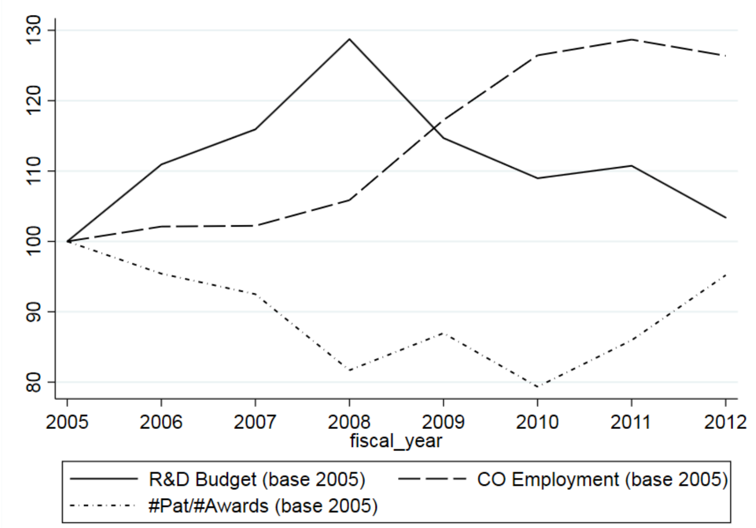
Once we have defined a valid measure for an office's workload and one for the outcome of an R&D contract, we need to consider how to identify the impact of the former on the latter. As a first step, we can match trends in R&D contracting personnel in those federal agencies that regularly purchase R&D from the private sector with the dynamics of R&D procurement spending and the number of patents associated with those contracts. Figure (1) shows these trends. As the figure shows, for general contracting, there were substantial increases in R&D expenditures (solid line) between 2005 and 2008, while the size of the acquisition forces in these offices (dashed line) remained essentially stable. In 2009-2012, the data show a reversal of trends: the size of R&D contracts consistently declined, while the size of the CO pool steadily increased between 2009 and 2012. Therefore, the workload of COs for R&D contracts seems to have increased until 2008 and decreased since then. Interestingly, the number of patents associated with contracts awarded by the offices (dotted line) seems to show a negative correlation with workload: it decreases significantly between 2005 and 2008 and reverses the trend after 2008 and especially since 2010. This could be considered as *prima facie* evidence of a negative impact of COs' workload on the performance of the R&D contracts they award. However, providing evidence that would support a causal interpretation of this relationship requires much more thought.

Ideally, we would look at all federal contracting offices that award contracts for the procurement of R&D and randomly divide them into different groups. We would then assign a different number of additional COs to each group; finally, we would evaluate whether the contracts awarded by the offices whose CO employment grows (shrinks), i.e., offices with a lower (higher) workload per officer, subsequently experience a higher (lower) likelihood of being associated with a patented invention. In such an ideal experiment, randomization of the treatment would ensure that the relationship between workload and contract performance is causal. Unfortunately, even with a small sample of contracting offices, conducting such an experiment is not feasible. Nonetheless, US federal agencies produce a wealth of observational data that, if properly used, would allow us to examine the existence of a causal relationship (if any) between COs' workload and contract performance. Section 3.1 describes the data in detail, but for now, it is sufficient to stress that our

agencies are cost-plus contracts (source: `usaspending.gov`).

[9] A discussion of the drawbacks of measuring R&D activity via generated patents and how we deal with them is covered in Section 3.1.

Figure 1: Workload and Patent Trends



Notes: Average annual R&D obligated budget (solid line), contracting force (dashed line), and number of patents connected to contracts (dotted line). The source is the FPDS database and all agencies awarding at least one R&D contract in the sample are considered.

data provide sufficient information to estimate the following linear probability model:

$$Patent_{i,k,t} = c + \beta \text{Log}\,\#\,\text{CO}_{\text{k,t}} + \gamma \text{X}_{\text{i}} + \delta Z_{k,t} + \zeta_k + \zeta_t + \epsilon_{i,k,t}, \tag{1}$$

where the variable $Patent_{i,k,t}$ reports whether R&D contract i awarded by the office k in the fiscal year t led at least to a patented invention and the $Log\,\#\,CO_{k,t}$ reports the *CO employment* in log-terms. X_i is a vector of controls for contract characteristics, whereas $Z_{k,t}$ is a vector that includes control variables at the office-fiscal-year level, such as the budget and purchases. The vectors ζ_k and ζ_t include a set of office- and fiscal-year fixed effects. A detailed description of the control variables is provided in Section 3.1.

Although the wealth of information available for each contract and awarding office allows us to control for several potential confounding factors that may affect the relationship between workload and contract performance, the use of observational data still presents some fundamental challenges for identification. First, unlike in the ideal experiment, CO employment is not randomly decided. In year $t-1$, the program manager of a particular federal contracting office plans the budget for purchases to be made for the fiscal year t. Specifically, in January of fiscal year $t-1$, federal agencies update the budget plan for t and submit it (as part of the DoD budget) to Congress, which approves it before the beginning of the new fiscal year (i.e., October of the same calendar year).[10] Thus, after taking into account the human resources available at $t-1$, the incoming workload of an office at

[10]The DoD budget is requested along with the budgets of all other departments and constitutes the budget of the US government, which the President submits in early February at $t-1$. Typically, agencies begin preparing their proposals about 18 months before the budget takes effect. The budget must be passed no later than $t-1$ on September 30, just before the start of the new fiscal year. Otherwise, the government will have no budget, it will shut down, and many functions will cease. Then Congress must pass a continuing resolution to temporarily fund the government. Therefore, each office must know its budget before the fiscal year begins in order to operate. The Department of the Air Force, which is the focus of our empirical analysis, sets department-level budget years in advance while also conducting detailed budget estimates for each office that plans for the near future. Source: `https://www.usa.gov/budget`.

time t is known in advance to the contracting office manager. Therefore, the manager can already decide at $t-1$ to hire additional contracting staff (including COs and other procurement staff) in case of an expected growing workload. If the office budget could perfectly measure workload, this would not pose a critical problem for identification as long as we can control for trends in the office budget. However, as discussed by Warren (2014), measuring workload by simply considering the budget would overestimate the workload of officials who draft contracts for the simplest tasks. This problem is particularly relevant in the context of technology procurement, especially in the procurement of R&D. The size of a contract could depend heavily on the fixed costs of the R&D project to be performed, but these fixed costs do not necessarily correlate with the complexity of the task, a contract characteristic that we cannot measure perfectly in our cross-R&D-category data. Although we can control for other contract-specific characteristics that might partially explain the technical complexity of a task, if the complexity level of the average contract awarded by a given office changes over time, the program manager can adjust the workforce accordingly. If the manager knows at $t-1$ that the office will need to award contracts to perform more complex tasks at t and knows that awarding more contracts for more complex tasks will require more clerk time and effort, the office manager can plan to hire additional COs in year $t-1$. The possibility of anticipating the increase in workload due to the increase in average task complexity of the contract implies a positive correlation between our main explanatory variable, $Log\,\#\,CO_{k,t}$, and the omitted factor, i.e., the average complexity of the contracts awarded by an office over time. At the same time, the complexity of the specific task for which a contract is awarded—which in turn is omitted—could be strongly correlated with our outcome variable. More complex tasks are almost by definition more uncertain, and their probability of success is lower than simpler tasks. In the context of R&D contracting, this means that a more complex research project is more likely to lead to inconclusive and thus unpatentable results. As omitted project complexity is negatively correlated with the variable $Patent_{i,k,t}$ and positively correlated with our variable of interest $Log\,\#\,CO_{k,t}$, we would expect the estimate of our coefficient of interest β in equation (1) to be downward biased.

A second unobserved factor that may introduce bias into our estimates involves the quality of the CO assigned to the contract. Even if we could rule out the possibility that the program manager could anticipate the workload and adjust the CO employment to such needs, the manager would still be able to make some adjustments to minimize the impact of an increase in workload. For example, even in the event of a sudden and unpredictable jump in workload, the manager could still assign the more complex projects that require extra effort and expertise to the higher quality (or more experienced) COs who have above-average productivity. Therefore, offices with a greater proportion of high-quality COs would be better able to respond to unforeseen shocks in workload. In such cases, we would then underestimate the potential effect of a shock in workload for offices staffed with more high-quality COs and overestimate the shock for offices with fewer high-quality COs.

To tackle the two threats to identification described above, we implement a two-step strategy. First, we focus on a specific set of offices for which we can obtain reliable information about the identity of the CO who awards a particular contract. In particular, we use data obtained from the FedBizOpps platform (described in detail in the next subsection). The website reports contact information about the CO in charge of a procurement process to facilitate communication between potentially interested contractors and the awarding agency. Unfortunately, for most federal agencies that award R&D contracts, contact information about the CO is reported sparingly and incompletely. However, for a few agencies, in particular the Department of the Air Force, the name and contact information of the relevant COs are reported quite systematically. Taking advantage of this, we focus on the Air Force contracting offices that award the vast majority of R&D contracts, and in particular those offices that are part of the AFRL. Then, for R&D contracts awarded by these offices, we can pinpoint the CO responsible for awarding the contract. This fact allows us to rewrite the equation (1) as

$$Patent_{i,k,t,o} = c + \beta \mathrm{Log}\,\#\,\mathrm{CO_{k,t}} + \gamma \mathrm{X_i} + \delta Z_{k,t} + \zeta_k + \zeta_t + \zeta_o + \epsilon_{i,k,t,o}, \tag{2}$$

where ζ_o is an additional vector of CO fixed effects, allowing us to account for the intrinsic quality—plus other time-invariant idiosyncrasies—of the CO awarding the contract.

To address the residual source of endogeneity provided by the omitted complexity, we adopt an IV approach. In our setting, the instrument must identify a shock in the workload that cannot be predicted by the program manager. Warren (2014) considers a similar problem and approach and proposes the number of retiring contracting employees as an IV. The main idea behind this choice is that, as discussed by Asch et al. (2005), in the US civil service, retirements are mainly driven by the rules that govern pension obligations. More specifically, the retirement decision is strongly influenced by the attainment of the threshold years of service that qualify an employee for immediate pension benefits. However, at time $t-1$, the head of a given contracting office knows how many of the employees are eligible for retirement benefits at t. As both the office's budget and purchases at t are established at $t-1$, the manager would anticipate fluctuations in both the workload and the workforce and make hiring decisions accordingly. If many COs are retirement eligible at t, the manager is likely to hire more staff at $t-1$ to compensate for future retirement. In addition, hiring at $t-1$ is especially important if the size of the retirement-eligible population at t requires it, as hiring a new CO could be a lengthy process. Despite the goal of 80 days set by the Office of Personnel Management (OPM) in 2008, federal agencies took an average of 106 days to hire a new employee in 2017 (127 days for the DoD), with little change from 2004 when the average was 103 days (GAO 2019). This pattern is confirmed in our data: the correlation of retirement eligible at $t-1$ with hiring at $t-1$ and hiring at t is 0.72 and 0.73, respectively.

