BUREAUCRATIC COMPETENCE AND PROCUREMENT OUTCOMES

Francesco Decarolis
Leonardo M. Giuffrida
Elisabetta Iossa
Vincenzo Mollisi
Giancarlo Spagnolo

Bureaucratic Competence and Procurement Outcomes
Francesco Decarolis, Leonardo M. Giuffrida, Elisabetta Iossa, Vincenzo Mollisi, and Giancarlo
Spagnolo
NBER Working Paper No. 24201
January 2018, Revised April 2018
JEL No. H11,H57,J45

ABSTRACT

Does a more competent public bureaucracy contribute to better economic outcomes? We address this question in the context of the US federal procurement of services and works by combining contract-level data on procurement performance and bureau-level data on competence and workforce characteristics. Using an instrumental variable strategy, we find that an increase in bureau competence causes a significant and economically important reduction in: i) delays, ii) cost overruns, and iii) number of renegotiations. Cooperation within the office appears to be a key driver of the findings.

Francesco Decarolis
Bocconi University
Via Sarfatti 25
Milan, 20100
Italy
and EIEF
francesco.decarolis@unibocconi.it

Leonardo M. Giuffrida
University of Rome Tor Vergata
Via Columbia, 2
Rome, 00100
Italy
leonardo.giuffrida@uniroma2.it

Elisabetta Iossa
Department of Economics and Finance
University of Rome Tor Vergata
Via Columbia 2
Rome, 00133
Italy
and IEFE-Bocconi and CEPR
elisabetta.iossa@uniroma2.it

Vincenzo Mollisi
Free University of Bozen-Bolzano
Piazza dell'Universita 1
Bozen-Bolzano 39100
Italy
vincenzo.mollisi@unibz.it

Giancarlo Spagnolo
SITE-Stockholm School of Economics
Sveavägen 65
Stockholm, SE-113 83
Sweden
and EIEF, University of Tor Vergata, and CEPR
spagnologianca@gmail.com

I Introduction

An inefficient bureaucracy can represent a major obstacle to economic activities. In a path-breaking study, De Soto [1990] documented how excessive government requirements for starting a business can dramatically slow down the entry of new enterprises. Djankov et al. [2002] notoriously expanded this work by measuring, for 85 countries, the number of procedures, the official time and the official cost that a start-up must bear before it can operate legally. These works laid the ground for the World Bank's Doing Business project which provides objective measures of business regulations and their enforcement across 190 economies and it is widely recognized as a fundamental competitiveness' indicator.

An inefficient bureaucracy can also affect the ability of private firms to do business with governments. The World Bank has recently began to release its Benchmarking Public Procurement, which examines public procurement laws and regulations across 180 economies. The report reveals the existence of considerable heterogeneity across countries, which may be associated with significant waste. Concerns on the competency of public buyers are also being expressed. Reporting data from the French Public Procurement Grouping Union, Saussier and Tirole [2015] lamented that 61 percent of French public buyers joined a purchasing department following a period of internal mobility with no prior experience in the field. Only 39 percent of public buyers undertook specific training, and less than a third felt familiar with the economic and industrial fabric. The resulting inefficiency may be significant, as procurement involves complex activities, especially for works and services. Public buyers must identify their needs, design and manage the award mechanism, balance risks and incentives in the contract, and monitor the contractor in the execution phase. These tasks require knowledge of product and market characteristics and of legal rules, as well as good management practices to time activities and coordinate resources, and strategic abilities to design the tender and negotiate the contract. With about 15 percent of world GDP spent every year on public procurement, reducing inefficiencies in public procurement might yield significant costs saving and better public services.

In this paper, we provide the first comprehensive quantification of the impact of bureau-

cratic competence on the outcome of US public contracts for works and services. Throughout the paper, we use the term "competence" to capture all those factors, from the availability of appropriate skills to good management practices, which affect the capacity of procurement offices to effectively perform their mission. Combining three large datasets on federal bureaus purchases, their internal functioning and workforce characteristics, we quantify the effects of bureaus' competence on the time and cost performance of public contracts and on the number of times they are renegotiated. Our identification strategy exploits the exogeneity of death events involving public officials to allow for a causal interpretation of bureau competence on procurement performance.

There are three main measurement challenges that our analysis seeks to overcome. The first regards the choice of the performance measure. Whilst unit price comparisons can be used for standardized goods, they are not suitable for works or services, where prices can be renegotiated and contract execution can be delayed. We therefore construct two proxies of performance based on cost overruns and time delays, using the Federal Procurement Data System (FPDS), a system tracking nearly every awarded federal contract, as well as every follow-on action. We take into account that both cost overruns and delays may be due to new or additional work requested by the public buyer, in which case they should not be viewed as indicative of a poor outcome. We therefore disregard these type of delays and cost overruns and consider only those which have occurred to deliver the work or service that was originally tendered. These performance measures are regularly used by governments in their procurement reports as well as by researchers in the field.[1] In addition, we use the number of renegotiation episodes to capture Williamson [1971]'s transaction or "haggling" costs, that are present whatever the reason behind the renegotiation and that recent research have shown to be sizable [Bajari, Houghton and Tadelis, 2014].

The other two measurement challenges regard bureaucratic competence. Translating the complex and multifaceted concept of competence into a variable entails some choices. We build measures of federal bureaus and agencies competence, by relying on a major survey, the Federal Employee Viewpoint Survey (FEVS), which has been administered for more than

---

[1]See, for instance, Bajari and Lewis [2011] for delays, and Mohamed, Khoury and Hafez [2011], Iimi [2013], and Bajari, Houghton and Tadelis [2014] for cost overruns.

ten years with the same questions to nearly all government agencies, drawing responses from about one fourth of all federal employees every year. While the source of data is extremely rich and the generality of the survey question we use to measure competence ("How would you rate the overall quality of work done by your work unit?") helps us to capture the broad nature of this concept, it follows that our variable is only a proxy for the underlying measure. In turn, this measurement error implies a likely downward bias in any association between this measure and procurement outcomes.

The third measurement problem is the association between more complex contracts and more competent buyers: a buyer may consistently show a poor performance simply because it has to deal with complex contracts. Thus, despite the richness of our data to control for contract complexity, since more complex contracts are intrinsically more likely to produce renegotiations, an omitted variable problem is thus likely to bias downward our estimates of the effects of competence. This point is well illustrated by a case we will discuss below: the performance of the agencies that are worst in terms of competence (the Department of Veterans Affairs and the Department of Justice) is superior to that of the two most competent agencies (the NASA and the Nuclear Regulatory Commission) in terms of both delays and cost overruns. This striking inversion of the relative ranking is a key feature of the economic environment that we analyze and it implies that any straightforward regression of performance on competence would grossly underestimate the impact of competence.

To handle both sources of bias, we develop an instrumental variable strategy exploiting exogenous changes in competence. For this purpose, we use a third dataset (FedScope) which contains detailed characteristics of the public workforce and build instruments for bureaus' competence based on death occurrences of specific types of employees. In particular, one instrument accounts for the death of bureau managers and other white collar employees who are relatively young and high wage and, hence, who are more likely to be relevant for all the initial design stages of a procurement event. The second instrument accounts for deaths in the local bureau where the contract execution is handled and renegotiations are decided.

Both instruments perform well in terms of their statistical properties and they allow us to estimate a causal effect of competence on procurement outcomes that is an order of

magnitude larger than the corresponding OLS estimate. A one standard deviation increase in competence reduces cost overruns by 29 percent and the number of days of delay by 23 percent. It also reduces by half the number of renegotiations. This implies that, if all federal bureaus were to obtain NASA's high level of competence (corresponding to the top 10 percent of the competence distribution), delays in contract execution would decline by 4.8 million days and cost overruns would drop by $6.7 billions over the entire sample analyzed.

Then, we present an attempt to understand what makes a bureau competent. From the FEVS data, we identify three different components of bureau competence: cooperation among employees, incentives and skills. Separately estimating their causal effects would be ideal, but this is unfeasible with instruments like the two described above: the validity of the exclusion restriction, which can be argued to be satisfied when measuring a broadly defined notion of bureau competence, is unlikely to hold for more specific components of competence. Nevertheless, we provide multiple pieces of evidence suggestive that cooperation is the key driver behind the positive effects of competence. The prominence of cooperation conforms with the view that successful procurement requires procurers to appropriately handle and coordinate a multiplicity of tasks involving different individuals and bureaus. The complexity of the environment implies that no one size can fit all: tender and contract design must take into account the often complex characteristics of each specific good, work or service acquired, the existing competition in that market and the characteristics of the pool of potential suppliers, besides the legal principles and available contract management ability and resources. A multidisciplinary approach requiring collaboration among employees with different skills is thus essential.

Finally, we consider the extent to which the role of cooperation is due to the presence of capable managers, able to lead a group to effective cooperation. We exploit the heterogenous effects obtained through instruments considering the deaths of different subgroups of employees, in the spirit of the recent work by Jäger [2017]. We show that the deaths that matter the most are those of relatively young and best paid white-collar employees. Moving along the age and salary dimensions, the estimates change in an intuitive way, with the death of older employees being less consequential in terms of changes in bureau competence.

4

Our quantification of the impact of competence on procurement outcomes confirms the importance of improving decision making within procurement organizations. In the US, efforts to improve procurement capabilities intensified considerably in 1976, when the Federal Acquisition Institute (FAI) was created with the objective of fostering the development of the federal acquisition workforce and certify its competence.[2] In Europe, recent policy initiatives see the introduction of qualification systems for public procurers as a necessary response to the greater discretion granted them by the 2014 Procurement Directives 24 and 25. Some European professional bodies had already developed voluntary qualifications systems for individual procurers (see, for example, the UK Chartered Institute of Procurement & Supply). Existing certification programs, however, have mainly targeted individual contracting officers. Our results on the role of cooperation suggest that, while certification of individual contracting officer's capabilities is certainly welcome and important, it may not be sufficient. Certification programs could be also useful at the level of the procuring office, and should include features such as the organization of the procurement process and the prevailing management practices, as it is often done for private firms.

II   Related literature

Our paper contributes to the recent and growing literature on the determinants of public procurement outcomes.[3] Aside from the obvious concerns about corruption risks, the theoretical literature has offered a variety of explanations for why more competent, higher quality procurers should improve procurement outcomes related to the buyers' involvement in the various stages of the procurement process: the ex-ante design of an adequate award

---

[2]The FAI coordinates several training programs and is complemented by agency-specific programs such as those offered by the Defense Acquisition Institute, that also offers a rich set of certification options for the Department's contracting officers. Other certification programs exist for those performing acquisition-related work in civilian agencies, e.g. the Universal Public Procurement Certification Council.

[3]A number of empirical papers have investigated the role of, for examples, bid preferences (Marion [2007], Krasnokutskaya and Seim [2011], Athey, Coey and Levin [2013]), scoring rule auctions (Lewis and Bajari [2011], Lewis and Bajari [2014]), minimum prices (Chassang and Ortner [2017]), contract duration (MacKay [2017]), electronic procurement (Lewis-Faupel et al. [2016]), transparency (Coviello and Mariniello [2014]), discretion (Coviello, Guglielmo and Spagnolo [2017]), contract renewal (Chong, Saussier and Silverman [2015]), and past performance (Banerjee and Duflo [2000] and Decarolis, Spagnolo and Pacini [2016]).

procedure and contract, the selection of participants and winner(s) at the award stage and, finally, the ex post contract management (see Spulber [1990], Manelli and Vincent [1995] and Bajari and Tadelis [2001]). Nevertheless, a systematic analysis of the hypothesis that buyers' characteristics matter has not yet been undertaken. Among the few studies in this area, the closest papers to ours are Bandiera, Prat and Valletti [2009], Best, Hjort and Szakonyi [2017] and Warren [2014].[4] Contrary to our study, the former two papers focus on the price that different buyers pay for highly standardized goods. Bandiera, Prat and Valletti [2009] estimate that Italian public buyers would save 21 percent of their expenditures if they all paid the same as the buyers at the 10th percentile of the estimated procurement price distribution. The saving amount could reach 1.6-2.1 percent of Italian GDP. Furthermore, they point out that bureaucratic inefficiency is the main cause of waste, accounting for 83 percent of total estimated waste, compared to a 17 percent due to corruption. In a similar vein, Best, Hjort and Szakonyi [2017] reports that 60 percent of within-product price variation in Russia in 2011-2015 can be ascribed to the bureaucrats and organizations in charge of procurement. Like them, we are interested in the extent to which public procurement is affected by the effectiveness of the bureaucracy. Unlike them, we consider complex procurements rather than standardized goods, and measure performance via contract renegotiations instead of purchase price. Thus, while their focus is on outcome measures at the contract design and award phases, our measure also involves the follow-up phase of contract management. Furthermore, our setting is substantially different than theirs, as we focus on the US system rather than Italy or Russia, with obvious economic and institutional differences, and we also are the first to investigate possible channels through which competence emerges.