At t, actual retirement is realized. Actual retirement could be higher or lower than predicted by the program manager, and according to several empirical papers, managers have little influence on idiosyncratic retirement decisions (Lewis and Pitts 2018). Therefore, we consider the difference between the counts of eligible retirement and actual retirement to be a good candidate measure of an unanticipated workload shock (controlling for budget, purchases, and officer quality). The larger the difference between the number of officers eligible to retire in the service at $t-1$ and the actual retirement in the same office at t, the larger the number of COs who (unexpectedly) decide to postpone retirement. As managers' hiring decisions are based on the number of expected retirees, the larger the difference, the larger the short-run positive shock to the number of COs active in an office at t.[11]

[11] In Section 4.2, we show that our results hold when the actual retirement counts among the contracting staff are used as an IV as in Warren (2014).

3 Data

3.1 Data sources and description of variables

The dataset developed for this study combines four data sources for the first time.

FPDS To retrieve contract-specific information, we rely on the FPDS, the source of procurement data of the US government used extensively in recent research, including studies by Liebman and Mahoney (2017), Warren (2014), Kang and Miller (2017), Giuffrida and Rovigatti (2018), Decarolis et al. (2020). Since fiscal year 2000, federal agencies have been required to complete procurement action reports, which in turn feed into the FPDS. The FPDS covers all federal contracting agency transactions related to an award above the federal micro-purchase threshold.[12] Like Decarolis et al. (2021), we focus only on the pool of R&D contracts from FPDS. The R&D code (and stage) specified for each award comes from the FPDS variable "Product or Service Code."[13]

Moreover, most of the other information we use in our empirical analysis to build covariates or fixed effects comes from FPDS. We observe the *Fiscal Year* of the project award; the expected cost at the award stage (*Award Amount*) and the final cost, computed as the cumulative sum of the *Award Amount* with all subsequent price renegotiations; the expected and actual duration of a project (*Expected Duration* and *Final Duration*, respectively). *Last Week* identifies a project if it is awarded in the last week of the fiscal year (i.e., the last seven days of September).[14]

3PFL de Rassenfosse et al. (2019) exploit the FAR to trace patented inventions directly related to federal contracts. The 3PFL database covers USPTO patents granted between 2005 and 2018. More specifically, we use the information contained in the 3PFL database to construct our performance measures for our sample of R&D contracts with the extensive margin of patents, i.e., *Patent*, which is a dummy indicating that the project is associated with at least one registered patent. We also observe the total number of patented inventions associated with a given federal R&D contract.

Two main concerns might cast doubt on the suitability of patents as a proxy for the innovative output of an R&D contract in our context. First, a contractor might choose to favor secrecy over

[12] The value was $3,000 during the period under analysis. In 2015, it was revised to $3,500; in 2020, it was revised again and increased from $3,500 to $10,000. The amounts for public R&D projects are typically very high, and we can confidently state that we observe the universe of these projects over the period under analysis.

[13] The variable consists of two alphabetic and two numeric digits. The first digit is always the letter "A" to identify R&D; the second digit is alphabetic "A" through "Z" to identify the major category of R&D; the third digit is numeric, 1 through 9, to identify a subdivision of the major category of R&D. The categorical variable *R&D Category* is defined according to the combination of the first three digits. The fourth digit is numeric, 1 to 7, to identify the corresponding level of R&D with: (1) Basic Research; (2) Applied Research and Exploratory Development; (3) Advanced Development; (4) Engineering Development; (5) Operational Systems Development; (6) Management and Support; (7) Commercialization. The categorical variable *Stage* is generated accordingly. The R&D usually includes the first six categories. According to the FPDS Product or Service Code Manual, the first stage (i.e., basic research) includes all scientific endeavors and experiments aimed at expanding the body of knowledge and forms part of the basis for subsequent applied research and exploratory and advanced development in the various disciplines, as well as new or improved functional capabilities. The second stage includes all efforts directed toward solving specific problems except major development projects. Advanced development includes all efforts directed at projects that have transitioned to hardware development for testing, for example. The primary outcome of this type of effort is proof of design concept and/or prototype.

[14] This control variable is similar to Liebman and Mahoney (2017), who highlight how the federal budget expiring at the end of the fiscal year creates incentives for government buyers to rush to spend resources on low-quality projects.

patenting to protect its invention. However, FAR 27.3 states that a contractor must timely file a patent application and disclose it to the government to retain title to an invention made under a government contract. If the contractor fails to do so, it risks losing the title to the invention as the government has the right to file a patent application on its behalf. So there are strong incentives for the contractor to file a patent application when an invention materializes. Second, the government itself may recommend the contractors keep the invention secret in the interest of national security. In such cases, even if the contractor has duly filed a patent application, the Patent Office imposes a secrecy order that halts (at least temporarily) the patent prosecution process.[15] Nevertheless, as discussed in de Rassenfosse et al. (2020) and Decarolis et al. (2021), the actual number of secrecy orders issued each year is quite small and only a limited number of these orders appear to target the output of federal R&D contracts.

FedBizOpps The federal acquisition process begins when an agency determines its requirements and how to procure them. If the agency's CO determines that the appropriate method for procuring the goods or services is a contract, and the expected value is greater than $25,000, then the contracting authority is required by the FAR to post a solicitation on the FedBizOpps platform.[16] The FedBizOpps can be thought of as the government's call for tenders point-of-entry, and its purpose is to collect, maintain, and disseminate information to the public about federal solicitations. System information is used to administer and manage access by federal buyers, maintain lists of interested vendors, and notify vendors of federal solicitations of business interest. Government contractors use FedBizOpps as a search engine to find immediate solicitations or bid opportunities as well as archived records.

A subset of solicitations on FedBizOpps reports an additional piece of information: the identity of the bureaucrat responsible for the solicitation process, i.e., the CO (and the associated contract identifier once awarded and tracked by FPDS). This point of contact is located at the bottom of the solicitation documents and includes the first and last name, title, phone number, and email address of the CO. For R&D procurement solicitations, this information is particularly rich for the activity of the AFRL, and we refer the reader to Section 3.2 for more details on the sample of contracting offices under study. We use this valuable information to calculate the total number of COs actively working on the procurement of R&D in AFRL purchasing units in a given fiscal year. We have already defined this variable as *# COs* and referred to it as CO employment. As discussed in Section 3.3, this variable is the main explanatory variable for the project innovation outcome under study. In addition, we define *Specialist* as a dummy variable indicating whether the CO is assisted by a contract specialist for a particular R&D process.[17]

[15] The Invention Secrecy Act of 1951 governs this process.

[16] FBO.gov moved to the beta version SAM.gov in November 2019—after our data collection— and is now known as Contract Opportunities. See also Carril (2021) for a discussion of the FedBizOpps' reporting threshold.

[17] CO and contract specialist fulfill different roles in the procurement process. A CO is an individual who can bind the US government to a contract and has signature authority as a government contract agent. In the federal procurement career, the contract specialist acts as a management consultant and assists the CO in planning to acquire needed goods and services. Only the CO is authorized to sign and administer the contract once awarded. Thus, the contract specialist is a lower-grade contract bureaucrat. A contract specialist is not always necessary, while a CO with signature authority is always necessary for a contract to be initiated.

FedScope The Office of Personnel Management—an independent federal agency that acts as the central human resources department for the executive branch— collects, maintains, and publishes data on approximately 96 percent of civilian federal employees. These data are published through the federal Human Resource database (FedScope), which is the most comprehensive source available on the size and scope of the US federal workforce.[18] Fedscope is the fourth data source we use. The data are divided into five subject categories (called "cubes"), of which we consider only the Employment cube and the Separations cube. The Employment cube contains various demographic characteristics and information about appointments and assignments, such as length of service, job category, pay grade, pay level, type of appointment, work schedule, and location of each employee. The Separations cube contains all separation events (inflows and outflows), that is, employees who are transferred to other offices or agencies, resigned voluntarily, retired, experienced a reduction in force, were terminated, or died during employment. In both cubes, we focus on GS11-02 employees, which is the government's job classification series for contracting and acquisition personnel, including COs and all second-order contracting jobs. We will label the sum of GS-1102 employees as *Contracting Employment*. This is an alternative and less conservative measure of contracting employment—used by (Warren 2014)—to check the robustness of our results to the definition of the endogenous employment variable.

Importantly, the FedScope dataset allows us to build our IV as outlined in Section 2.2. The retirement eligibility of federal civilian employees is determined by age and the number of years of creditable service. People that have reached the minimum retirement age are eligible for immediate retirement benefits provided they have at least 10 years of creditable service. The minimum retirement age at the beginning of the period we consider (i.e., 2005) was 55 years and 6 months, and it was 56 at the end of the period (i.e., 2012).[19] To identify the number of retirement-eligible COs in a given office—which we define as *Retirement Eligible*— we exploit the information about the age group and the years of creditable service of the employees in the GS-1102 category as reported in the Fedscope database employment cube.[20] Moreover, *Retirement Actual* instead indicates the *Contracting Employment* who retire during a given fiscal year. Finally, *Non-retirement* is defined as the difference between lag Retirement Eligible and Retirement Actual, and is our IV. Before showing the results, we describe in detail our sample of AFRL's bureaus and present some relevant descriptive facts about the data used to connect FedScope and the other data sources.

3.2 The Air Force Research Lab

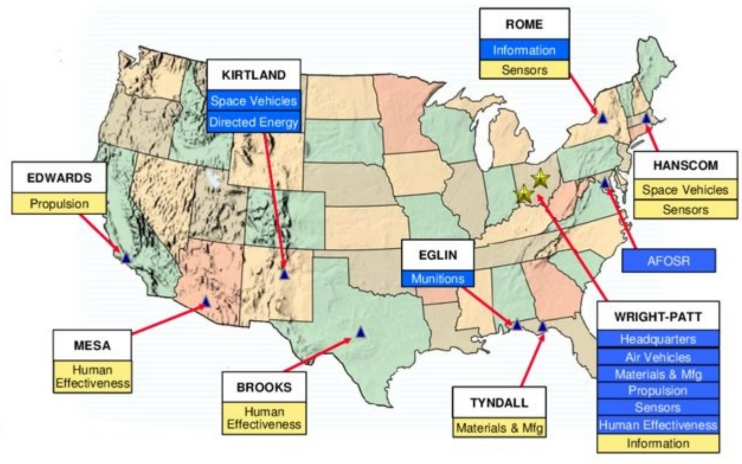
In this paper, we require a metric for workload within the contracting office. FedScope data are available at the subagency level—referred to as "AGYSUB"—but the geographic information in FedScope allows us to determine the location (i.e., state) of each federal employee. Since FedScope information does not allow us to pinpoint offices below the level of the subagency, we must use the

[18]The database has exclusions that affect, for example, some national security and intelligence agencies and the US Postal Service. The data are already used by Decarolis et al. (2020), who also provide a detailed description of FedScope.

[19]For details, see: `https://www.opm.gov/retirement-services/fers-information/eligibility/`.

[20]We count the number of employees in the GS-1102 category that are eligible for immediate retirement benefits, i.e., employees older than 55 years of age and with more than 10 years of creditable service, or older than 62 years of age and five years of service, or older than 65 years of age. Data available at: `https://www.fedscope.opm.gov/employment.asp`.

Figure 2: AFRL Sites



Notes: Operating locations of the AFRL in 2006. Source: Mait (2005).

geographic information to link FedScope to the other data sources.

For the sake of clarity, we emphasize once more that, in our study, we focus on different contracting offices all belonging to one subagency, namely the AFRL, which is a subunit of the Department of the Air Force. The AFRL is a scientific research organization operated by the US Air Force Materiel Command dedicated to the discovery, development, and integration of air and space combat technologies, planning and executing the Air Force science and technology program, and providing warfighting capabilities to the air, space, and cyberspace forces of the US. The laboratory is divided into eight technical directorates and the Air Force Office of Scientific Research, based on various research areas. The latter is primarily a funding body for external research, whereas the other directorates conduct research internally or on behalf of external entities. Figure (2) shows a map of AFRL sites throughout the US territory.

According to the FPDS data, procurement of R&D activities at the AFRL—which in the FPDS raw data amounts to $2.22 billion per year in 2005-2012—is conducted through six different contracting offices in different branches of the AFRL. AFRL headquarters is located at Wright-Patterson Air Force Base in Ohio. Its primary functions are leadership, policy, and guidance. The AFRL's only contracting office in Ohio, i.e., FA8650, as coded in FPDS, is located there. The Space Vehicles Directorate is one of the branches of the AFRL. Its mission is to develop and implement space technologies for more effective and less costly warfighting missions. The Directorate has two headquarters located at different Air Force Bases: Kirtland, New Mexico, and Hanscom, Massachusetts. Both Directorate headquarters conduct R&D procurements, which we track through FPDS with two separate contracting offices (FA9453 and FA8718, respectively). Rome Laboratory is the Air Force "superlab" for command, control, and communications R&D and is responsible for planning and executing the USAF science and technology program. The contracting office FA8750 is installed at Rome Laboratory. AFRL's only R&D procurements in Ohio, New Mexico, Massachusetts, and New York are performed by those offices. Tyndall and Eglin are two Air Force bases, both located in Florida. The AFRL's Florida R&D procurement are conducted by the contracting offices of the two bases (i.e., FA8651 and FA9200). Because these two contracting offices are located in the same state, we are unable to link FedScope information on separations and employment to specific AFRL contracting offices, so they are excluded from our sample.[21] AFRL's four purchasing units provide a diverse science and technology portfolio, ranging from basic and applied research to engineering and operational systems development. Combining FPDS and FedScope data, we define *Same State* as a binary variable—which we include as a covariate in our empirical model—that signals a job performed in the same state where the contracting office is located.

3.3 Sample selection and descriptive analysis

The merging process is as follows. Our starting point is FPDS R&D data. We start by splitting the raw transaction records—i.e., all transactions between government procurement offices and private suppliers—into two main groups: base contracts records and amendment records. The former refer to the first transaction between a procurement office associated with a given contracting process and a supplier and correspond to our unit of observation for this study, the reported characteristics of which constitute the base agreement information. The latter capture all revisions, modifications, or corrections to the contract. All transactions associated with a contract are identified by a unique procurement instrument identifier ("PIID") that marks a signed contract and all its future modifications; therefore, we can track the contracts' entire history from award to completion (or close-out) and link each contract to its revisions. Second, FPDS is combined with 3PFL and FedBizOpps at the contract level. This is very straightforward as both 3PFL and FedBizOpps report the contract PIID. Finally, the intermediate dataset is merged with FedScope. As the level of observation of FedScope is the subagency-state-year, the data must be merged at that level. The nomenclature of FedScope bureaus differs from that of FPDS, but we have relied on an external dictionary that maps the variable "Contracting Office Agency ID" in FPDS to the variable "AGYSUB" of Fedscope.[22] Following the discussion above, we limit our focus to the combined project-level information associated with AFRL awarding agencies FA9453, FA8650, FA8718, FA8750, which represent 88 and 83 percent, respectively, of the spending and contract counts in the AFRL raw sample.

We further restrict the sample according to the following rules: R&D activities conducted within US borders; award amount greater than $25,000; expected contract end date before the end of the sample; no Small Business Innovation Research (SBIR) contracts; 2005-2012; R&D preceding the commercialization phase.[23] This ultimately leaves us with a sample of 1,970 R&D

[21]For a full description of the sample selection, see Section 3.3.

[22]FedScope releases are monthly. To ensure temporal consistency with FPDS and FedScope, we employ the September snapshot of the FedScope cubes as a reference for the closing fiscal year.

[23]The $25,000 threshold is the lowest contract value associated with a contract publicized in FedBizOpps, as described above. R&D contracts are usually very large and this selection prompts the loss of very few observations. Regarding the exclusion of SBIR contracts, these contracts are intended to assist certain small businesses in conducting innovative activities aimed at their eventual commercialization, not their patentability (Howell 2017; Bhattacharya 2018). Contracts awarded before the fiscal year 2005 (i.e., October 1, 2004) are very few and of poorer quality, according to Liebman and Mahoney (2017). Those awarded after 2012 (September 30, 2012) are excluded because public R&D activity in our data lasts more than three years on average and, once completed, potentially produces a patent 18 months later on average. Since 3PFL tracks patents registered through 2018, contracts awarded from fiscal year 2013 onward may not have a patent due to the limited time horizon and not as a poor contract outcome. Finally, contracts in the commercialization phase are excluded from the analysis because they do not

contracts, with a total value of $9.6 billion, 12,020 bids submitted, and 579 unique winners (of which 87 were universities or other higher education institutions). The final sample includes 275 distinct COs, whose associated contracts yielded a total of 522 patents (5 percent of contracts with one patent, 4 percent with two or more).

Table (1) shows the details of the R&D activities included in the sample. Each cell reports the number of contracts for each combination of procurement category and R&D stage and the total number of associated patents in parenthesis. Most contracts and patents are observed for the first three stages of the R&D process, i.e., basic research, applied research (and exploratory development), and advanced development.

Table 1: Cross-tabulation of Contracts and Patents per R&D Category and Stage

Category	Stage 1	Stage 2	Stage 3	Stage 4	Stage 5	Stage 6	Total
AC1: Defense System (Aircraft)	98 (13)	96 (51)	92 (27)	1 (0)	1 (0)	1 (0)	289 (91)
AC2: Defense System (Missile/Space Systems)	88 (48)	38 (1)	7 (1)	0 (0)	3 (6)	3 (3)	139 (59)
AC5: Defense System (Weapons)	0 (0)	1 (0)	1 (0)	0 (0)	0 (0)	0 (0)	2 (0)
AC6: Defense System (Electronics/Communication Equipment)	110 (72)	257 (104)	53 (11)	14 (1)	5 (0)	6 (0)	445 (188)
AC9: Defense System (Miscellaneous Hard Goods)	2 (0)	11 (1)	5 (0)	0 (0)	1 (0)	1 (0)	20 (1)
AD2: Defense Other (Services)	2 (0)	2 (4)	0 (0)	0 (0)	0 (0)	0 (0)	4 (4)
AD6: Defense Other (Construction)	1 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	1 (0)
AD9: Defense Other (Miscellaneous)	128 (6)	572 (93)	169 (59)	5 (0)	33 (2)	0 (0)	907 (160)
AE3: Economic Growth (Manufacturing Technology)	0 (0)	0 (0)	2 (0)	0 (0)	38 (6)	0 (0)	40 (2)
AJ4: General Science/Technology: Engineering	0 (0)	0 (0)	1 (0)	0 (0)	0 (0)	0 (0)	1 (0)
AJ9: General Science/Technology (Other)	23 (2)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	23 (2)
AN1: Medical (Biomedical)	0 (0)	1 (1)	0 (0)	0 (0)	0 (0)	0 (0)	1 (1)
AR1: Space (Aeronautics/Space Technology)	3 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	3 (0)
AZ1: Other R&D	33 (8)	49 (0)	9 (0)	3 (2)	1 (0)	0 (0)	95 (10)
Total	488 (149)	1,027 (255)	339 (98)	23 (3)	82 (14)	11 (3)	1,970 (522)

Notes: Cross-tabulation of the total number of contracts and associated patents (in parenthesis) for each R&D category and stage in our dataset.

Contract amounts are relatively large and highly skewed: 50 percent of contracts have an award price below $998,000, while 10 percent of contract spending is on contracts worth more than $7.4 million. The average award amount is approximately $4 million, but the average total cost, including all subsequent modifications, averages $4.9 million. Correspondingly, the average expected and final contract durations, including any delays, are 1,000 and 1,113 days, respectively. The significant cost increase is typical of R&D activity, which is compounded by the cost-plus nature of most contracts in our sample (95 percent). The prevalence of cost-plus contracts in DoD procurement is well documented (Carril and Duggan 2020). It is explained by the DoD's interest in achieving timely completion of contracts whose cost is highly uncertain at the time of bidding.[24] However, we find that most tendering processes are characterized by full and open competition (81 percent), consistent with the statistics on the entire population (Decarolis et al. 2021). The main characteristics of these contracts—depending on whether they are associated with at least one patent—are presented in Table (2). R&D contracts that lead to one or more patents are on average larger, last longer, and receive more bids. This is consistent with de Rassenfosse et al. (2019), who show that the size and duration of a contract are positively associated with the total number of patents associated with the contract.