---

[4]Other relevant papers in this area are Bajari, McMillan and Tadelis [2009], Decarolis [2014], and Bucciol, Camboni and Valbonesi [2017]. Bajari, McMillan and Tadelis [2009] analyzes auctions versus negotiations. Employing a dataset of private sector building contracts awarded in Northern California during the years 1995-2000, they find that project characteristics affect the choice of the award mechanism and that auctions are used more often by more experienced buyers (i.e., those in organizations that are larger and procure contracts more frequently). Decarolis [2014] studies procurement outcomes in terms of ex post contract renegotiations and shows that they depend on the choice of the procurement mechanism and on the level of bid screening undertaken by the buyer. Large buyers, who are the most experienced, are better able to screen offers, as shown by the better outcomes in terms of time and cost renegotiations for given contract choices. Our paper complements these studies by analyzing procurer quality features that go beyond the mere frequency of tendering and organizational size; we measure their impact and investigate the specific channels through which these features affect procurement outcomes. Bucciol, Camboni and Valbonesi [2017] tries to capture the effects of buyer quality through buyer fixed effects.

The IV strategy that we devise is also novel for the procurement literature where variations driven by managers' deaths have not been previously exploited as a source of identification. Related but different is Warren [2014], which uses retirement-induced workload spikes for procurement specialists to document an economically important effect of shortages in these specific employees on civil agencies' procurement outcomes. The focus is therefore on the quantity of public employees, while our work looks at how bureaucratic quality rather than quantity, affects procurement outcomes.

At a more general level, our results are relevant to the growing literature documenting the heterogeneity of employees and organizations that implement state policies within and across countries. Besley and Persson [2009], Besley and Persson [2010] and Acemoglu, Garcia-Jimeno and Robinson [2015] have stressed the importance of "state capacity," the ability of the state to effectively provide the fundamental public goods necessary for the private economy to flourish and lead to growth. Part of this literature, like Dal Bo, Finan and Rossi [2013], Bai and Jia [2016] and Bertrand et al. [2016], has focused on the determinants of government performance related to the selection and recruitment of personnel, incentives, and monitoring activities (see Finan, Olken and Pande [2015]). This literature mostly focused on developing countries. Our paper contributes to it by providing an assessment of the importance of public sector management quality for a large developed country like the US.

Finally, our focus on different types of bureau competence connects our work to the recent literature on the role of managerial practices in the public sector (Bloom et al. [2014], Bloom et al. [2015]). Closer to our theme, Rasul and Rogger [2016] show that management practices affecting autonomy correlate robustly with public project completion in Nigeria, while practices related to incentives/monitoring of bureaucrats are negatively associated with completion rates. Consistent with the findings in this literature, we document a variation in the quality of US procurement agencies, but in contrast to Rasul and Rogger [2016], we do not find a clear negative effect of incentives. Incentives in the public sector might thus play a different role in strong and weak institutional environments. Finally, our findings on cooperation also square well with results in Blader, Gartenberg and Prat [2016] on the benefits of "cooperative" managerial practices relative to high powered individual incentives.

III Institutional Background

Our analysis combines procurement data at the individual contract level with competence data, which are at the bureau level. We indicate as *bureaus* the sub-units of the U.S. federal government agencies. All federal agencies, whether executive (i.e., analogous to ministers common in parliamentary or semi-presidential systems) or independent, will be indicated as *agencies* throughout this study. Each agency has its own organizational structure according to which its power is exercised through different sub-units, the bureaus. Bureaus are charged with a specific mission depending on the agencies they are affiliated to.

The bureaus in charge of contract procurement typically have an organizational structure that we will exploit to devise our instrumental strategy. As source selection guides reflect, the planning and competition process often calls for input and involvement by both central and regional bureau officials, the latter being those located in the area of contract execution.[5] Albeit with exceptions, the general practice in U.S. federal procurement is to have the tendering and award processes overseen by a "procurement contracting officer" (PCO) in a regional office. The local PCO can receive extensive input from attorneys, program personnel and other experts in the central headquarters, who can even overrule the regional officer's decision. Thus, the training and judgment of both the regional officer and the headquarters personnel may have a material impact on the quality of the contract award. Instead, the actual management of the contract is typically done only by a regionally located contracting officer. The PCO decides on contract renegotiations by handling the claims for additional costs or time from contractors during the course of contract execution.[6] We will return to the distinction between central and regional bureaus when discussing our instruments.

---

[5]A useful example of a guide mapping out a typical source selection process is the Army Source Selection Guide, which in turn complements the "master" guidance, the Defense Department's general source selection procedures, which are called out in the Defense Federal Acquisition Regulation Supplement (DFARS).

[6]Contracts are typically competed and signed by the PCO, in a regional office. There may be a division of roles - the source selection authority (SSA) on a major procurement, for example, may be a senior officer and not the PCO, or the PCO and the administrative contracting officer (ACO) may be different people - but in general, the roles and responsibilities are centered on one official, the PCO, who typically serves as the SSA. The SSA is an extremely important figure - he or she is the pivotal official who must exercise discretion in the contract award and management. We are grateful to Christopher Yukins for all his precious advices, which proved to be fundamental to understanding the working of federal procurement organizations.

IV Data

This section presents our three data sources. We first discuss the survey data measuring bureau competence, then the procurement data from which we construct the performance outcomes, and finally the federal employees' characteristics data used for the IV strategy.

A. Federal Bureau Competence: FEVS Data

The principal explanatory variables that we use to measure bureau competence come from the Federal Employee Viewpoint Survey (FEVS). Since the early 2000s, the Office of Personnel Management has called on federal employees to provide their opinions on all aspects of their employment, including evaluations of their supervisors, bureaus, agencies and, more generally, of their work experience. The goal is to measure government employees' perceptions of whether, and to what extent, conditions characterizing successful organizations are present in their bureaus and agencies and, ultimately, to influence change in their workplace. The beginning of this survey dates back to 2002 when it was first administered under the name "Federal Human Capital Survey" as an essential tool of the George W. Bush administration's agenda for a managerialization of the public administration. Since then, the survey has been mainly used for internal human resources management recommendations from the Office of Personnel Management to the agencies. This office uses the FEVS to monitor human capital management initiatives and outcomes and to provide guidance, resources, and technical assistance to the entire federal government. Despite the existence of published works based on FEVS data (see the survey review of Fernandez et al. [2015]), ours is the first to reconcile them with the procurement data discussed next.

We focus on all bureaus that in a year procure at least one contract, over the 2010-2015 period. By focusing on this period, we can use yearly data since before 2010 the FEVS was run every other year. There is a total of 96 bureaus from 23 agencies. The agencies that are invited to participate account for 97 percent of the executive branch workforce with about half of the employees randomly selected to participate in the survey and an average 47% response rate. The FEVS consists of 85 questions divided into five different sections which appear to respondents in the following order: my work experience, my work unit,

my agency, my satisfaction and work/life. The section "my work unit" begins with eight questions pertaining to different features of the bureau and ends with a ninth question aiming to capture the overall effectiveness of the job done in the office.[7] This is the only question in the survey that can proxy for a self-evaluation of the overall work conducted by individual work units within each agency. Therefore, we use this variable as our main measure of overall bureau competence and label it *competence*. To distinguish bureau features from agency features, we will also use the summary question from the section "my agency" which asks whether "The workforce has the job-relevant knowledge and skills necessary to accomplish organizational goals". We label this variable *Ag.competence*.

For all questions, employees' responses are in five ordered levels of intensity. For the typical question, the possible responses are: very poor, poor, fair, good, very good.[8] We first transform these answers into numerical values from zero to four, then we aggregate answers at the bureau level (by using the FEVS' representative weights), and finally we normalize the resulting variables to be between zero and one. The top panel of Table 1 reports summary statistics for the main FEVS variables: *competence* and *Ag.competence*, as well as three additional variables that will be analyzed as the components of bureau competence in the final part of this study and that we indicate as cooperation, incentives and skills.

B. Procurement Outcomes: FPDS Data

To construct measures of procurement performance and retrieve other contract-specific information, we use the Federal Procurement Data System (FPDS), the source for U.S. government-wide procurement data. Since fiscal year 2000, federal bureaus complete reports on procurement contract actions that feed the FPDS.[9] The data track every transaction between federal contracting bureaus and sellers. The system contains detailed information on contract actions over $3,000 (fiscal year 2004 and later data).[10] Information is of two

---

[7]As reported in the introduction, this question asks: "How would you rate the overall quality of work done by your work unit?". The full list of questions composing this section is reported in Table 11.

[8]The respondent can also report "do not know" or leave the question unanswered, but both occurrences are rare (typically less than 2 percent of the responses for each of these two cases).

[9]These data have been used to research key features of the US public procurement system in several studies, including Liebman and Mahoney [2017], Warren [2014], Kang and Miller [2017] and Giuffrida and Rovigatti [2017].

[10]Data are downloadable at *https://usaspending.gov*.

Table 1: Summary statistics

	Mean	Median	S.D.	N
Bureau Characteristics (FEVS Data)				
Competence (Q28)	0.81	0.81	0.03	445
Ag.Competence (Q29)	0.69	0.69	0.04	445
Cooperation (Q20)	0.73	0.73	0.04	445
Skill (Q21)	0.54	0.54	0.05	445
Incentive (Q23)	0.45	0.45	0.05	445
Contract Characteristics (FPDS Data)				
Contract Amount	531.7	87.0	3595.7	122533
Expected Duration	244.0	212	208.0	122533
Cost Performance	0.85	1	0.25	122533
Time Performance	0.73	1	0.33	122533
Total Cost	891.6	109.2	7127.4	122533
Total Time	485.7	364	703.4	122533
No. of Cost Ren.	1.29	0	4.58	122533
No. of Time Ren.	1.17	0	4.01	122533
No. of Tot. Ren.	2.47	1	8.25	122533
Bureau Experience	26.3	4.30	52.5	122533
No. of Offers	3.84	2	6.17	122533
Works	0.19	0	0.39	122533
$Bu.Performance^{C}$	0.84	0.89	0.15	112658
$Bu.Performance^{T}$	0.73	0.73	0.18	112658
Bureau Experience	26.3	4.30	52.5	122533
Bureau Size	1.14	0.44	1.37	122533
No. of Offers	3.84	2	6.17	122533

Notes: The top panel presents summary statistics for the FEVS data. The unit of observation is a bureau-year. Not all 96 bureaus are observed for all years due to organizational changes within the 23 agencies covered. The bottom panel presents summary statistics for the FPDS data. *Bureau Experience* is scaled down by thousands of US dollars.; *Contract Amount* and *Total Cost* are expressed in thousands of US dollars; *Expected Duration* and *Total Time* are expressed in days; *Cost Performance*, *Time Performance*, *Competence*, $Bu.Performance^{C}$, and $Bu.Performance^{T}$ are bounded between 0 and 1; all variables are described in the main text.

kinds: *a)* data concerning the contract and the awarding stage, and *b)* data concerning the subsequent life of the project (i.e., contract amendments) which are also classified according to the reason for the modification.

We focus on the procurement of services and works where, compared to that of goods, the extent of ex-post cost uncertainty makes post-award amendments, with the high haggling cost they imply, a useful proxy of contract performance (Tadelis [2002]).[11] Since not

---

[11]The web appendix discusses these sample selection choices. In the literature, post-award modifications are widely used as a proxy for wasteful spending. Spiller [2008] argues that given the formal nature of public contracting, any terms renegotiation would add adjustment costs, providing weaker incentives to adapt for both contractors and public authorities. Bajari, Houghton and Tadelis [2014] provide support to this hypothesis by quantifying in 8 to 14% of the winning bid the adaptation costs in their construction

all modifications are equally problematic, we split the set of amendments to two broad categories: *in-scope* and *out-of-scope* revisions.[12] In line with other studies that use FPDS data, we consider *in-scope* amendments only.[13] The quantitative relevance of modifications is evident from the summary statistics reported in the bottom panel of Table 1. Our sample ranges from 2010 to 2015 and consists of 122,533 projects, associated to 821 categories (i.e., the typology of work or service procured). Although the overall value of the contracts is $65.2 billion using the initial awarding price, it increases to $109.3 billion if cost overruns are included. The distribution of contract amounts is highly skewed: fifty percent of contracts are for amounts below $87,000, while 10 percent of contract spending is accounted for by contracts worth more than $757,000. The average award per contract is $531,700, while the total cost, inclusive of any cost overrun, is $892,000. Correspondingly, the average contractual duration is 244 days, while the final contract duration inclusive of any delay is 486 days. In both cases, the medians are lower than the means.