Table (3) shows the characteristics of the contracting units. Each office spends an average of $0.54 billion per fiscal year on 65 different R&D projects. Contracting Employees are 292, 32 of which are certainly COs according to available FedBizOpps records. Retirement Eligible represents

consist of an R&D process, only the commercialization of the output. However, their share of the raw transaction data is only approximately 3 percent.

[24]See Bajari and Tadelis (2001) for a detailed study of the trade-off between time and cost to completion induced by contract pricing format.

Table 2: Summary Statistics: Project Level

	No Patents			With Patents		
	Mean	Median	St.Dev.	Mean	Median	St.Dev.
Award Amount ($,000)	3432.95	959.83	14728.58	10703.81	3071.42	23041.69
Total Cost ($,000)	4092.92	1042.87	17227.72	13314.69	4009.57	26729.48
Expected Duration (days)	967.09	823.00	632.40	1333.86	1194.00	785.99
Total Duration (days)	1075.48	1003.50	633.12	1495.86	1462.00	786.95
# Patents	0.00	0.00	0.00	2.97	1.00	4.97
Cost Plus	0.94	1.00	0.23	0.96	1.00	0.20
Fully Competed	0.81	1.00	0.39	0.79	1.00	0.41
Last Week FY	0.09	0.00	0.29	0.09	0.00	0.28
Same State (Buyer/Performance)	0.17	0.00	0.38	0.16	0.00	0.37
Also a Specialist	0.81	1.00	0.39	0.81	1.00	0.39
# Bids	6.70	2.00	15.89	9.47	2.00	25.24
N	1794			176		

Notes: The level of observation is the contract. The share of contract associated with at least one patent is 9.81 percent.

an average of 22 percent of the contracting workforce. Actual retirement counts are low during the period (3 percent of the contracting workforce and 15 percent of the Retirement Eligible). This results in high non-retiree counts, 54 per office on average.[25]

Table 3: Summary Statistics: Office

	Mean	Median	S.D.
R&D Budget ($,000)	536270.00	394194.22	468614.23
# R&D Contracts	65.67	74.00	47.19
# COs	32.11	26.86	25.55
# GS-1102	292.10	111.00	321.25
Experienced GS-1102	2.33	2.08	1.50
Non-retirement	53.83	21.50	51.40
Retirement Eligible	65.17	26.00	62.55
Retirement Actual	10.50	5.00	12.67
N	30		

Notes: The level of observation is the contract office and fiscal year.

4 Results

4.1 Baseline analysis

We begin the presentation of our results with Table (4), which displays the estimates corresponding to the binary version of Equation (2), that is

$$\begin{aligned} Pr\left(Patent_{i,k,t,o} = 1 | \mathrm{Log\#CO_{k,t}}, \mathrm{X_i}, Z_{k,t}, \zeta_t\ , \zeta_o\right) = \\ \Phi\left(c + \beta \mathrm{Log\#CO_{k,t}} + \gamma \mathrm{X_i} + \delta Z_{k,t} + \zeta_t + \zeta_o\right), \end{aligned} \tag{3}$$

[25] Since workforce is planned in the previous fiscal year, the fair comparison between Retirement Eligible at t-1 is with Retirement Actual at t. Non-retirement follows this scheme and that is the reason why the average of Non-retirement in Table (3) differs from the mean difference between Retirement Eligible and Retirement Actual.

where Φ is the cumulative standard normal distribution function. Thus, we estimate the binomial response to our variable of interest β via a probit model.[26] In this specification, as compared to Equation (2), we exclude office fixed effects to avoid perfect collinearity because we observe that in our panel COs are grouped in contracting offices and never switch.

Moving from left to right expands the set of controls included. Column 1 reports the most parsimonious model specification and only includes budget and purchases as covariates controlling for office size. Holding the office's budget and purchases in a given fiscal year helps us interpret each additional CO colleague as an actual reduction in the office's total procurement workload. Column 2 contains CO and fixed effects for the fiscal year. The former is key to our identification, as discussed in Section 2.2, while the latter accounts for the government budget cycle affecting all offices simultaneously, with the resulting time-varying sources of bias. Column 3 includes controls for project value and duration to capture unobserved features of the underlying R&D activity associated with the project scale, which may predict project success and potentially correlate with the office-level workload. The construction of these variables proceeds as follows. Using the universe of R&D contracts sourced from FPDS, we evaluate within each of the R&D category cells the empirical distribution of final costs and final duration of contracts. We then assign the final cost and duration of the contract in our sample to the respective decile of the category-specific distribution. We include this classification with fixed effects for both dimensions. To control for another shared layer of unobserved characteristics, column 4 also includes fixed effects for the procurement category and procurement phase of R&D. Controlling for procurement typology is useful for controlling for time-invariant unobserved characteristics related to the probability of generating a patent. In addition, controlling for the stage of R&D activity is particularly important because contracts awarded to conduct basic research may be characterized by a higher degree of uncertainty and have a lower probability of being associated with a patent than contracts for applied research. Finally, column 5 contains the dummies *Last Week*, *Same State*, and *Specialist*, which capture possible dimensions that may be simultaneously correlated with outcome and treatment. This is our favored specification.[27] To facilitate interpretation of the estimates, we report as coefficients the average marginal effects with robust standard errors. In Appendix B, we discuss how our results are virtually unchanged when standard errors are assumed to be homoscedastic or clustered at different levels.

In line with the descriptive evidence, a naive association between workload and patents (column 1) leads to a positive and statistically significant estimate; however, the coefficient loses significance once additional controls are included. In particular, this already happens in column 2, where we add officer and fiscal year fixed effects. Finally, adding more controls increases the magnitude of the estimates but not their precision. Despite the inclusion of these controls, the problem of

[26] An alternative approach is to estimate only a linear probability model. This robustness check of the methodology is discussed in detail in the appendix. The well-known weakness of the LPM is that the fitted values need not lie in the unit interval, so predicted probabilities below zero or above one are common. We do not consider the linear probability model appropriate in our context as the underlying distribution of patents per contract is quite sparse, with many zeros for the contracts without patents, leading to a mean of 0.14 for the binary patent metric, quite far from the interval 0.4-0.6 that would accommodate well both methodologies.

[27] Please note that the working sample amounts to 1,173 observations instead of the 1,970 presented in Table (2). This sample reduction is mostly due to the CO fixed effects predicting success or failure perfectly. For the sake of comparability, all model specifications are executed on the working sample. Table (B2) in Appendix B displays the robustness of our IV probit's results to the use of a 2SLS using the full sample of contracts.

Table 4: Structural Model - Probit

	$\mathbb{1}$(# Patents > 0)				
	(1)	(2)	(3)	(4)	(5)
Log-# COs	0.16***	0.049	0.085	0.088	0.084
	(0.048)	(0.11)	(0.11)	(0.12)	(0.12)
CO FEs	No	Yes	Yes	Yes	Yes
Fiscal Year FE	No	Yes	Yes	Yes	Yes
Cost and Duration FEs	No	No	Yes	Yes	Yes
R&D Category FEs	No	No	No	Yes	Yes
R&D Stage FEs	No	No	No	Yes	Yes
Controls	No	No	No	No	Yes
N	1173	1173	1173	1173	1173

Notes: Coefficients report average marginal effects. In addition to the reported fixed effects, controls for budget and purchases in the unit-year are included in all specifications. Controls include indicators for last week of fiscal year, seller and buyer located in the same state, assistance from a contract specialist. Robust standard errors are in parentheses. *** $p < .01$.

potential downward bias in the estimates of the structural probit workload remains, as discussed in Section 2.2. To address these concerns, we implement an IV strategy based on non-retirement as the instrument.

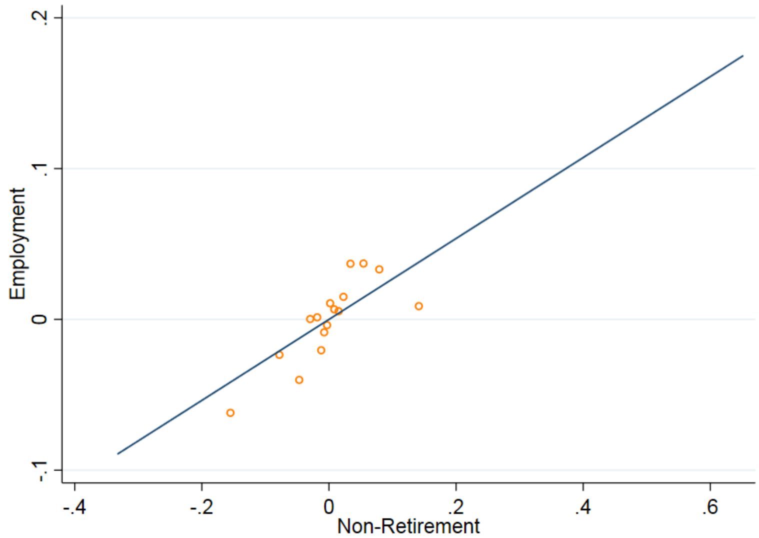
For this approach to be valid, we need to satisfy two conditions. First, the instrument must cause the variation in the treatment variable; in other words, the variation in non-retirement rates over time must have some causal power in explaining the variation in CO employment. Second, the instrument must not affect the outcome variable directly but only indirectly through the treatment variable. For this exclusion restriction to be satisfied, we need to show that the instrument affects the outcome variable only through the treatment (i.e., the non-retirement must relate to the patents because it correlates with CO employment), conditional on other possible confounding effects (i.e., the instrument must be independent of potential outcomes).

As for the strength of the instrument, Table (5) shows the first-stage results. Non-retirement is expressed in log-terms and enters with a positive and significant term. Sudden non-separations trigger a positive employment effect, as expected. In numbers, a one percentage point increase in non-retirement induces a 0.3 percentage point increase in CO employment. The elasticity of the effect is less than one because we regress CO employment on the non-retirement counts for all contracting employees—not only the non-retirement for COs.[28] Standard statistical tests of the performance of our instrument—reported at the bottom of Table (5)— reject weak and under identification, and advocate a strong first-stage relation. Figure (3) provides visual evidence of how Non-retirement affects # COs. The variables are residualized by including the variables from column 5 of Table (4) as controls and grouping them in binner scatterplots as in Cattaneo et al. (2019). More specifically, each dot represents the residualized IV's mean statistic and the residualized endogenous variable within each bin. This graphical evidence further stresses the existence of a positive effect of our IV on project innovativeness.