To operationalize the data on time and cost renegotiations into a proxy for contract performance we proceed as follows. We define: *Time Overrun* as the difference - in days - between the actual completion date and the estimated completion date, and *Cost Overrun* as the sum - in thousands of dollars - of all renegotiated amounts. In order to compare the two overrun measures with the initial expected outcomes - that is, the time/cost of completion specified in the contract terms - we specify two indexes for contract performance like: $performance_{ijt}^{g} = \frac{expected\ outcome_{ijt}^{g}}{expected\ outcome_{ijt}^{g}\ +\ overrun_{ijt}^{g}}$, where the superscript $g = \{T, C\}$ distinguishes between the time and cost measures, the subscripts (i, j, t) refer to contract, bureau and time, *expected outcome* is the initial contract value (in dollars for cost and days for time) and *overrun* is either the cost overrun or the delay. Each performance measure ranges be-

---

data. Markups from private information and market power, the focus of much of the literature, are typically much smaller. For related arguments on the waste associated with time and cost renegotiations in public contracts see also Lewis and Bajari [2011], Bajari, Houghton and Tadelis [2014], Guasch, Laffont and Straub [2008] and De Silva et al. [2017].

[12] According to the FPDS data dictionary, we label as *out-of-scope* all amendments classified as "Additional Work (new agreement, FAR part 6 applies)", "Novation Agreement", "Vendor DUNS or name change - Non-Novation" and "Vendor Address Change". We consider all other amendments as being *in-scope*.

[13] An alternative based on a categorization used in a recent work by Kang and Miller [2017] is discussed in the appendix. Essentially, they exclude some *in-scope* revisions, but also retain some of the *out-of-scope* revisions. When we adopt this alternative definition we find very similar results to those in our baseline estimates (see Table A.3 in appendix).

tween zero, worst performance, and one, perfect performance (i.e., no overrun). In the data, about half of the observations show no cost or time overruns. The coefficient of the linear correlation between the two equals 0.52 with a Spearman ρ of 0.57.

As mentioned in the introduction, additional outcome measures will also be the number of renegotiations, both overall and separately for cost and time purposes. Table 1 reports summary statistics for these variables as well as for other FPDS variables that will be used as controls. Among these variables, *Bureau Experience* and *Bureau Size* will be particularly interesting as past studies have often used them as proxy for buyers' competence. The first variable measures the number of times a bureau has appeared in the past in the data for the same contract category, while the second measures the cumulative value of contracts a bureau has awarded in the same year in the same service category.

Figure 1: State of Contract Performance
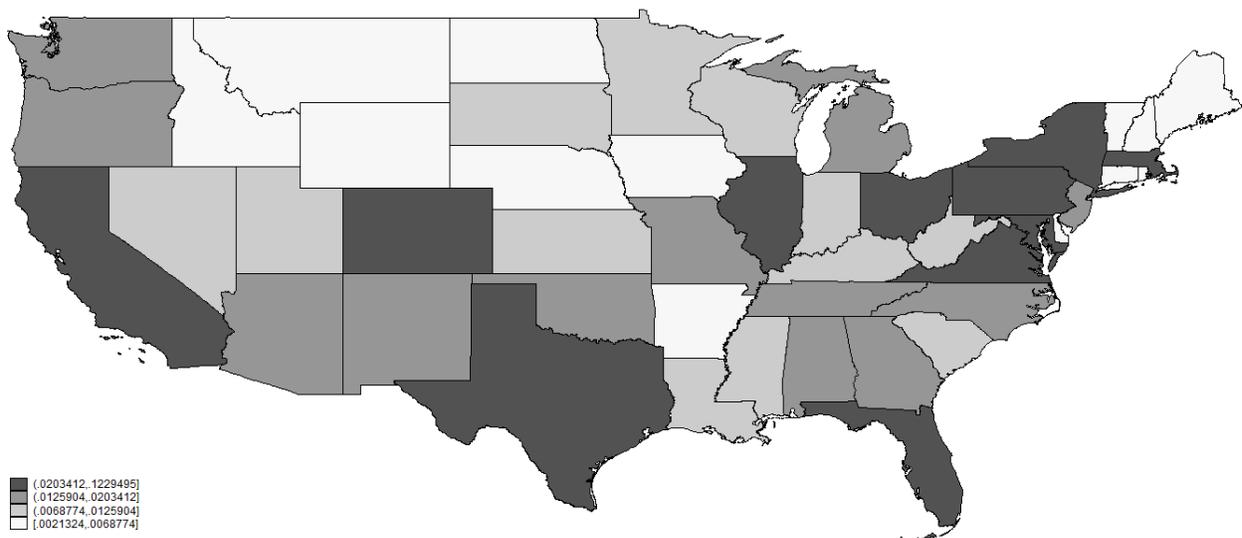
Notes: percentage of contracts associated to each state across our sample. Colors represent the quartiles of the distribution (white 1st quartile to dark grey, 4th quartile).

The data also exhibits geographical variation in the place of contract execution that we document in Figure 1. More contracts take place in more densely populated states (12% of all contracts are in California), but all states have at least some contracts. This matters because, as mentioned earlier, central and local procurement offices play different roles in terms of contract design and contract management, with the latter delegated to local offices.

C. Public Workforce Characteristics: FedScope Data

The Office of Personnel Management (OPM) is an independent federal agency that functions as the central human resources department of the executive branch. In fulfilling its mission, OPM collects, maintains, and publishes data on a large portion of the federal civilian workforce. In FY 2010, OPM established a system called the Enterprise Human Resources Integration Statistical Data Mart (EHRI-SDM). This system provides access to personnel data for 96% of federal civilian executive branch employees.[14] These data are released through the Federal Human Resource (Fedscope) database, which represents the most comprehensive resource available on the size and scope of the federal workforce. Fedscope is the third data source that we use by merging it with FPDS at bureau level.[15] FedScope data are presented in five subject categories, called "cubes." For the purposes of the present paper, we only consider the "Employment" cube and the "Separations" cube for the years 2010-2015.

The Employment cube contains several demographic characteristics along with information on appointments and tasks, e.g. length of service, occupation category, pay grade, salary level, type of appointment, work schedule, and location of each single employee. The Separations cube contains all the separation occurrences in the public workforce: employees who transferred to other bureaus or agencies, voluntarily resigned, retired, experienced a reduction-in-force, were terminated, or died while employed. The IV variables that we will use are based on the occurrence of death events in the bureaus. This is achieved by combining the two cubes in order to obtain, for each bureau and year, the combination of deaths by age and salary. Moreover, since the Employment cube allows distinguishing managers and other white-collars workers from the other employees, we will focus on the former group of employees, whose separations from a bureau is most likely to have an impact on the bureau's competence. In Table 2, we report quantiles of age and salary of the managers and other white-collar employees: a total of 2.5 million employees, subdivided into 96 bureaus

---

[14]The database does have exclusions involving, for example, some national security and intelligence agencies and the Postal Service.

[15]This is possible through an external dictionary which maps the variable "*Contracting Office Agency ID*" in FPDS to the variable *AGYSUB* of Fedscope. To ensure temporal coherence with FPDS and FEVS, we employ the September snapshot of FedScope's "Employment" cube.

Table 2: Quantiles of Age and Salary

	Managers		Other White-Collar Employees	
	Age	Salary	Age	Salary
1 %	25-29	$40,000 - $49,999	20-24	$20,000 - $29,999
5 %	30-34	$50,000 - $59,999	25-29	$30,000 - $39,999
10 %	35-39	$50,000 - $59,999	25-29	$30,000 - $39,999
25 %	40-44	$70,000 - $79,999	35-39	$40,000 - $49,999
50 %	50-54	$90,000 - $99,999	45-49	$50,000 - $59,999
75 %	55-59	$120,000 - $129,999	50-54	$80,000 - $89,999
90 %	60-64	$150,000 - $159,999	60-64	$110,000 - $119,999
95 %	60-64	$160,000 - $169,999	60-64	$120,000 - $129,999
99 %	65 or more	$180,000 or more	65 or more	$170,000 - $179,999
Obs	1,342,306	1,342,306	7,099,127	7,099,127
Std. Dev.	1.78	3.53	2.36	3.29
Av. # employees	648	648	3,379	3,379
Md. # employees	106	106	477	477
Employees Std. Dev.	1,795	1,795	13,345	13,345
Local Av. # employees	50	50	190	190
Local Md. # employees	8	8	16	16
Local Employees Std. Dev.	155	155	778	778

Notes: The table reports the distribution of age and salary separately for two groups of employees, managers and other white-collar employees during the time window. The sample is that of the employees in the 96 bureaus that we observe in the FPDS and FEVS, which represent more than 90 percent of the entire workforce covered by FedScope. 1 point S.D. in *Age* represents 5 years; 1 point S.D. in salary $10,000.

that have on average 648 managers and 3,379 other white-collar employees at the national level and 50 managers and 190 other white-collars employees at the local level. Finally, the geographical information in FedScope enables us to match the location (state) of each single federal employee with that of contract performance.[16] More details on the specific ways in which these data are used to construct our instruments are presented in section VI. Before that, however, in section V we present some relevant descriptive facts about the data that serve to establish the link between the FEVS and FPDS data.

V   Descriptive Evidence

Before trying to assess any causal effect of bureau competence on procurement outcomes, it is useful to explore the data to establish two facts. First, we show that the relevant variation in performance occurs at the bureau and not at the agency level. Second, we argue that

---

[16]In the appendix we provide a full list of states where bureaus have employees; see Figure A.1.

the naive association between the competence measures from the FEVS and the procurement performance proxies is likely to underestimate the benefits of greater competence on procurement.

Figure 2: Procurement Performance across Bureaus and Agencies
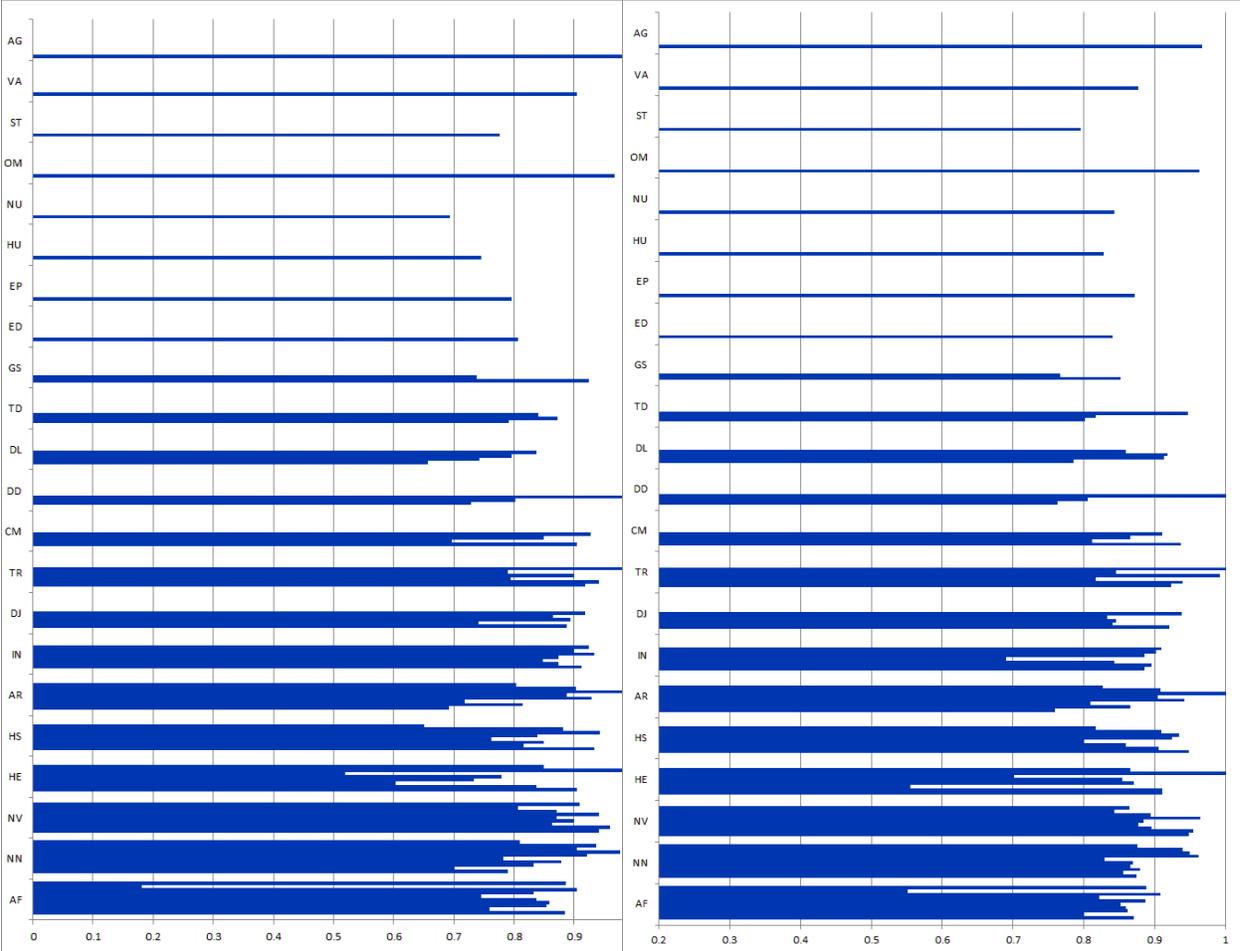
(a) $Bu.Performance^C$ (b) $Bu.Performance^T$

Notes: The table reports the distribution of average $Bu.Performance^C$ and $Bu.Performance^T$ across all bureaus and agencies. For each agency, we report the overall average performance of all bureaus with which the agency appears in the FPDS. The length of the horizontal lines measure the performance of $Bu.Performance^C$ (left column) and $Bu.Performance^T$ (right column).