In terms of the exclusion restriction of our instrument, we believe that the sudden non-

[28]We emphasize again that the instrument is constructed using FedScope data and the GS-1102 classification, which do not distinguish between COs and other procurement bureaucrats. Instead, the endogenous metric for employment is constructed using FedBizOpp records that assign a responsible CO to each contract. However, we control for project-level assistance from a contract specialist in our base specification.

Figure 3: Visual Representation of the First Stage



Notes: Graphical representation of the relationships between # COs and Non-retirement. The variables are residualized, including as controls those from column 5 of Table (5). Each graph is a binned scatterplot. This means that each point represents the mean statistic of the residualized IV and the residualized endogenous variable inside each bin. The selected number of bins is 122, and it is optimal in minimizing the (asymptotic) integrated mean-squared error following Cattaneo et al. (2019).

Table 5: First Stage Regressions

	Log-# COs				
	(1)	(2)	(3)	(4)	(5)
Log-(Non-retirement)	0.151***	0.261***	0.265***	0.268***	0.268***
	(0.00748)	(0.0588)	(0.0589)	(0.0630)	(0.0626)
Weak Id.	407	20	20	18	18
Under Id.	265	32	33	32	32
CO FEs	No	Yes	Yes	Yes	Yes
Fiscal Year FE	No	Yes	Yes	Yes	Yes
Cost and Duration FEs	No	No	Yes	Yes	Yes
R&D Category FEs	No	No	No	Yes	Yes
R&D Stage FEs	No	No	No	Yes	Yes
Controls	No	No	No	No	Yes
N	1173	1173	1173	1173	1173

Notes: We report the Wald F statistic for weak identification (Kleibergen-Paap) and LM test statistic for under identification (Kleibergen-Paap). In addition to the reported fixed effects, controls for budget and purchases in the unit-year are included in all specifications. Controls include indicators for last week of fiscal year awards, seller and buyer located in the same state, assistance from a contract specialist. Robust standard errors are in parentheses. *** $p < .01$.

separation refers to an existing experienced employee who has already reached full productivity and not to a newly hired officer who has not yet reached full productivity. As a result, we expect this exogenous labor surplus as good for the R&D office outcomes, but only through a variation in workload. This is confirmed by the reduced-form relationship between patent and the instrument, as the coefficients on the instrument tend to enter with a positive and significant effect on our outcome variable (see Table 6). More specifically about the exclusion restriction, we need

Table 6: Reduced-form Regressions

	$\mathbf{1}$(# Patents > 0)				
	(1)	(2)	(3)	(4)	(5)
Log-(Non-retirement)	0.024**	0.24*	0.24**	0.28**	0.28**
	(0.011)	(0.13)	(0.12)	(0.13)	(0.13)
CO FEs	No	Yes	Yes	Yes	Yes
Fiscal Year FE	No	Yes	Yes	Yes	Yes
Cost and Duration FEs	No	No	Yes	Yes	Yes
R&D Category FEs	No	No	No	Yes	Yes
R&D Stage FEs	No	No	No	Yes	Yes
Controls	No	No	No	No	Yes
N	1173	1173	1173	1173	1173

Notes: The coefficients report average marginal effects. In addition to the reported fixed effects, controls for budget and purchases in the unit-year are included in all specifications. Controls include indicators for last week of fiscal year, seller and buyer located in the same state, assistance from a contract specialist. Robust standard errors are in parentheses. * $p < .1$, ** $p < .05$.

to consider that non-separation could determine a surplus of skills in the form of knowledge and timely managerial decisions, which can also positively impact the quality of work. Once eligible for retirement, the CO could postpone the decision to retire for some time. As long as the primary determinant of retirement now versus later is idiosyncratic and depends on personal circumstances, in addition to unobserved office- and year-level circumstances shared by the other employees, we can include the fixed effects, and the instrument will be valid.

Thus, the validity of the instrument depends on the unobserved office features being as good as random. Of course, time-varying office-level characteristics may also influence retirement decisions. If the decision to postpone retirement changes due to office tasks and characteristics changes, the exogeneity assumption would not be satisfied, and the instrument would not be valid. We use variation in the set of employees eligible for retirement and need to test its orthogonality to their workplace features. A crippling condition for us would be that the workplace characteristics affect the individual decision to postpone retirement. This would create a potential reverse causality problem, in particular when large changes in workload somehow induce people to stay at work even though eligible for retirement. To test this, we collapse the data at the office-year level and run an auxiliary regression analysis—presented in Table (7)—to detect possible observable determinants of our instruments and provide evidence for our exogeneity argument. Based on the way we construct the instrument, we find that sudden postponement of retirement is mechanically associated with contracting employment metrics (i.e., # GS-1102 and # Exp. GS-1102). There is no clear pattern of association between non-retirement counts and any of the other potential office-level predictors that we include through our data and that appear in Table (7): none reach statistical significance across model specifications. Some unobserved change in a qualitative factor of contracts may still drive retirement, undermining identification—in addition to scale variables—, but we cannot detect much from observable factors. A further, detailed discussion on the exclusion restriction of our IV is presented in Appendix A.

We can now turn to the presentation of the second-stage relationship between patent and workload. The structural relationship from the probit model depicted by Table (4) shows that the estimated effects of decreasing workload in the patentability of R&D contracts are not significant.

Table 7: Non-retirement Predictors

	Log-(Non-retirement)					
	(1)	(2)	(3)	(4)	(5)	(6)
(mean) University	-0.69	-0.72			-1.01	-0.93
	(0.64)	(1.15)			(0.58)	(0.56)
(mean) # Bids	0.00074	-0.013			-0.00062	-0.0045
	(0.0031)	(0.018)			(0.0040)	(0.0035)
(mean) Same State	-0.50	-4.33*			-0.41	0.15
	(0.60)	(2.12)			(0.55)	(0.37)
(mean) Last Week	0.47	-0.21			0.17	-0.45
	(0.51)	(3.03)			(0.52)	(0.84)
(mean) Log-Budget			-0.031	-0.15**	0.23	0.15
			(0.23)	(0.058)	(0.28)	(0.17)
(mean) Log-Purchases			-0.019	0.21**	-0.18	-0.075
			(0.074)	(0.075)	(0.13)	(0.12)
(mean) # GS-1102			0.57**	0.90***	0.50*	0.84***
			(0.20)	(0.041)	(0.26)	(0.081)
(mean) # Exp. GS-1102			-0.019	-0.021	-0.071*	-0.12**
			(0.016)	(0.036)	(0.035)	(0.051)
Constant	5.36***	6.84	1.91	1.09	-2.98	-3.66
	(0.89)	(7.30)	(4.46)	(0.84)	(5.78)	(2.82)
Office FEs	Yes	No	Yes	No	Yes	No
Fiscal Year FE	No	Yes	No	Yes	No	Yes
R-Squared	0.98	0.78	0.98	0.98	0.99	0.99
N	30.00	30.00	30.00	30.00	30.00	30.00

Notes: The table presents two sets of possible predictors of the office-year non-retirement instrument. Columns (1) and (2) include contract and contractor characteristics demeaned at the office-year level. Columns (3) and (4) include office features. Columns (5) and (6) nest the set of covariates. OLS estimates include, alternatively, office and year fixed effects. Robust standard errors are in parentheses. * $p < .1$, ** $p < .05$, *** $p < .01$

This result would suggest that additional CO colleagues in the office do not affect the innovativeness of their purchases. Offices with more COs are not more or less likely to generate patents given the same budget and purchases. However, the IV probit results shown in Table (8) suggest that the structural probit results are misleading when exogenous changes in the number of COs are taken into account. The set of controls is identical in all columns and is the same as those of Table (4). According to the baseline IV probit estimates from column 5, in our sample an additional CO in the office—corresponding to a 3 percent increase in average CO employment—leads to an increase in the probability of generating patents by about two percentage points.[29] Compared to the probit

[29] From Table (2), the baseline probability of patenting is 0.0981. The interpretation of the baseline result of our probit model is as follows: A one-unit increase in the log number of COs is associated with a 111 percentage point increase in the probability of patent equal to 1 (or equivalently, an increase in the expected probability of $0.0981 + 1.11 = 1.2081$). Such an implausible number results from the huge increase in the underlying predictor. From the perspective of the CO employment variable itself, # COs, being natural-log-transformed, is multiplied by approximately 2.718, representing an approximately 172 percent increase in average CO employment. To make

estimates from column 5 of Table (4), the magnitude of the IV probit estimates is more than one order of magnitude larger than and exceeds the 95 percent confidence interval of the structural relationship. To provide a more transparent economic interpretation of the estimates, we can consider what would happen if we used them to infer the effect of raising the workload of all offices to the level of the office with the largest backlog in our sample. To do so, we first collapse our information at the office-year level. Then, we regress employment in contracting offices on budget and purchases. Finally, we rank the office year in terms of residualized employment (interpretable as backlog). In our data, the office-year pair with the highest workload is the purchasing unit Rome Laboratory in 2012. If we bring all office-year pairs up to its level, this implies a reduction in the number of patents by 13 percentage points on average per contract (i.e., about 50 percent less likely), or about 33 total across all contracts in the dataset on a yearly basis).

Table 8: Second Stage - IV Probit

	$\mathbb{1}$(# Patents > 0)				
	(1)	(2)	(3)	(4)	(5)
Log-# COs	0.14**	0.96*	0.94**	1.09**	1.11**
	(0.072)	(0.50)	(0.47)	(0.48)	(0.49)
CO FEs	No	Yes	Yes	Yes	Yes
Fiscal Year FE	No	Yes	Yes	Yes	Yes
Cost and Duration FEs	No	No	Yes	Yes	Yes
R&D Category FEs	No	No	No	Yes	Yes
R&D Stage FEs	No	No	No	Yes	Yes
Controls	No	No	No	No	Yes
N	1173	1173	1173	1173	1173

Notes: Coefficients report average marginal effects. In addition to the reported fixed effects, controls for budget and purchases in the unit-year are included in all specifications. Controls include indicators for last week of fiscal year, seller and buyer located in the same state, assistance from a contract specialist. Robust standard errors are in parentheses. * $p < .1$, ** $p < .05$.