To illustrate the first point, we begin by constructing a bureau-level performance measure based on the procurement data only. Thus, we aggregate *time performance* and *cost performance* into two performance measures at the bureau-level: $Bu.Performance_t^g$ with $g = \{C, T\}$ for cost and time performance, respectively. These are constructed by aggregating the contract-level performance measures for all contracts i that, at any given date t, the

16

bureau had previously procured for the same contract category j: $Bu.Performance_{ijt}^{g} = \sum_{ij\{\mathbf{1}|t'<t\}} w_{ijt'} * performance_{ijt'}^{g} / \sum_{ij\{\mathbf{1}|t'<t\}} w_{ijt'}$, where w are weights that are larger for more recent contracts. We use these two performance measures to establish what follows.

First, we seek to show that the bureau is the right unit of analysis with which to link the FEVS and FPDS data. Since the FEVS data contain questions at both the bureau and the agency level, it is important to understand whether the bureau is indeed the most relevant unit of observation. Figure 2 shows why aggregating at agency level would result in missing a substantial share of the variation in performance. There we report the distribution of the bureau-level performance measures across all bureaus and agencies. For each agency, we report the performance of all bureaus with which the agency appears in the FPDS. The length of the horizontal lines measures the performance of $Bu.Performance^{C}$ (left) and $Bu.Performance^{T}$ (right). From this it is clear that, although there is some variation at the agency level, most of the action takes place between bureaus within agencies. This is particularly the case for the time performance measure.

Table 3: BEST AND WORST AGENCIES (COMPETENCE)

Agency	Competence	Ag. Competence	$Bu.Performance^{C}$	$Bu.Performance^{T}$
NRC	.86	.76	.60	.59
NASA	.86	.74	.75	.68
DVA	.79	.67	.86	.71
DOJ	.76	.69	.85	.73

Notes: Average agency scores for *Competence*, *Ag.Competence*, *Bu.Performance^{C}*, and *Bu.Performance^{T}* reported for the two best agencies in terms of average Competence - Nuclear Regulatory Commission (NRC) and National Aeronautics and Space Administration (NASA) - in the two top rows and the two bottom agencies - Department of Veteran Affairs (DVA) and Department of Justice (DOJ) - in the two bottom rows.

Second, to better understand the relationship between these two competence variables, as well as between them and contract performance, we present the case of the four agencies at the extremes of the bureau competence measure. This case study will be illustrative of the downward bias concern driving our IV strategy in the next section. Table 3 reports competence and performance measures of the top two agencies in terms of bureau competence - averaged across all the bureaus in the agency - which are the NRC (Nuclear Regulatory Commission) and NASA, both with an average *competence* equal to 0.86, and the worst two, which are DVA (Department of Veteran Affairs) and DOJ (Department of Justice),

with an average *competence* equal to 0.79 and 0.76, respectively. The corresponding values of *Ag.competence* across these four agencies in Table 3 also indicate a marked difference between the top and bottom two agencies. The last two columns of Table 3 report the values of the two performance measures for the four agencies considered.

Figure 3: Dynamics of the Main Measures
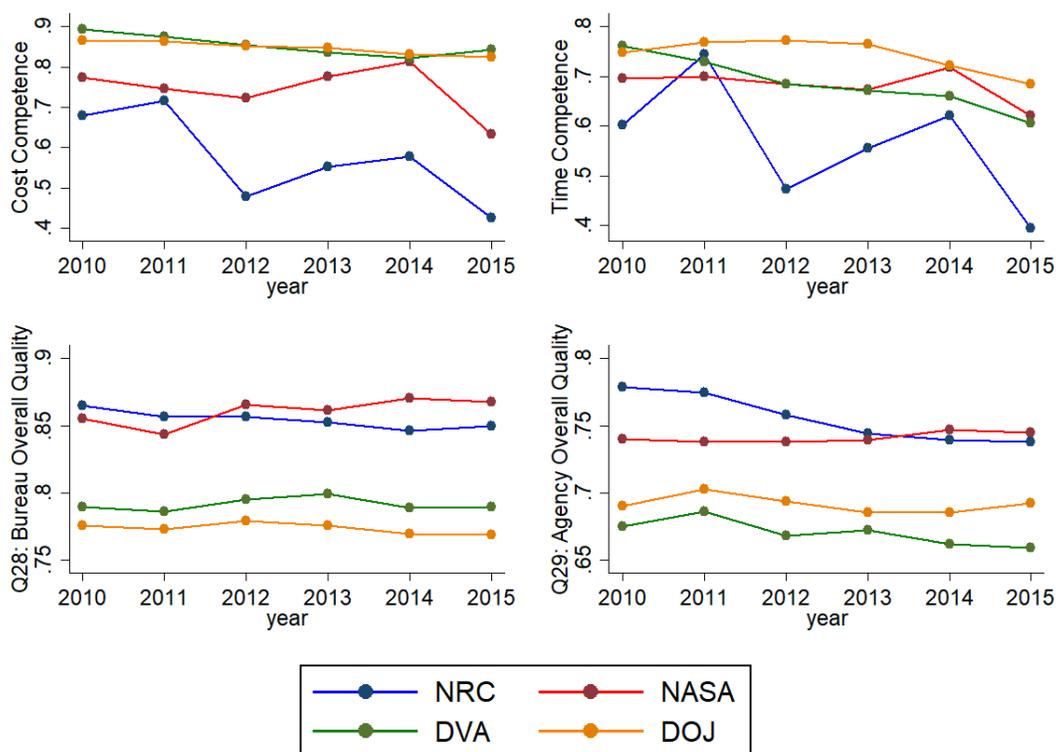
Notes: Evolution of yearly average agency scores for - from top left to bottom left, clockwise - $Bu.Performance^C$, $Bu.Performance^T$, $Ag.Competence$, and $Competence$ - reported for the two best agencies in terms of overall average *competence* (Nuclear Regulatory Commission (NRC) and National Aeronautics and Space Administration (NASA)) and the two worst agencies (Department of Veteran Affairs (DVA) and Department of Justice (DOJ)).

Figure 3 shows the evolution over time of the four variables for each of these four agencies. It reveals that the evidence based on the sample averages reported in Table 3 is persistent over time. Thus, by comparing the relative rankings of the four agencies across the four columns, it is impossible to see any positive association between bureau (or agency) competence and contractual performance. Indeed, the performance of the agencies that are worst in terms of *competence* (DVA and DOJ) is superior to that of the two most competent agencies (NASA and NRC) in terms of both time and cost. This striking inversion of the relative ranking is a key features of the economic environment that we analyze and around which we construct

our empirical strategy: more competence is associated with more complex contracts, which are intrinsically associated with higher levels of delays and cost overruns.

VI   Empirical Analysis

To assess the relationship between bureau competence and procurement performance, we begin by estimating the following linear regression model:

$$performance^g_{ijkct} = \beta\, competence_{jt} + \theta\, \mathbf{X_{ij}} + \kappa_k + \zeta_z + \tau_t + \varepsilon_{ijkct}, \tag{1}$$

where $g = \{C, T\}$ indicates whether the outcome variable is cost or time performance, i, j, k, z, and t indicate contract, bureau, agency, service category and year, respectively; $\mathbf{X_{ij}}$ is a matrix of contract- and bureau-level covariates, and ι_j, ζ_z, and τ_t indicate agency, service category and year fixed effects, respectively. In the estimates we also include state fixed effects. Bureau fixed effects, instead, are not included as the high degree of persistency of competence over time makes it unfeasible to identify competence when these fixed effects are included.[17] This has the important implication that the source of identification of the coefficient of interest β - the effect of the bureau competence on contract performance, conditional on the other regressors - is cross-sectional. In one of the robustness checks, we show how the estimates change when bureau fixed effects are included.

There are several challenges in interpreting the OLS estimate as a causal effect. First, our survey measure of competence is likely to be a noisy proxy for the set of characteristics that would ideally measure a bureau's competence. Individuals could misreport their bureau quality for a variety of reasons ranging from simple biased perceptions to sophisticated strategies to exploit how the OPM ensuing recommendation might benefit them. Moreover, measurement error may also arise from surveying recording errors, sampling errors, and differences between the true and respondent's reported judgments that are associated with

---

[17]The persistency of the competence measures is evident for the case of the four agencies shown in Figure 3. A more in depth analysis of this aspect is presented in an earlier NBER WP version of this study and is available upon request from the authors.

the coarseness of the possible answers. Furthermore, and more crucially for this study, as discussed above for the case of the two most/least competent agencies, competence and performance might move in opposite directions due to the mere association of more competent bureaus with more complex procurement projects.[18]

Our approach to addressing these potential concerns is twofold. First, we exploit the richness of our data to include in the model specifications all observable characteristics likely contributing to explaining performance. In particular, we always include agency and service category fixed effects to capture the differences in the types of procurement across both agencies and contracts. We also control for the contract initial amount and duration to proxy for contract complexity. Then, we gradually include controls for *Bureau Experience* and *Bureau Size*, for the motives mentioned above, and additional fixed effects for the state where the contract is based. There are, however, multiple features of the project design and management that most likely we cannot observe and that pose the risk of an omitted variable bias in our estimate of β.

Therefore, the second element of our strategy is an instrumental variable (IV) approach. The variables we employ as instruments are derived from FedScope, through which we observe bureau employees' deaths. We exploit the richness of the data to evaluate the public workforce under different aspects and construct two instruments that capture the distinct roles that central and local procurement officers can have on the procurement process.

Table 4: INSTRUMENTS SUMMARY STATISTICS

	Mean	Median	S.D.	N
Relevant Deaths	0.91	1	0.29	122533
Proximal Deaths	0.64	1	0.48	122533

Notes: The table presents summary statistics of the instruments employed in the IV analysis. Both *Relevant Deaths* and *Proximal Deaths* are dummy variables.

First, inspired by the vast literature on CEO deaths, we focus on deaths of those employees more likely to have positive roles for the productivity of their office. We thus look

---

[18]One might also worry about reverse causation, but this is unlikely to be an issue because the respondents to the FEVS survey are not limited to workers dealing with procurement. Hence, the performance of procured contracts should not directly affect the typical survey respondent. Nevertheless, any remaining risk of reverse causation bias is addressed by our IV strategy presented next.

at white-collar employees of an age no higher than the median and with a salary no lower than the median, relative to the distributions of these variables for other white-collar employees. According to Table 2, this implies looking at employees with a salary of \$50,000 or more and an age of 50 years or less. Such thresholds value are able to capture 95% of the manager population and the upper half of the other white-collar employees. We thus build our first instrument as a dummy indicating whether a death of at least one employee in this age/salary groups occurred within a bureau-year:

$$Relevant\ deaths_{jt} = \mathbb{1}Death[age <= 49, salary >= 50k]_{jt}, \tag{2}$$

where j is the bureau and t the year. Table 4 reports the summary statistics for this instruments which are most easily understood through Figure 4a. This figure illustrates for all the bureaus-years in the sample, the distribution of the share of deaths within the relevant age/salary population. It reveals a well-behaved distribution with 9% of the bureau-year observations being zero deaths and only a few extreme observations (to the exclusion of which our estimates are robust). The reason for the effectiveness of this variable as an instrument for competence can be deducted from Figure 4b. In this figure, we report the median value of *deaths* for each combination of age and salary levels. The median value of *deaths* increases monotonically in age, with salary having little effect (especially below the \$100,000 salary, where most observations lie). This implies that for the group of individuals that we consider to be important for the well functioning of a bureau (i.e., young with a relative high salary), deaths are particularly unlikely. Thus their occurrence will be particularly disruptive. We return to this aspect after having introduced the other instrument.