4.2 Robustness analysis

We conclude this section with a brief summary of the findings from the main robustness checks. To simplify the presentation, we present the findings by categorizing them into three groups. First, to address potential concerns about the definition of the workload index, we execute the regression shown in equation (3) with alternative endogenous employment measures sourced from the FedScope's workforce data. Second, to assess the soundness of the identification strategy, we explore alternative definitions of the instrument by again using the FedScope dataset. Finally, we use additional variables from the FPDS to test for possible omitted variables that could bias our results. In essence, these additional results further exploit the richness of the dataset. While the overall qualitative results prove robust, these additional findings play an important role in strengthening the quality and depth of the analysis. We refer the reader to Appendix B for

the interpretation more tractable and realistic, we may want to see the impact on project innovation of adding one additional CO. To do this, we need only divide the marginal effect of 111 percentage points by one log-units of CO employment. Since the average number of CO -employed in our sample before the log transformation is 32.11, the marginal increase by one log unit of employment amounts to 32.11 * (2.718 - 1), which is an absolute increase of about 55 employees. Proportionally, one additional CO in the office—implying an average increase of 1/32.11 = 3.11 percent of # COs—corresponds to an increase in the probability of patenting of 1.11/55.16 = 2 percentage points.

additional robustness checks.

The combination of the FPDS and FedBizOpps allows us to identify the COs in charge of the procurement process and count the distinct officers active in a given contracting unit-year. We want to test the robustness of our results against alternative and less conservative specifications of CO employment. By relying on available information from FedScope, we are able to provide an alternative and less conservative count of CO employment. As stressed in Section 3.1, the GS-1102 count (i.e., contracting employment) includes all COs and other contracting employees involved in the procurement process at different levels of the hierarchy and with different tasks. In Table (9), we show how the estimates change relative to our baseline from Table (8)—reported in column 1—when we change only the endogenous variable. The coefficients from column 2 suggest that replacing our baseline index of CO employment with the total count of GS-1102 has no statistical difference in terms of its effect on the outcome. To capture the effect heterogeneity that arises from GS-1102 having managerial responsibilities, we condition GS-1102 in column 3 on having at least pay grade 14.[30] Although qualitatively the same, the magnitude of the effect is one-third of the baseline but more precisely estimated.

Our instrument exploits the unexpected gaps between actual and expected retirement. In the baseline analysis, we construct this variable as the log difference between the two. As we rely on the same underlying variables, we want to test the qualitative stability of our results when we use alternative specifications of IV with a similar interpretation. In Table (10), we benchmark the results in column 1, where we report the baseline. In column 2, we report the ratio of Retirement Eligible to Retirement Actual. In column 3, we use the logarithm of this ratio. The second stage results are stable and statistically indistinguishable across all alternative linear or log-linear specifications of the non-retirement counts. Finally, column 4 uses log counts of total retirements as in Warren (2014) as an alternative instrument for the workload. Again, the results are qualitatively and quantitatively stable.

The results of Warren (2014) suggest that the decision to leave contracts less complete may also affect other procurement terms, in particular the extent to which a project is competed at the bidding phase and the pricing structure of the contract offered. Less complete contracts benefit less from the competition, so busier COs use less competitive mechanisms. Cost-plus contracts facilitate the management of contract renegotiations and are therefore preferred by COs in the current and foreseeable backlog. The author shows that an increased workload for COs due to backlog spikes leads to fewer complete contracts and, consequently, higher use of noncompetitive and cost-plus agreements. Following these arguments, we believe that other dimensions of the design process observed via FPDS may be affected by workload. Although in the R&D realm contracts are highly incomplete by design, some variation at the intensive margin could still be captured by contract pricing (i.e., cost-plus vs. fixed price) and the choice of the officer to make the notice full and open to competition or to exclude some sources. Another decision that the CO makes in the solicitation is the bidding process. In standard procurement processes, the bidding process usually boils down to a choice between a sealed low-bid auction and a negotiated proposal format. According to FAR, bidding procedures in R&D procurements can vary, and the path

[30] GS-14 is the 14th pay grade on the General Schedule, the salary scale used to set salaries for most government employees. Pay grade 14 is reserved for top positions such as supervisors, high-level technical specialists, and top advanced degree holders.

chosen depends heavily on the nature of the research features being procured. Whether pricing, competition, and tendering procedures also affect our outcome variable is again an open question for which there is neither empirical evidence nor theoretical modeling. However, we believe it is relevant to test the robustness of our results against the inclusion of these problematic controls that capture contract completeness decided by CO and could bias our baseline results. In column 2 of Table (11), we add two binary variables to the baseline model specification—in column 1—indicating the cost-plus nature of the contract (as opposed to a fixed price) and open (vs. restricted) competition. The coefficient is insignificant, while the coefficient on our main variable is only marginally affected. As we do not know the difference between solicitation procedures in the R&D context, we take a data-driven approach in column 3 and include a set of fixed effects for the different categories of procedures in the baseline model specification. Specifically, 81 percent is basic research (FAR 6.102), 11 percent is negotiated proposal/quotes, whereas the remainder is split between sealed bids, single sourcing (FAR 13.106), and multiple award fair opportunities (FAR 16.505). Again, the main coefficient of interest is positive and significant, with an indistinguishable magnitude compared to the baseline analysis for the same sample. Finally, column 4 includes both dummy variables and fixed effects with similar results. The results prove to be very robust to the inclusion of decision variables that are up to the discretion of the CO, in particular in the R&D contracting realm.

Table 9: Alternative Specifications of Endogenous Employment

	$\mathbf{1}$(# Patents > 0)		
	(1)	(2)	(3)
Log-# COs	1.11**		
	(0.49)		
Log-# GS-1102		0.94**	
		(0.45)	
Log-# Exp. GS-1102			0.31***
			(0.10)
N	1173	1173	1173

Notes: Baseline results—column 5 of Table (8), reported in column 1—are replicated with alternative measures of contracting employment. # COs is replaced as endogenous variable by # GS-1102 and # Experienced GS-1102 in columns 2 and 3, respectively. Robust standard errors in parenthesis. **p < .05, ***p < .01

5 Workload and Contract Specifications

The results of the analyses presented in the previous sections clearly show that the workload of the officer in charge of the award of an R&D procurement process harms the contract performance as measured by patents. However, the data at hand do not allow us to unambiguously identify the channel through which the additional workload worsens contracting output. As discussed above, Warren (2014) analyzes the impact of workload in the context of regular procurement and finds that workload negatively affects contract completeness and, as a consequence, the contract performance. Warren (2014) measures contract completeness mainly by looking at the type of contract pricing

Table 10: Alternative Instruments

	$\mathbf{1}$(# Patents > 0)			
	(1)	(2)	(3)	(4)
Log-# COs	1.11**	1.74**	1.76***	0.75**
	(0.49)	(0.70)	(0.60)	(0.30)
N	1173	884	884	1173

Notes: Baseline results—column 5 of Table (8), reported in column 1—are replicated with alternative IVs: Retirement Eligible $t-1$ / Retirement Actual t in Column 2; log(Retirement Eligible $t-1$ / Retirement Actual t) in Column 3; Log(Retirement Actual t) in Column 4. Coefficients report average marginal effects. Robust standard errors are in parenthesis. **p < .05, ***p < .01

Table 11: Inclusion of Endogenous Omitted Controls

	$\mathbf{1}$(# Patents > 0)			
	(1)	(2)	(3)	(4)
Log-# COs	1.11**	0.96**	0.98**	0.88**
	(0.49)	(0.45)	(0.46)	(0.41)
Cost Plus Pricing		-0.025		-0.024
		(0.047)		(0.047)
Open Competition		0.043		0.050
		(0.042)		(0.042)
Solicitation Proc. FEs	No	No	Yes	Yes
N	1173	1126	1172	1125

Notes: Baseline results—column 5 of Table (8), reported in column 1—are replicated with the inclusion of additional covariates: Cost Plus and Open Competition dummies in column 2; Solicitation Procedures fixed effects in column 3; Cost Plus and Open Competition dummies plus Solicitation Procedures fixed effects in column 4. Coefficients report average marginal effects. Robust standard errors in parenthesis. **p < .05.

selected by the contracting officers, i.e., a cost-reimbursement vs. fixed-price contract. Although we can assume contract specification to be a relevant channel in the context of R&D contracting, we cannot test this hypothesis using variables such as contract pricing. As described in Section 1, R&D contracts are by definition focused on goals for which work or methods cannot be precisely described in advance, and FAR openly recommends using cost-reimbursement contracts, i.e., cost-plus contract instead of fixed-price contracts. Our data confirms COs generally comply with this recommendation as 95 percent of the contracts in our sample are awarded using a cost-plus pricing scheme.

Nevertheless, we are inclined to believe that the impact of workload on contract performance passes through tender and contract specifications. A CO's main problem in the award of an R&D contract is describing the work requirement in a way that is clear to the prospective contractors, even in situations in which the CO does not have a complete understanding of the work in advance. The CO is responsible for translating a rather abstract idea into a language that is contractually clear to the prospective contractors (US Air Force 1967). To succeed in such a complex task, the CO generally closely interacts with the internal scientific staff responsible for the issuance of the tender and other technical associates and must develop an ad-hoc approach for each new procurement activity. Clearly, a higher workload for the CO may harm her capacity to maintain adequate levels of interaction with the technical staff and especially her ability to translate tender-specific but uncertain technical objectives into contractually binding behavior for prospective contractors.

Although we can neither observe nor proxy the actual effort the CO put in for specific procurement activities and, therefore, we cannot determine if the increases in workload lead to a reduction in the effort provided for a single contract, the heterogeneity of the R&D activities in our sample allows us to provide indirect evidence of the importance of the COs in drafting solicitations and contractual agreements nonetheless. The contracts in our sample are awarded for the performance of work at different stages of the R&D process. Table (2) reports that 25 percent of the contract was awarded for the performance of basic research, 52 percent for the performance of applied research, and the remaining 23 percent for development research.[31] The importance of the CO in the tender specification is likely to change with the stage of the R&D work procured. In the procurement of basic research whose direct applicability is still uncertain, agencies often rely on a Broad Agency Announcement (BAA).[32] BAAs are broad in their subject matter and may be used by agencies to fulfill their requirements for scientific study and experimentation directed toward advancing the state-of-the-art or increasing knowledge or understanding rather than focusing on a specific system or hardware solution. Generally, a BAA describes the agency's research interest, either for an individual program requirement or for broadly defined areas of interest, but does not include a clear and complete work statement concerning the area of exploration and the end objectives of a contract. In this context, the importance of the CO in drafting a request for proposal and a subsequent contract that would provide the right incentive to the prospective contractor appears to be rather limited. However, contracts for the procurement of R&D work that happens at more advanced stages of development are relatively definable. In such cases, it is rather straightforward for the CO to draft a request for proposal and the resulting contract in a clear and complete fashion.