For a second instrument, we follow Bruce, Figueiredo and Silverman [2017] who suggest that the spatial proximity of a death event in the procurement agency can be relevant to contractual performance. By exploiting this variation, we construct our second instrument, *proximal deaths*: a binary variable indicating whether at least one death event among white-collars employees of the bureau awarding the contract has occurred in the same state of the contract's place of performance and in the same year of the contract awarding. To avoid

Figure 4: Count of Death events divided by the workforce base

(a) Histograms at contract level
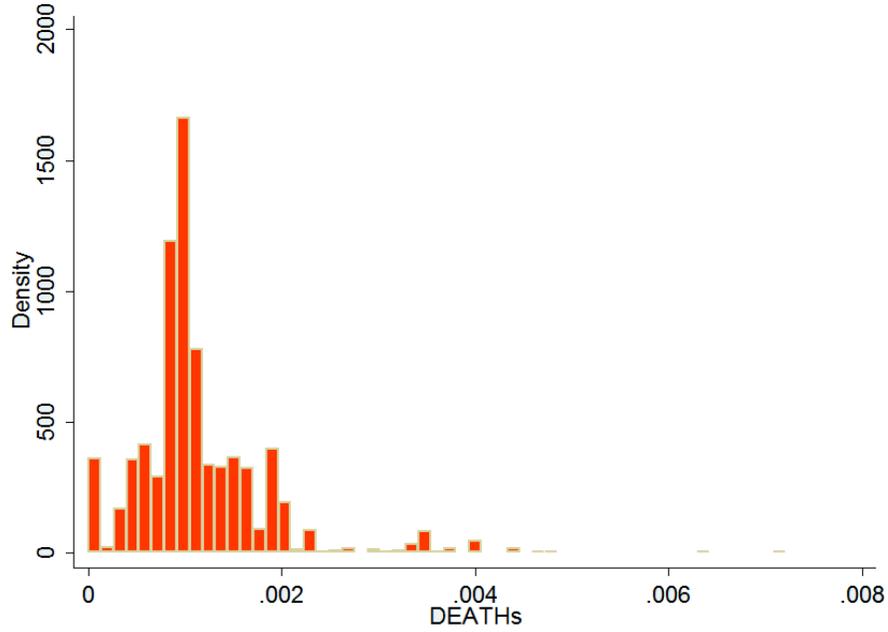
(b) Median frequency by Age and Salary
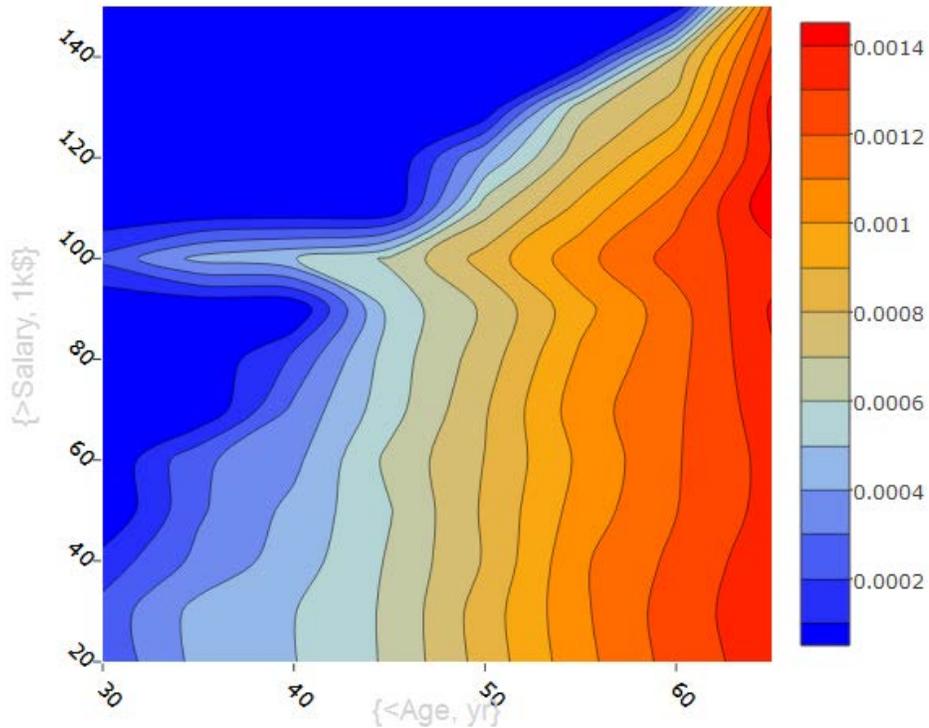
Notes: In panel (a), we report the histogram of the ratio between the count of death events and the workforce dimension for each bureau. In panel (b), we report the median value of the ratio for each combination of age and salary.

ambiguities in interpreting a value of zero for this instrument, we exclude from the sample all the contracts that are performed in a state in which no employees of the awarding bureau are located (around 4% of the working sample).

The relationship between deaths and competence is apparent from the "visual first stage" reported in Figure 5. This figure shows the relationship between our two instruments, *relevant deaths* and *proximal deaths* and *competence*. A clear negative association is present in both panels. This evidence supports the presence of a powerful first-stage relationship that will be more formally assessed below.

Figure 5: Visual Representation of the First-Stage
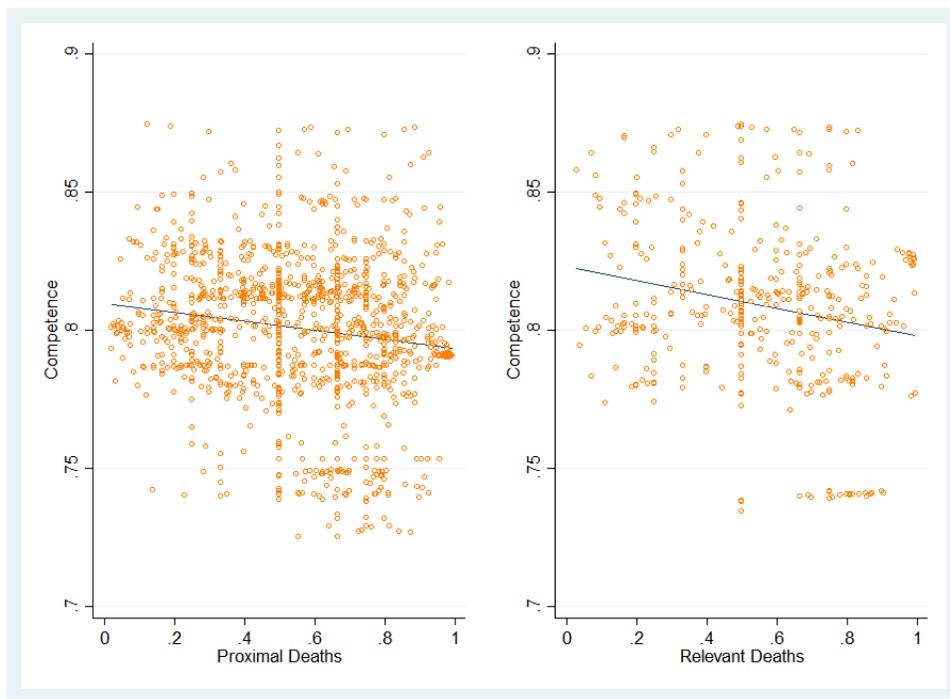
Notes: Graphical representation of the relationships between *Competence* with *Proximal Deaths* - left panel - and *Relevant Deaths* - right panel. Observations are collapsed at cluster level (bureau and service category).

Before presenting the IV results, however, we conclude this section with a discussion of the instruments. While we are unaware of other studies on procurement exploiting the deaths of public officials as a shock to bureau competence, the use of death occurrences (or inability to work) of CEOs and their relatives as instrumental variables for the productivity of firms

23

has a long tradition in economics.[19] This literature suggests that individuals in charge of high-responsibility tasks, such as managers, supervisors and team leaders, account for most of the competence in an institution. The validity of the instrument is supported by the fact that as-good-as-random separations of office managers negatively affect the competence level of the whole office through two obvious channels. First, a sudden separation determines a vacancy of skills in terms of knowledge and prompt decisions of management. Second, the managerial literature evaluates the so-called onboarding effect, and estimates as the time a newly hired officer needs to reach full productivity to be eight months. The latter feature is clearly present in the federal workforce, as new hirings are notoriously slow due to the need to resort to public evidence procedures while transfers of workers are hindered by the limited ability to negotiate financial incentives.

Although the potentially endogenous relationship between workplace quality and deaths might create a concern, there are two pieces of evidence suggesting this is not the case. First, even though FedScope does not allow to distinguish between death causes, we use different statistical sources to assess suicide rates. Suicides are a good proxy for deaths associated with stress and depression, which could be driven by features of the procurement process. But both the Survey of Occupational Injuries and Illnesses and the Census of Fatal Occupational Injuries show zero suicides among federal managers in our sample years. Second, we perform a regression analysis (see Table 5) to identify the determinants of our two instruments. We find that while these deaths are associated in an intuitive way with education, salary and health security of the workplace, they are not associated with any of the procurement measures appearing at the top of Table 5: the total available budget, the number of pledged contracts, or the average amount of pledged contracts. The magnitude of coefficients is nearly zero and none reaches statistical significance.

---

[19]For recent instances, see Becker and Hvide [2013], Bennedsen, Pérez-gonzález and Wolfenzon [Forthcoming] and references therein. See also Jäger [2017] for a detailed account of the spillover effects of an employee's death on coworkers. Other related papers include Azoulay, Zivin and Sampat [2011] on the spillover effects of research superstars, and Jones and Olken [2005] to evaluate the role of national leaders.

Table 5: DEATH OCCURRENCE PREDICTORS

	Proximal Deaths				Relevant Deaths			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Budget	0.00	0.00	0.00	0.00	-0.00	-0.00	-0.00	-0.00
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
N of contracts	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Mean Amount	-0.00	-0.00	-0.00	-0.00	-0.00	-0.00	-0.00	0.00
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Median Age	0.02	0.02	0.02	0.02	-0.03	-0.02	-0.03	-0.02
	(0.02)	(0.02)	(0.02)	(0.02)	(0.05)	(0.05)	(0.06)	(0.06)
Median Education	-0.01	-0.01	-0.01	-0.01	-0.07**	-0.07**	-0.07**	-0.07*
	(0.01)	(0.01)	(0.01)	(0.01)	(0.03)	(0.03)	(0.03)	(0.03)
Median LOS	-0.00	-0.00	-0.00	-0.00	-0.09*	-0.07	-0.08*	-0.07
	(0.02)	(0.02)	(0.02)	(0.02)	(0.04)	(0.05)	(0.05)	(0.05)
Median Salary	-0.02*	-0.02	-0.02*	-0.02*	-0.01	-0.01	-0.01	-0.01
	(0.01)	(0.01)	(0.01)	(0.01)	(0.06)	(0.06)	(0.06)	(0.06)
Median WF Composition	-0.01	-0.01	-0.01	-0.01	-0.06	-0.07	-0.07	-0.07
	(0.03)	(0.03)	(0.03)	(0.03)	(0.10)	(0.10)	(0.10)	(0.10)
Accomplishment		0.33	0.73	1.14*		1.13	3.87	2.61
		(0.60)	(0.73)	(0.65)		(2.34)	(3.16)	(3.35)
Appreciation		-0.31	-0.35	0.08		-1.53	-2.28	-2.15
		(0.84)	(0.91)	(0.96)		(3.18)	(3.21)	(3.45)
Level of Workload		-0.01	0.06	0.16		-0.72	0.01	-0.10
		(0.33)	(0.35)	(0.36)		(1.52)	(1.54)	(1.46)
Physical condition workplace		-0.03	0.03	-0.03		-2.78**	-2.66**	-2.60**
		(0.31)	(0.34)	(0.33)		(1.20)	(1.26)	(1.29)
Integration policy			0.22	0.16			-0.23	-0.40
			(0.54)	(0.51)			(2.02)	(2.04)
Health Security			0.05	-0.01			0.21	0.21
			(0.43)	(0.43)			(1.48)	(1.57)
Good Place to work			-0.50	-0.28			-1.41	-2.94
			(0.42)	(0.65)			(1.71)	(3.24)
Balance wotk/life			0.38	0.31			-1.23	-1.14
			(0.72)	(0.73)			(2.32)	(2.60)
Respect and Self esteem			-0.47	-0.27			-0.86	-0.97
			(0.93)	(0.93)			(3.12)	(3.16)
Job Satisfaction				-1.73*				3.40
				(0.92)				(3.41)
Pay Satisfaction				0.34				-0.45
				(0.32)				(1.20)
Organization Satisfaction				0.49				0.23
				(0.68)				(4.19)
Healthcare Program				0.01				0.41
				(0.10)				(0.38)
R-squared	.0025	.0026	.0028	.0035	.073	.097	.11	.11
N	6920	6920	6920	6920	445	445	445	445

Notes: The table presents four nested sets of possible predictors (1)-(4) of the bureau-year proximal death instrument. OLS estimates include bureau fixed effects. In addition, the table presents four nested sets of possible predictors (5)-(8) of the bureau-year-state relevant death instrument. OLS estimates include bureau-state fixed effects. The specification contains year fixed effects and Age, Education, Length of Service, Salary and WorkForce Gender Composition' interquantile ranges as controls. * $p < .1$, ** $p < .05$, *** $p < .01$

VII   Results

We begin the presentation of our results from Table 6 where we show the OLS estimates corresponding to equation (1). We first present all the results for time and cost performance, then in Table 9 we present those for the number of renegotiations. The first five columns in Table 6 display the results for *cost performance*, while the latter five report those for *time performance*. From these two sets of estimates, moving across columns from left to right entails an expansion in the set of controls included in the model specification.[20] To facilitate the interpretation of the estimates, both the outcomes and endogenous regressors are replaced throughout all the regressions by their z-scores, i.e. the variables have been rescaled to have a mean of zero and a standard deviation of one.

Table 6: OLS regressions Competence

	Cost Performance					Time Performance				
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Competence	-0.01	0.04***	0.04***	0.05***	0.05***	-0.01	0.02**	0.02**	0.03**	0.03***
	(0.02)	(0.01)	(0.01)	(0.02)	(0.02)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Bureau Experience			0.00	-0.02	-0.02			-0.05	-0.05	-0.05
			(0.03)	(0.03)	(0.03)			(0.04)	(0.04)	(0.04)
Bureau Size			0.03**	0.04**	0.04**			0.04**	0.03*	0.04*
			(0.02)	(0.02)	(0.02)			(0.02)	(0.02)	(0.02)
R-squared	122526	122526	122526	122526	122526	122526	122526	122526	122526	122526
Observations	.13	.14	.14	.14	.15	.11	.12	.12	.12	.12
Amount FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Duration FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Agency FEs	No	Yes	Yes	Yes	Yes	No	Yes	Yes	Yes	Yes
Year FEs	No	No	No	Yes	Yes	No	No	No	Yes	Yes
State FEs	No	No	No	No	Yes	No	No	No	No	Yes

Notes: Both contract outcomes and bureau characteristics are replaced by their standard scores. Standard errors are clustered by bureau and service category and are in parentheses. Amount FEs and Duration FES represent deciles for contract value and duration. * Significant at the 10 percent level; ** Significant at the 5 percent level; *** Significant at the 1 percent level.

In line with the descriptive evidence, a naive association between competence and performance (columns 1 and 6) results in an estimate that is negative (but close to zero) and not statistically significant. But the coefficient turns positive and significant as soon as additional controls are included. In particular, this is what happens in column 2 and 7 where we add agency fixed effects. This is not surprising given the very different nature of the

---

[20]The level of clustering is bureau and service category; the number of clusters is 2,073.

contracts that different agencies procure. Adding *Bureau Experience* and *Bureau Size* has, instead, no impacts on competence, thus confirming the difference between our measure of competence relative to these other proxies used in past studies. Finally, adding year and state fixed effects further increases the estimates' magnitude. Nevertheless, the magnitude remains economically small with a one standard deviation increase in competence amounting to an improvement in cost performance of 5 percent of a standard deviation (3 percent in the case of time performance).

Table 7: First stage and Reduced Form Regressions

	Cost Performance (RF)			Time Performance (RF)			Competence (FS)		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Proximal Deaths	-0.07***		-0.06***	-0.06***		-0.05***	-0.14***		-0.12***
	(0.01)		(0.01)	(0.01)		(0.01)	(0.02)		(0.02)
Relevant Deaths		-0.05*	-0.04		-0.07***	-0.06**		-0.20***	-0.18***
		(0.03)	(0.03)		(0.02)	(0.02)		(0.04)	(0.04)
Observations	122526	122526	122526	122526	122526	122526	122526	122526	122526
R-squared	.15	.15	.15	.12	.12	.12	.61	.61	.61

Notes: Columns 1-6 reports reduced-form regressions of *cost performance* and *time performance*, respectively, on the instruments. In Columns (7) to (9) we present the first stage for each IV regression from table 8. Both contract outcomes and bureau characteristics are replaced by their standard scores. Standard errors are clustered by bureau and service category and are in parentheses. All models include controls for contract features (cost plus format and solicitation procedure), buyer characteristics (experience and yearly procurement budget), fixed effects for service category, agency, deciles for contract value and duration, year, and U.S. state of performance. * Significant at the 10 percent level; ** Significant at the 5 percent level; *** significant at the 1 percent level.

Despite the inclusion of these controls, a concern with the potential downward bias in the OLS competence estimates remains. To address this concern, we implement an IV strategy based on the two instruments presented above. Table 7 reports the reduced-form and first-stage estimates. For the first-stage regressions, these estimates confirm what the visual IV showed in terms of a negative and significant effect of both instruments on competence. For the reduced form regression, the coefficients on both instruments tend to enter with a negative and significant effect, both when used individually and jointly. The exception being that for cost performance one of the two instruments - relevant deaths - is either only marginally significant when entered in isolation (column 2) or insignificant when entered jointly (column 3). The reduced form estimates are an interesting result on their own:

deaths of well paid white collars or managers negatively impact contractual performance. The impacts are similar for the two instruments and the two outcomes, which is not ex ante obvious given the different type of variation that the two instruments capture (one is across bureaus and the other across bureaus-states) and their low mutual correlation (15 percent). Crucial for the validity of our instruments is that it is only through competence that deaths affect procurement outcomes. In our context, this hinges on how employees interpret the wording of the FEVS question. In this regard, the specific nature of the question and its position within the survey at the end of the "my work unit" section make unambiguous that employees should here evaluate all elements affecting the proper functioning of their bureau. Thus, any effect that deaths might have should be captured by the answer to this question, guaranteeing that the exclusion restriction is satisfied. Standard statistical tests on the performance of these instruments are reported at the bottom of Table 8 where we report the IV estimates.

The first three columns of Table 8 report the results for *cost performance*, while the latter are for *time performance*. Across all columns, the set of controls is identical and corresponds to that of column 5 (and 10) of Table 6. For each outcome, the three estimates reported are obtained using first one instrument at the time and then both jointly. According to the baseline estimates with both instruments, one standard deviation increase in *competence* causes an increase of 0.38 and 0.37 standard deviation of *cost performance* and *time performance*, respectively. Compared to the OLS estimates of column 5 (and 10) of Table 6, the magnitude of all IV estimates is substantially larger, always exceeding the OLS 95% confidence interval.[21] Under the IV estimates, a one standard deviation increase in competence induces an increase in cost performance between one half and one fourth of a standard deviation (between one third and one half in the case of time performance).

The estimates remain quite similar between cost and time performance. Interestingly, despite the two instruments having a relatively low mutual correlation (15%), the estimates

---

[21]Building on the earlier discussion on the limited extent of reverse causality bias in the OLS estimates, the fact that the IV estimates exceed the OLS ones also indicates that the source of upward bias, if any, is less relevant than that of downward bias. Nevertheless, it is also worth noticing that the possibility of reverse causality means that an IV approach is preferable to a different approach based on first regressing performance on bureau fixed effects and, subsequently, regressing these fixed effects on bureau competence.

in column 1 and 2 are statistically identical. This is suggestive of these estimates plausibly representing an average treatment effect and not a LATE. Indeed, IV estimates differing when using different instruments, is an indication of heterogeneous treatment effects due to different compliers associated with the instruments (Angrist, Imbens and Rubin [1996]). Possible compliers in our setting are bureaus increasing or decreasing competence if and only if they experience some deaths; this is unlikely because accurate recruiting, attention to the training of personnel, and other human capital policies result in very standardized practices across federal bureaus.

Table 8: IV regressions

	Cost Performance			Time Performance		
	(1)	(2)	(3)	(4)	(5)	(6)
Competence	0.49***	0.26*	0.38***	0.41***	0.32**	0.37***
	(0.11)	(0.15)	(0.10)	(0.11)	(0.14)	(0.09)
Bureau Experience	-0.05	-0.03	-0.04	-0.07*	-0.07	-0.07
	(0.03)	(0.03)	(0.03)	(0.04)	(0.04)	(0.04)
Bureau Size	-0.01	0.02	0.01	-0.00	0.01	0.00
	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
Centered R-squared	.07	.13	.11	.07	.09	.08
Observations	122526	122526	122526	122526	122526	122526
Weak Id. F-Test	39.64	21.76	29.38	39.64	21.76	29.38
Underid. F-Test	40.21	18.65	54.2	40.21	18.65	54.2
Overid. F-Test	0	0	1.71	0	0	.29
Amount FEs	Yes	Yes	Yes	Yes	Yes	Yes
Duration FEs	Yes	Yes	Yes	Yes	Yes	Yes
Agency FEs	Yes	Yes	Yes	Yes	Yes	Yes
Year FEs	Yes	Yes	Yes	Yes	Yes	Yes
State FEs	Yes	Yes	Yes	Yes	Yes	Yes

Notes: Instruments are: *Proximal Deaths* and *Relevant Deaths*. Columns (1) and (4) report IV with *Proximal Deaths*; columns (2) and (5) report IV with *Relevant Deaths*; columns (3) and (6) report IV with both *Proximal Deaths* and *Relevant Deaths*. Both contract outcomes and bureau characteristics are replaced by their standard scores. Standard errors are clustered by bureau and service category and are in parentheses. All models include controls for contract features (cost plus format and solicitation procedure), buyer characteristics (experience and yearly procurement budget), fixed effects for service category, agency, deciles for contract value and duration, year, and U.S. state of performance. * Significant at the 10 percent level; ** Significant at the 5 percent level; *** Significant at the 1 percent level. The weak identification test employed is that of Pflueger and Montiel Olea [2013]. The underidentification test is an LM test of whether the equation is identified, i.e., that the excluded instruments are relevant, meaning correlated with the endogenous regressors. The test is essentially the test of the rank of a matrix: under the null hypothesis that the equation is underidentified, the matrix of reduced form coefficients on the L1 excluded instruments has rank=K1-1 where K1 is the number of endogenous regressors. Under the null, the statistic is distributed as chi-squared with degrees of freedom equal to (L1-K1+1). A rejection of the null indicates that the matrix is full column rank (model is identified). The Sargan statistic is calculated as N*R-squared from a regression of the IV residuals on the full set of instruments.

To offer a more transparent economic interpretation of the estimates, we can then consider what would happen if we were to use them to infer the effect of lifting the level of *competence* from all bureaus to that of the bureaus at the 90th percentile of this distribution. This implies a reduction in cost overruns of $54,679 on average per contract, or around $6.7 billions in total across all contracts in the dataset. Moreover, this would imply a saving of 39.2 days in effective execution time, corresponding to 4.8 million days across all the contracts in the dataset. The amounts are economically sizable and compare well to what the literature has indicated could be achieved by optimizing either the incentives given to suppliers (for instance through the choice between cost plus and fixed price contracts) or the type of awarding procedures (for instance through the selection of negotiations versus competitive auctions).

Table 9: Number of Renegotiations - IV Estimates

	# Time Reneg.			# Cost Reneg.			# Tot. Reneg.		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Competence	-0.71**	-0.28	-0.50**	-0.81**	-0.84**	-0.83***	-1.53***	-1.12*	-1.33***
	(0.29)	(0.26)	(0.22)	(0.32)	(0.39)	(0.28)	(0.58)	(0.62)	(0.47)
Centered R-squared	.1	.11	.11	.1	.1	.1	.1	.11	.11
Observations	122526	122526	122526	122526	122526	122526	122526	122526	122526
Weak Id. F-Test	39.64	21.76	29.38	39.64	21.76	29.38	39.64	21.76	29.38
Underid. F-Test	40.21	18.65	54.2	40.21	18.65	54.2	40.21	18.65	54.2
Overid. F-Test	0	0	1.56	0	0	0	0	0	.3

Notes: IV models of columns (1) to (3) of table 8 are replicated with the number of time renegotiations (columns 1 to 3), the number of cost renegotiations (columns 4 to 6), and the total number of renegotiations (columns 7 to 9) as substitutes for cost performance and time performance. No of time renegotiations stands for the number of contract modifications related to an amendment of the final contract duration; No. of cost renegotiations is instead related to the number of amendments of contract price. Standard errors are clustered by bureau and service category and are in parentheses. All models include controls for contract features (cost plus format and solicitation procedure), buyer characteristics (experience and yearly procurement budget), fixed effects for service category, agency, deciles for contract value and duration, year, and U.S. state of performance. * Significant at the 10 percent level; ** Significant at the 5 percent level; *** significant at the 1 percent level.