The relevance of the CO's ability in translating an abstract idea in a contractually clear work statement reaches its maximum for contracts awarded for intermediate stages of development, especially for the applied research and exploratory development phase. Hence, if the negative effect of increases in workload on contract performance is connected to a reduction in the effort of the CO in the key moments of solicitation and contract drafting, we should expect our results to be stronger for contracts awarded for applied research than for the other stages. To test this hypothesis, we split our sample into three groups based on the R&D stages, Basic Research, Applied Research, and Advanced Development, and ran the same model as in the focal analysis. Table (12) reports the results of the split-sample analysis. As the table shows, the overall negative baseline effect seems to be largely driven by contracts awarded for applied research. Although we are aware this result does not provide conclusive evidence, it strongly hints at the proposed idea that backlogged COs are indeed unable to devote enough time to tender and contract specification, resulting in a reduction in the performance of the contracts they award.

[31] Due to the high comparability in their nature, we group the subsequent development stages (i.e., stages 4 to 6) together with stage 3 within the broad development stage umbrella.

[32] See FAR 35.016.

Table 12: Sensitivity Analysis with R&D Stages

	1(# Patents > 0)		
	(1)	(2)	(3)
Log-# COs	-1.03	2.43***	1.60
	(4.30)	(0.84)	(2.18)
CO FEs	Yes	Yes	Yes
Fiscal Year FE	Yes	Yes	Yes
Cost and Duration FEs	Yes	Yes	Yes
R&D Category FEs	Yes	Yes	Yes
R&D Stage FEs	No	No	Yes
Controls	Yes	Yes	Yes
N	482	1014	452

Notes: Baseline analysis is replicated in the subsample of basic research (i.e., stage 1), applied research (i.e., stage 2), development (i.e., stage 3-6) contracts in columns 1, 2, and 3, respectively. R&D stages are omitted in specifications 1 and 2 due to collinearity. The coefficients are estimated via 2SLS due to insufficient power. Robust standard errors are in parenthesis. ***p < .01.

6 Conclusions

As modern economies have evolved from industrial- to knowledge-based, the activity of public administrations has also changed over time, and the governance of technology has become a crucial part of their activity. In public procurement, these structural changes are compelling governments to procure increasingly complex and costly products and services. In the US, this increasing complexity is going hand in hand with an increasing number and value of procurements. In response, the government has focused on streamlining acquisition rules and giving more discretion to front-line officials (Carril 2021; Calvo et al. 2019; Giuffrida and Rovigatti 2018). However, human resources are lagging behind due to the inability to retain and recruit talent at all levels of the administration. The result is a backlogged contracting workforce, which generates a major capacity constraint for effective public spending in general, and investment in R&D in particular.

To provide the first quantification of the latter bottleneck, this paper sheds light on the microeconomic implications for innovative activities due to the workload of public buyers of R&D. Our research is in the spirit of Furman and Stern 2011's as we are interested in understanding how institutions shape new knowledge and what frictions they encounter, and complements Warren 2014's findings on how a backlogged contracting personnel affects standard procurements' performance outcomes. We combine several different data sources to link tender, contract, patent, and office records to the identity of the CO. We focus on contracts awarded by the contracting offices that compose the AFRL to effectively count the number of officials actively involved in procuring R&D projects in a given fiscal year: we use this measure as an inverse indicator of the office's workload, after controlling for the budget and purchases. To overcome endogeneity in the R&D contract outcomes and buyer's workload relationship, we implement an IV strategy combined with CO fixed effects to identify the effect of the latter on the former. The identification comes from unanticipated retirement shifts among COs, which we use as an instrument for the workload. Our results are robust to several modifications and stress that a large increase in patenting at the extensive margin occurs when the same officer is exposed to a declining workload.

References

AAP (2007): "Report of the Acquisition Advisory Panel to the Office of Federal Procurement Policy and the United States Congress," *Executive Office of the President.*

Acemoglu, D. (2005): "Politics and economics in weak and strong states," *Journal of Monetary Economics*, 52, 1199–1226.

Asch, B., S. J. Haider, and J. Zissimopoulos (2005): "Financial Incentives and Retirement: Evidence from Federal Civil Service Workers," *Journal of Public Economics*, 89, 427–440.

Bajari, P. and S. Tadelis (2001): "Incentives versus Transaction Costs: A Theory of Procurement Contracts," *RAND Journal of Economics*, 32, 387–407.

Baltrunaite, A., C. Giorgiantonio, S. Mocetti, and T. Orlando (2021): "Discretion and Supplier Selection in Public Procurement," *The Journal of Law, Economics, and Organization*, 37, 134–166.

Bandiera, O., A. Prat, and T. Valletti (2009): "Active and Passive Waste in Government Spending: Evidence from a Policy Experiment," *The American Economic Review*, 99, 1278–1308.

Best, M. C., J. Hjort, and D. Szakonyi (2017): "Individuals and Organizations as Sources of State Effectiveness," NBER Working Paper 23350, National Bureau of Economic Research, Cambridge, MA.

Bhattacharya, V. (2018): "An Empirical Model of R&D Procurement Contests: An Analysis of the DOD SBIR Program," Working Paper, Northwestern University, Evanston.

Bruce, J. R. and J. M. de Figueiredo (2020): "Innovation in the U.S. Government," Working Paper 27181, National Bureau of Economic Research.

Bruce, J. R., J. M. de Figueiredo, and B. S. Silverman (2019): "Public Contracting for Private Innovation: Government Capabilities, Decision Rights, and Performance Outcomes," *Strategic Management Journal*, 40, 533–555.

Bucciol, A., R. Camboni, and P. Valbonesi (2020): "Purchasing Medical Devices: The Role of Buyer Competence and Discretion," *Journal of Health Economics*, 74.

Calvo, E., R. Cui, and J. C. Serpa (2019): "Oversight and efficiency in public projects: A regression discontinuity analysis," *Management Science*, 65, 5651–5675.

Carril, R. (2021): "Rules versus Discretion in Public Procurement," Economics Working Papers 1765, Department of Economics and Business, Universitat Pompeu Fabra.

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.

Cattaneo, M. D., R. K. Crump, M. H. Farrell, and Y. Feng (2019): "On binscatter," *arXiv preprint arXiv:1902.09608.*

de Rassenfosse, G., A. Jaffe, and E. Raiteri (2019): "The procurement of innovation by the US government," *PloS one*, 14, e0218927.

de Rassenfosse, G., G. Pellegrino, and E. Raiteri (2020): "Do Patents Enable Disclosure? Evidence from the Invention Secrecy Act," Working Papers 9, Chair of Innovation and IP Policy.

Decarolis, F. (2014): "Awarding Price, Contract Performance and Bids Screening: Evidence from Procurement Auctions," *American Economic Journal: Applied Economics*, 6, 108–132.

——— (2018): "Comparing Public Procurement Auctions," *International Economic Review*, 59, 391–419.

Decarolis, F., G. de Rassenfosse, L. Giuffrida, E. Iossa, V. Mollisi, E. Raiteri, and G. Spagnolo (2021): "Buyers' Role in Innovation Procurement: Evidence from U.S. Military R&D Contracts," *The Journal of Economics & Management Strategy*, Forthcoming.

Decarolis, F., L. M. Giuffrida, E. Iossa, V. Mollisi, and G. Spagnolo (2020): "Bureaucratic Competence and Procurement Outcomes," *The Journal of Law, Economics, and Organization*, 36, 537–597.

Edquist, C. (2015): "Innovation-related Public Procurement as a Demand-oriented Innovation Policy Instrument," Papers in Innovation Studies 2015/28, Lund University, CIRCLE - Center for Innovation, Research and Competences in the Learning Economy, Lund.

Fleming, L., H. Greene, G. Li, M. Marx, and D. Yao (2019): "Government-funded research increasingly fuels innovation," *Science*, 364, 1139–1141.

Furman, J. L. and S. Stern (2011): "Climbing atop the Shoulders of Giants: The Impact of Institutions on Cumulative Research," *American Economic Review*, 101, 1933–63.

GAO (2017): "High-Risk series 2017," Tech. rep., GAO-17-317, US Government Accountability Office.

——— (2019): "DOD Increased Use of Human Capital Flexibilities but Could Improve Monitoring," Tech. rep., GAO-19-509, US Government Accountability Office.

Gerardino, M. P., S. Litschig, and D. Pomeranz (2017): "Can Audits Backfire? Evidence from Public Procurement in Chile," Working Paper 23978, National Bureau of Economic Research.

Giuffrida, L. M. and G. Rovigatti (2018): "Can the Private Sector Ensure the Public Interest? Evidence from Federal Procurement," ZEW Discussion Paper No. 18-045, ZEW - Centre for European Economic Research, Mannheim.

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

Kang, K. and R. A. Miller (2017): "Winning by Default: Why is There So Little Competition in Government Procurement?" Working paper, Carnegie Mellon University, Pittsburgh.

Lewis, G. B. and D. Pitts (2018): "Deciding to Retire from the Federal Service," *Review of Public Personnel Administration*, 38, 49–82.

Liebman, J. B. and N. Mahoney (2017): "Do Expiring Budgets Lead to Wasteful Year-End Spending? Evidence from Federal Procurement," *American Economic Review*, 107, 3510–49.

Mait, J. N. (2005): "Making IT Happen: Transforming Military Information Technology," Tech. rep., National Defense University Washington DC Center for Technology and National.

Moretti, E., C. Steinwender, and J. Van Reenen (2019): "The Intellectual Spoils of War? Defense R&D, Productivity and International Spillovers," Working Paper 26483, National Bureau of Economic Research.

Peña, V., S. V. Howieson, B. Lal, J. R. Behrens, B. L. Zuckerman, M. V. Merrill, and J. L. Zhu (2017): *Early Stage Research and Technology at US Federal Government Agencies*, Institute for Defense Analyses.