Before exploring the robustness of these estimates, we present analogous IV estimates for three different outcomes measuring the number of renegotiations. In Table 9, the first three columns use as outcome the number of times that the end date of the contract was modified, the next three columns regard the number of times the final cost was modified and the latter three regard the total number of times either the completion time or cost was modified. For each outcome, the three estimates reported are obtained with the same model specification

of Table 8 and using first one instrument at the time and then both jointly. The main finding is that, despite some differences in magnitudes and significance, we observe a consistently negative effect of competence on the number of renegotiations. One standard deviation increase in competence causes 0.5 (39%) and 0.8 (71%) fewer cost renegotiations and time renegotiations, respectively (1.3 - 55% - fewer in total). Since each negotiation episode is likely to be associated with some waste - transaction costs -, this additional evidence strongly supports the main takeaway from this study: enhancing bureau competence can significantly improve the effectiveness of public procurement even in a developed country like the US.

We conclude this section with a brief exploration of robustness of the estimates by focussing on time and cost performance. A more in depth analysis is offered in the web appendix. We focus on four types of concerns: i) the econometric model formulation, ii) the sample selection, iii) the measurement of both performance measures and competence components, and iv) inference. First, we assess the robustness of the 2SLS estimates presented above to two alternative estimation approaches: i) a limited information maximum likelihood (LIML) estimator, and ii) a fractional probit within a control function method. As is standard practice, the former is used to verify that indeed we do not face a situation of weak instruments. The latter, instead, is used to check, through the Wooldridge [2002] method, that the particular shape of the distribution of the performance measures (bounded between zero and one and with a mass point at one) is not compromising the analysis. In both cases, the estimates obtained (reported in Table A.1 and A.2 in the appendix) are very close to the baseline IV estimates presented above.

Second, we consider the robustness to alternative subsamples. Table 10 reports the IV obtained when estimating on different subsamples models analogous to those in columns 3 of Table 8. For convenience, we report in the first column of each panel the baseline estimate from Table 8. In the second and third column, we restrict the sample to exclude the most extreme observations with cost and time performance lower than 0.1 and 0.25, respectively.[22] While the magnitude of the estimates declines with a more stringent definition of the sample, the signs and significance are broadly in line with those in the baseline estimates. In columns

---

[22]According to our index of contract performance, the values of 0.1 and 0.25 represent overruns of 9 and 4 times the expected outcome, respectively.

4-6, we iteratively discard one-per-time all the observations belonging to the three largest agencies: DOD, DVA, and GSA, respectively. All estimates appear robust. Column 7 restricts the sample to tenders receiving at least two offers. Finally, in Column 8 we present results when agency fixed effects are replaced by bureau fixed effects. By subsampling contracts we still find that greater competence has a strong, positive effect, in line with our baseline estimates. In the latter case, instead, as we discussed the high degree of persistency of competence over time, coefficients are consistently non-significant.

Table 10: ROBUSTNESS CHECKS: SAMPLE SELECTION

Panel A: Competence - Cost Performance

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Q28	0.38***	0.33***	0.15**	0.29***	0.31***	0.35***	0.35***	0.02
	(0.10)	(0.09)	(0.06)	(0.10)	(0.10)	(0.09)	(0.11)	(0.13)
Observations	122526	117274	102061	96846	89382	113242	74711	122526
Centered R-squared	.11	.11	.11	.11	.14	.11	.15	.16
Weak Id. F-Test	29.38	28.46	27.53	23.79	30.88	30.17	24.85	23.58
Underid. F-Test	54.2	52.93	52	42.53	56.97	55.59	49.01	52.77
Overid. F-Test	1.71	2.46	3.74	.54	1.13	.86	1.02	15.31

Panel B: Competence - Time Performance

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Q28	0.37***	0.35***	0.17***	0.22**	0.25***	0.34***	0.33***	0.22
	(0.09)	(0.08)	(0.06)	(0.09)	(0.09)	(0.09)	(0.11)	(0.14)
Observations	122526	117274	102061	96846	89382	113242	74711	122526
Centered R-squared	.08	.07	.11	.13	.13	.09	.12	.13
Weak Id. F-Test	29.38	28.46	27.53	23.79	30.88	30.17	24.85	23.58
Underid. F-Test	54.2	52.93	52	42.53	56.97	55.59	49.01	52.77
Overid. F-Test	.29	.42	.78	.4	.07	0	.14	4.62

Notes: Column 1 reports the baseline estimate for *Cooperation* or *PCA Cooperation* from table 8. In Column 2 and 3 we do not consider contracts associated to cost and time performance lower than 0.1 and 0.25, respectively. In columns 4,5, and 6, we discard DOD, DVA, and GSA, respectively. Column 7 restricts the sample to tenders receiving at least two offers. Finally, in Column 8 we present results when agency fixed effects are replaced by bureau fixed effects. Both contract outcomes and bureau characteristics are replaced by their standard scores. Standard errors are clustered by bureau and service category and are in parentheses. All models include controls for contract features (cost plus format and solicitation procedure), buyer characteristics (experience and yearly procurement budget), fixed effects for service category, agency, deciles for contract value and duration, year, and U.S. state of performance. * Significant at the 10 percent level; ** Significant at the 5 percent level; *** significant at the 1 percent level.

The third set of robustness checks involves alternative measures of procurement performance. Contracts amendment records in the FPDS data are classified according to the reason for contract modification, which can be *in-scope* or *out-of-scope* revisions.[23] In line

---

[23]See footnote 12.

with other studies, we have considered *in-scope* amendments only.[24] Kang and Miller [2017] have recently proposed a different measure of renegotiations by excluding some *in-scope* revisions, but also retaining some of the *out-of-scope* revisions. When we follow this alternative definition (see Table A.3), we find similar results to those in our baseline estimates.

The fourth and final robustness check involves inference. Recent research by Young, Alwyn [2017], indicates that IV studies sometimes have inference problems driven by the finite sample estimator performance. Our large sample size limits this problems but, nevertheless, in the appendix we report standard estimates obtained via bootstrap. We replicate our IV analysis by drawing 500 bootstrap samples in a fashion consistent with the error dependence within our cluster of observations (bureau and service category) and independence across observations. The findings confirm the baseline estimates presented above.

VIII   Channels: Cooperation, Skills or Incentives?

The FEVS data contains several questions that might help to disentangle what forces are behind the effects of competence on procurement. Table 11 reports the full list of questions composing the "my work unit" section. The one at the bottom of the table (Q28) is the one we used so far, i.e. *competence*. The eight questions that precede it cover several aspects of the bureau characteristics that we group into three categories: cooperation (two questions), incentives (four questions) and skills (two questions). Understanding to what extent these three channels contribute to explain our earlier findings is important in order to design the right policies to improve bureau competence and, through that, procurement outcomes.

Causally identifying the individual contribution of each channel would require instruments, or other sources of variation, separately moving each of them. Instead of attempting this route, we follow a more descriptive approach based on two strategies. First, we illus-

---

[24]Before initiating a modification, the contracting officer must determine if the proposed effort is within the scope of the existing contract or is a new acquisition outside of the scope. A new requirement outside of the scope of the existing contract must be processed as a new acquisition. Contract scope means, in simple terms, that the contemplated change must be generally related to the work originally contracted for. If a contract was awarded for the design (and only the design) of an automated information system, it could not be later modified to have the contractor provide and install hardware.

Table 11: List of FEVS Questions Composing the "My Work Unit" Section

Q#	Question	Classification	PCA Skill/Incentive Factor 1 Weights	PCA Cooperation Factor 2 Weights
My Work Unit:				
20	The people I work with cooperate to get the job done.	Cooperation	0.02	0.36
21	My work unit is able to recruit people with the right skills.	Skills	0.16	0.01
22	Promotions in my work unit are based on merit.	Incentives	0.16	0.07
23	In my work unit, steps are taken to deal with a poor performer who cannot or will not improve.	Incentives	0.15	0.09
24	In my work unit, differences in performance are recognized in a meaningful way.	Incentives	0.19	0.07
25	Awards in my work unit depend on how well employees perform their jobs.	Incentives	0.15	0.10
26	Employees in my work unit share job knowledge with each other.	Cooperation	0.03	0.22
27	The skill level in my work unit has improved in the past year.	Skills	0.14	0.07
28	How would you rate the overall quality of work done by your work unit?	Competence	-	-

Notes: The complete set of nine questions in the FEVS section dedicated to the employees' assessment of their work unit. The numbering in column one reflects that in the FEVS. The last two columns report the percentage contributions that each variable assumes through the weights calculated by the factor analysis.

trate how - purely within the FEVS data - competence correlates with cooperation, skills and incentives. Here we use Q20, Q21 and Q23 to measure *cooperation*, *skills*, and *incentives*, respectively. The wording of these questions is unambiguous and their correlation with competence in the regressions described next is stronger than that of the remaining questions.[25] The first four columns of Table 12 show OLS estimates obtained by regressing bureau competence on the three components, first separately and then jointly after collapsing the observations at bureau and year level. This gives us a first, clear indication of the extent to which the three different channels contribute to explain the variability of competence across bureaus. Cooperation is evidently a key driver of competence: when entered by itself the R^2 is 0.83 and the coefficient is close to one. The corresponding figures are smaller for incentives (0.68 and 0.55 respectively) and for skills (0.55 and 0.40). Indeed, when entered jointly in column 4, both the coefficient on cooperation and the regression's R^2 remain close to those in column 1, while the coefficients of both incentives and skills drop substantially relative to columns 2 and 3.

Second, we replicate the OLS regressions of Table 6 using the three channels instead of competence. Thus we regress time and cost performance on the competence channels

---

[25]To limit the arbitrariness of this choice, in the appendix we report results using the other questions and also results using the whole set of eight questions through a principal components analysis. The analysis reveals that two factors are sufficient to explain 84 percent of the common variance among cooperation, skills and incentives. The last two columns of Table 11 reports these weights. The first factor has essentially a 5% contribution of the two questions involving cooperation (Q20 and Q26) and nearly an equal contribution of all the remaining six questions. The second factor, instead, gives 56% of the weight to the two cooperation questions. These factors explain 47% of the total variance each and are also strongly correlated with competence (Q28).

Table 12: Cooperation, Skills and Incentives - OLS Estimates

	Competence				Cost Performance		Time Performance	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Cooperation	1.08***			0.94***	0.08***	0.11***	0.05***	0.06***
	(0.02)			(0.03)	(0.02)	(0.03)	(0.02)	(0.02)
Incentive		0.68***		0.09***		0.00		0.05***
		(0.03)		(0.03)		(0.02)		(0.02)
Skill			0.55***	0.06**		-0.04**		-0.08***
			(0.03)	(0.03)		(0.02)		(0.02)
Observations	441	441	441	441	122526	122526	122526	122526
R-squared	.83	.40	.55	.84	.15	.15	.12	.13
Amount FEs	No	No	No	No	Yes	Yes	Yes	Yes
Duration FEs	No	No	No	No	Yes	Yes	Yes	Yes
Agency FEs	No	No	No	No	Yes	Yes	Yes	Yes
Year FEs	No	No	No	No	Yes	Yes	Yes	Yes
State FEs	No	No	No	No	Yes	Yes	Yes	Yes

Notes: The FEVS data is the sample used for the estimates in the first four columns. The depended variable is competence, while the regressors are cooperation (Q20), skills (Q21) and incentives (Q23). In the following four columns, the sample is our baseline sample, obtained by combining FPDS and FEVS data. The dependent variables are time and cost performance. The model specification is identical to that in column 5 Table 6, but for the subsitution of competence with its three components, as detailed in the table.

(and the other covariates as in column 5 Table 6). The results are reported in the latter columns of Table 12. Given the prominence of cooperation, we first enter this variable alone (columns 5 and 7) and then jointly with incentives and skills (columns 6 and 8). The estimates for cooperation are always positive and significant. Their magnitude, especially for time performance, is also rather close to that of competence in Table 6. The evidence is more mixed on the effect of incentives and skills: conditioning on cooperation, the marginal effect of the former is estimated to be zero for cost competence and positive and significant for time performance, while the marginal effect of the latter is negative and significant for both performance measures. In the appendix, additional estimates using different FEVS variables, as well as their principal components, to measure the three channels confirm the main qualitative finding of cooperation being a key driver of competence.

IX Conclusions

Our paper represents the first comprehensive study of the impact of bureaucratic competence on public procurement outcomes for works and services. By combining three large datasets

on U.S. federal bureaus purchases, their internal functioning and workforce characteristics, we quantify the effects of bureaus' competence on the time and cost performance of public contracts, and on the number of times they are renegotiated as a proxy of haggling costs. Our identification strategy exploits the exogeneity of death events involving public officials to allow for a causal interpretation of bureau competence on procurement performance.