Raiteri, E. (2018): "A Time to Nourish? Evaluating the Impact of Public Procurement on Technological Generality through Patent Data," *Research Policy*, 47, 936–952.

Rau, C. A. and P. J. Stambersky (2009): "Management and oversight of services acquisition within the United States Army," Tech. rep., Naval Postgraduate School, Monterey, CA.

Rendon, R. G., U. M. Apte, and A. Apte (2012): "Services acquisition in the DoD: A comparison of management practices in the Army, Navy, and Air Force," Tech. rep., Defence Acquisition University.

Simcoe, T. and M. W. Toffel (2014): "Government green procurement spillovers: Evidence from municipal building policies in California," *Journal of Environmental Economics and Management*, 68, 411–434.

US Air Force (1967): "Air Force Research and Development Contracting Officers' Handbook," AFSC Pamphlet 70-2, Air Force System Command.

Warren, P. L. (2014): "Contracting Officer Workload, Incomplete Contracting, and Contractual Terms," *The RAND Journal of Economics*, 45, 395–421.

Weber, M. (1921): *Economy and Society: An Outline of Interpretive Sociology (1921; German: Wirtschaft und Gesellschaft. Grundriß der verstehenden Soziologie)*, vol. 1, University of California Press (1978).

A Appendix: Discussion on the Exclusion Restriction

We further explore the soundness of the exclusion restriction by presenting a suggestive complementary exercise and running IV regressions with an alternative definition of the endogenous variable. For our instrument to be valid, the effect of unexpected variation in employment must be mediated by a shift in workload; in our baseline setting, we (inversely) proxy workload with CO employment and use office scale via purchases and budget as included instruments. If we fix the latter two dimensions, we interpret an additional CO colleague as a reduction in the backlog of work. However, this empirical design can be replicated by replacing each time CO employment with one of two measures of office size as the endogenous workload index and controlling for the remaining two variables as measures of scale. If the IV tests and results hold in the other two cases, the exclusion restriction is not satisfied by construction, and other variables of scale size can substitute for the workload. This auxiliary exercise is shown in Table (A1) for the working sample. In columns 1 and 2, the endogenous workload variable is the number of awards and annual procurement spending, respectively, while CO employment is as a covariate. In the first case, we interpret an additional dollar spent, holding CO employment and purchases constant, as a proxy for workload; in the second case, we interpret an additional contract, holding the associated budget and CO employment constant, as a proxy for the workload. The instrumented coefficients are in bold. In both cases, tests for instrument relevance indicate the non-validity of these alternative IV analyses. The results confirm our idea that non-retirement affects the innovativeness of awards via workload, but only via the employment dimension of the latter.

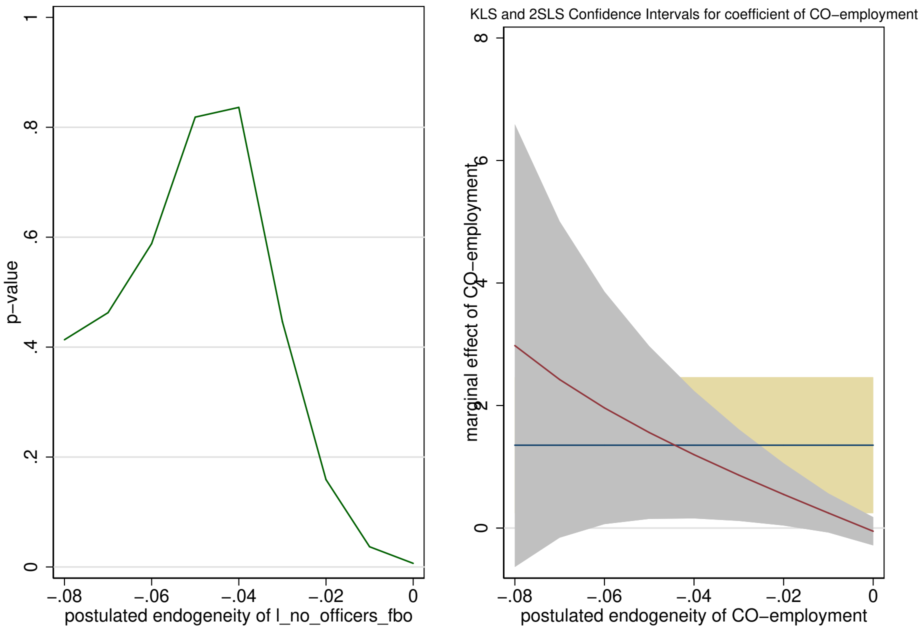
Table A1: Second Stage with Alternative Endogenous Workload Definition

	1(# Patents > 0)	
	(1)	(2)
Office R&D Budget	**-56.2*****	0.62
	(0.19)	(0.55)
# Office R&D Awards	-1.91***	**3.99**
	(0.73)	**(2.65)**
Log-# COs	16.0***	-2.86
	(1.06)	(1.87)
Weak Id.	0	3
Under Id.	0	3
CO FEs	Yes	Yes
Fiscal Year FE	Yes	Yes
Cost and Duration FEs	Yes	Yes
R&D Category FEs	Yes	Yes
R&D Stage FEs	Yes	Yes
Controls	Yes	Yes
N	1173	1173

Notes: We replicate column 5 of Table (8) by replacing the instrumented variable, which is in bold. We report the Wald F statistic for weak identification (Kleibergen-Paap) and LM test statistic for under identification (Kleibergen-Paap). Control for budget and purchases in the unit-year always included. Robust standard errors in parenthesis. *** $p < .01$.

To corroborate further our IV strategy, we propose a state-of-the-art exercise hinging on the

Figure A1: Exclusion Restriction Test as in Kiviet (2020)



Notes: Values of single just-identifying exclusion restriction tests and inference of 2SLS and KLS based on non-orthogonality conditions.

recent contribution from Kiviet (2020) to seek to actually *test* the exclusion restriction. Kiviet (2020) presents an approach by which, without exploiting any external instruments, general linear coefficient restrictions can be tested in a multiple regression model with an arbitrary number of endogenous regressors. The strategy requires a flexible assumption on the degree of endogeneity of all regressors. This approach allows generating statistical evidence on the tenability of exclusion restrictions. When this yields an acceptable just-identifying or over-identifying set of instruments, it provides the essential underlying building block for a standard or a series of incremental Sargan–Hansen tests. We follow the Kiviet (2020) approach in our just-identified IV analysis to produce further insights into the tenability of our exclusion restriction hypothesis. The left graph of Figure (A1) shows different values of downward bias in Equation (3) (i.e., negative values of postulated endogeneity), the p-values of the single just-identifying exclusion restriction tests for CO employment. The instrument's validity seems quite likely when it is close to -0.04, and it holds over most of the negative space. In the right graph, Figure (A1) shows the 2SLS asymptotic 95 percent confidence interval (the yellow area) for β, which is invariant regarding the endogeneity and centered at the 2SLS estimate 1.14 (solid blue line). It also shows the KLS estimator (the solid red line), which varies with the postulated endogeneity, and the KLS asymptotic 95 percent confidence interval (the grey area). The graph also shows that the 95 percent 2SLS confidence interval, which is contingent on the validity of the instruments, conforms in width to a conservative KLS-based interval contingent on the supposition and exogeneity of the instrument. This evidence suggests that, with the likely size and direction of endogeneity, the 2SLS and KLS inference are similar, and the exclusion restriction underlying our IV strategy's validity is satisfied.

B Appendix: Additional Robustness Analysis

This appendix reports additional robustness checks. For convenience, these results are subdivided into two groups depending on whether the robustness analysis involves 1) the specification of the standard errors; 2) the estimation method.

1. Robustness to the definition of the dependent variable. Baseline results—column 6 of Table (9), reported in column 1—are replicated with different specifications of the standard errors: homoscedastic standard errors in column 2; clusterization at the contracting office level in column 3; clusterization at the R&D category level in column 4; clusterization at the level of the state of performance in column 5.

2. Robustness to the estimation method. The baseline model—column 5 of Table (8), reported in column 1—is replicated with different estimators using the sample from column 5 of Table (8). Columns 2 and 3 report results of a 2SLS estimation using the working sample and the full sample, respectively; column 4 retains perfect predictor variables in the maximization process of the IV probit. This option is typically not used and may introduce numerical instability. Normally, IV probit drops any endogenous or exogenous variables that perfectly predict success or failure in the dependent variable. The associated observations are also dropped. Results are robust to the employment of these alternative estimation methods.

Table B1: Robustness to the Standard Errors Definition

	$\mathbf{1}$(# Patents > 0)				
	(1)	(2)	(3)	(4)	(5)
Log-# Officers FedBizOpp	1.11**	1.11**	1.11	1.11***	1.14**
	(0.49)	(0.50)	(0.73)	(0.21)	(0.58)
Observations	1173	1173	1173	1173	1170

Notes: Baseline results—column 5 of Table (8), reported in column 1—are replicated with different specifications of the standard errors: homoscedastic standard errors in column 2; clusterization at the office level in column 3; clusterization at the R&D category level in column 4; clusterization at the level of the state of performance in column 5. ** $p < .05$, *** $p < .01$.

Table B2: Robustness to the Estimation Method

	$\mathbf{1}$(# Patents > 0)			
	(1)	(2)	(3)	(4)
Log-# Officers FedBizzOpp	1.11**	1.35**	0.93**	1.11**
	(0.49)	(0.62)	(0.46)	(0.49)
N	1173	1173	1948	1173

Notes: The baseline model—column 5 of Table (8), reported in column 1—is replicated with estimators other than the IV probit. Columns 2 and 3 report results of a 2SLS estimation using the working sample and the full sample, respectively; column 4 retains perfect predictor variables in the maximization process of the IV probit. ** $p < .05$.

Download ZEW Discussion Papers from our ftp server:

http://ftp.zew.de/pub/zew-docs/dp/

or see:

https://www.ssrn.com/link/ZEW-Ctr-Euro-Econ-Research.html
https://ideas.repec.org/s/zbw/zewdip.html

//
IMPRINT

ZEW – Leibniz-Zentrum für Europäische Wirtschaftsforschung GmbH Mannheim
ZEW – Leibniz Centre for European Economic Research

L 7,1 · 68161 Mannheim · Germany
Phone +49 621 1235-01
info@zew.de · zew.de