Our main result lies in quantifying the effects of competence heterogeneity across US federal bureaus on their procurement performance. The size of these effects would be expected in a weak institutions environment, but are rather surprising in our view for the country with arguably the world's most efficient public (and private) management practices. They suggest that even in advanced countries, there is considerable scope for improving public service provision by improving competence in public bureaucracies.

Our second main result, to be taken somewhat more cautiously in terms of causal interpretation, is that cooperation in the bureau seems to be by far the most important component of bureau competence in terms of the effects on procurement performance. This second result is, in our view, linked to the complexity and multidisciplinarity typical of procurement. The need to master legal, engineering, economic/strategic and merceological skills for different types of goods, works and services and to coordinate the various phases of the procurement cycle (market analysis, tender design and implementation, contract management and evaluation) makes good procurement primarily the outcome of team-work. Cooperation among employees is therefore a crucial ingredient for a well functioning procurement office.

We see several avenues for further research. First, given the crucial role we have identified for competence, it would be important to develop a deeper understanding of what factors can promote this trait within public offices, especially with regard to the ability to maintain cooperation among employees. Although a detailed exploration of this issue is beyond the scope of this paper, our data are indicative of the key role played by young managers. To further explore this aspect, we report in Figure 6 plots of how our baseline estimates would differ with instruments constructed by altering the definition of the *relevant deaths* instrument. In the baseline estimates, the median values of age and salary are the cut offs used to select relatively low age and high salary employees. In Figure 6 we report the IV

estimates interactively replacing the *relevant deaths* instrument with an analogue dummy variable constructed for different sets of employees: those that are either above or below the median salary, and then for each of these two subgroups we report all possible age cutoffs in the IV construction. The results in the figure indicate that for all age cutoffs up until the age of 50, deaths of workers with higher than median salary produce estimated effects of competence on performance that are statistically larger than the corresponding ones for below median salary workers. Above age 50, the estimates become statistically identical. This evidence is further corroborated by a full heterogeneity analysis presented in the appendix where we explore all combinations of age and salary. This analysis is indicative of interesting heterogenous effects across employees that might even offer a simple policy prescription to help low-performing bureaus to improve: infuse relatively young, competent and well paid managers.[26] A similar policy prescription is offered by [Bertrand et al., 2016], although for a rather different type of country.

Figure 6: Heterogeneity of IV Estimates for Competence
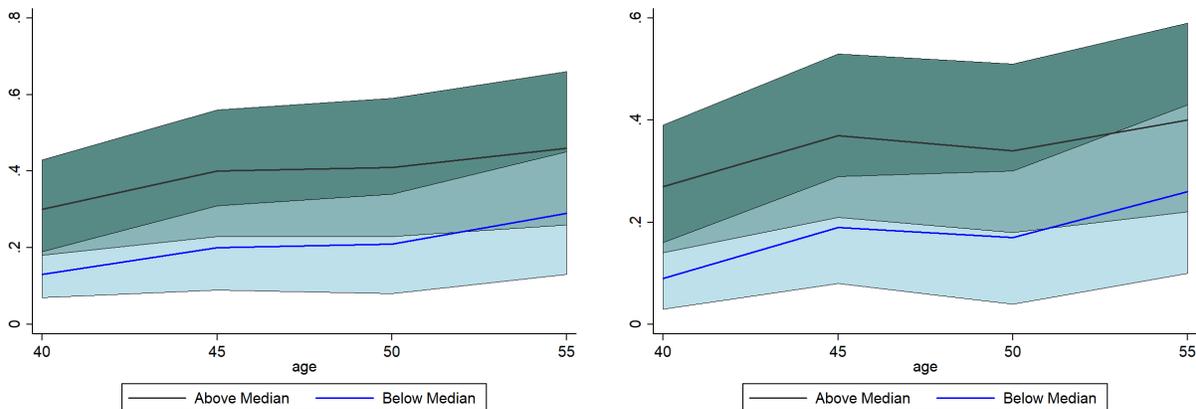
(a) Cost Performance: Competence

(b) Time Performance: Competence

Notes: IV estimates of the effects of competence on cost performance (panel a) and time performance (panel b). The model specification is the same of the model 4 in Table 8. The only difference relative to that model is that the *relevant deaths* instrument is replaced with an analogue dummy variable constructed for different sets of workers: workers that are either above or below the median salary, and then for each of these two subgroups we report all possible age cutoffs in the IV construction.

---

[26]This type of analysis thus also allow us to relate our estimates to an important strand of the literature that tries to quantify how specific groups of employees affect outcomes of the teams/units/firms they work for (see Jäger [2017] for a recent study in this area).

References

Acemoglu, Daron, Camilo Garcia-Jimeno, and James A. Robinson. 2015. "State Capacity and Economic Development: A Network Approach." *The American Economic Review*, 105(8): 2364–2409.

Angrist, Joshua D., Guido W. Imbens, and Donald B. Rubin. 1996. "Identification of Causal Effects Using Instrumental Variables." *Journal of the American Statistical Association*, 91(434): 444–455.

Athey, Susan, Dominic Coey, and Jonathan Levin. 2013. "Set-asides and subsidies in auctions." *American Economic Journal: Microeconomics*, 5(1): 1–27.

Azoulay, Pierre, Joshua S. Graff Zivin, and Bhaven N. Sampat. 2011. "The Diffusion of Scientific Knowledge Across Time and Space: Evidence from Professional Transitions for the Superstars of Medicine." *NBER Working Paper Series*, 16683.

Bai, Ying, and Ruixue Jia. 2016. "Elite Recruitment and Political Stability: The Impact of the Abolition of China's Civil Service Exam System." *Econometrica*, 84(2): 677–733.

Bajari, Patrick, and Gregory Lewis. 2011. "Procurement with Time Incentives: Theory and Evidence (with Patrick Bajari)." *Quarterly Journal of Economics.*

Bajari, Patrick, and Steven Tadelis. 2001. "Incentives versus Transaction Costs: A Theory of Procurement Contracts." *The RAND Journal of Economics Journal of Economics*, 32(3): 387–407.

Bajari, Patrick, Robert McMillan, and Steven Tadelis. 2009. "Auctions versus negotiations in procurement: An empirical analysis." *Journal of Law, Economics, and Organization*, 25(2): 372–399.

Bajari, Patrick, Stephanie Houghton, and Steven Tadelis. 2014. "Bidding for Incomplete Contracts: An Empirical Analysis of Adaptation Costs." *American Economic Review*, 104(4): 1288–1319.

Bandiera, Oriana, Andrea Prat, and Tommaso Valletti. 2009. "Active and passive waste in government spending: Evidence from a policy experiment." *American Economic Review*, 99(4): 1278–1308.

Banerjee, Abhijit, and Esther Duflo. 2000. "Reputation Effects and the Limits of Contracting: A Study of the Indian Software Industry." *Quarterly Journal of Economics*, 115(3): 989–1017.

Becker, Sascha O., and Hans K. Hvide. 2013. "Do Entrepreneurs Matter?" , (4088).

Bennedsen, Morten, Francisco Pérez-gonzález, and Daniel Wolfenzon. Forthcoming. "Do CEOs matter: Evidence from CEO Hospitalization Events." *Journal of Finance.*

Bertrand, Marianne, Robin Burgess, Arunish Chawla, and Guo Xu. 2016. "The Costs of Bureaucratic Rigidity: Evidence from the Indian Administrative Service." *mimeo.*

Besley, Timothy, and Torsten Persson. 2009. "The Origins of State Capacity: Property Rights." *American Economic Review*, 99(4): 1218–1244.

Besley, Timothy, and Torsten Persson. 2010. "State Capacity, Conflict, and Development." *Econometrica*, 78(1): 1–34.

Best, Michael Carlos, Jonas Hjort, and David Szakonyi. 2017. "Individuals and Organizations as Sources of State Effectiveness, and Consequences for Policy." *National Bureau of Economic Research Working Paper 23350.*

Blader, Steven, Claudine Madras Gartenberg, and Andrea Prat. 2016. "The Contingent Effect of Management Practices." *CEPR Discussion Paper*, DP11057.

Bloom, Nicholas, Renata Lemos, Raffaella Sadun, and John Van Reenen. 2015. "Does Management Matter in Schools?" *Economic Journal*, 125(584): 647–674.

Bloom, Nicholas, Renata Lemos, Raffaella Sadun, Daniela Scur, and John Van Reenen. 2014. "The New Empirical Economics of Management." *Journal of the European Economic Association*, 12(4): 835–876.

Bruce, Joshua R., John M. De Figueiredo, and Brian S. Silverman. 2017. "Public Contracting for Private Innovation : Government Capabilities , Decision Rights , and Performance Outcomes." *mimeo.*

Bucciol, Alessandro, Riccardo Camboni, and Paola Valbonesi. 2017. "Buyers' Ability in Public Procurement: A Structural Analysis of Italian Medical Devices." *Working Paper.*

Chassang, Sylvain, and Juan Ortner. 2017. "Collusion in Auctions with Constrained Bids: Theory and Evidence from Public Procurement." Princeton University, Department of Economics, Econometric Research Program. Working Papers.

Chong, Eshien, Stephane Saussier, and Brian S. Silverman. 2015. "Water Under the Bridge: Determinants of Franchise Renewal in Water Provision." *Journal of Law, Economics and Organization*, 31(1): 3–39.

Coviello, Decio, and Mario Mariniello. 2014. "Publicity Requirements in Public Procurement: Evidence from a Regression Discontinuity Design." *Journal of Public Economics*, 109: 76–100.

Coviello, Decio, Andrea Guglielmo, and Giancarlo Spagnolo. 2017. "The Effect of Discretion on Procurement Performance." *Management Science*, 63(5).

Dal Bo, Ernesto, Frederico Finan, and Martin Rossi. 2013. "Strengthening State Capabilities: The Role of Financial Incentives in the Call to Public Service." *Quarterly Journal of Economics*, 128(3): 1169–1218.

Decarolis, Francesco. 2014. "Awarding Price, Contract Performance, and Bids Screening: Evidence from procurement auctions." *American Economic Journal: Applied Economics*, 6(1): 108–132.

Decarolis, Francesco, Giancarlo Spagnolo, and Riccardo Pacini. 2016. "Past Performance and Procurement Outcomes." National Bureau of Economic Research Working Paper 22814.

De Silva, Dakshina G., Timothy Dunne, Georgia Kosmopoulou, and Carlos Lamarche. 2017. "Contract Modification and Bidding in Highway Procurement Auctions." *mimeo.*

De Soto, Hernando. 1990. *The Other Path.* New York, NY: Harper and Row.

Djankov, Simeon, Rafael La Porta, Florencio Lopez-de Silanes, and Andrei Shleifer. 2002. "The Regulation of Entry." *The Quarterly Journal of Economics*, 117(1): 1–37.

Fernandez, Sergio, William G. Resh, Tima Moldogaziev, and Zachary W. Oberfield. 2015. "Assessing the Past and Promise of the Federal Employee Viewpoint Survey for Public Management Research: A Research Synthesis." *Public Administration Review*, 75(3): 382–394.

Finan, Frederico, Benjamin A. Olken, and Rohini Pande. 2015. "The personnel economics of the state." *NBER Working Paper Series*, 21825.

Giuffrida, Leonardo M., and Gabriele Rovigatti. 2017. "Can the Private Sector Ensure the Public Interest? Evidence from Federal Procurement." *mimeo.*

Guasch, J. Luis, Jean-Jacques Laffont, and Stephane Straub. 2008. "Renegotiation of concession contracts in Latin America: Evidence from the water and transport sectors." *International Journal of Industrial Organization*, 22(7): 1267–1294.

Iimi, Atsushi. 2013. "Testing Low-Balling Strategy in Rural Road Procurement." *Review of Industrial Organization*, 43(3): 243–261.

Jäger, Simon. 2017. "How Substitutable Are Workers? Evidence from Worker Deaths." *mimeo.*

Jones, Benjamin F, and Benjamin A Olken. 2005. "Do Leaders Matter? National Leadership and Growth Since World War II." *Quarterly Journal of Economics*, 120(3).

Kang, Karam, and Robert. A. Miller. 2017. "Evaluating Discretion in Government Procurement." *mimeo.*

Krasnokutskaya, Elena, and Katja Seim. 2011. "Bid preference programs and participation in highway procurement auctions." *American Economic Review*, 101(6): 2653–2686.

Lewis-Faupel, Sean, Yusuf Neggers, Benjamin A. Olken, and Rohini Pande. 2016. "Can Electronic Procurement Improve Infrastructure Provision? Evidence from Public Works in India and Indonesia." *American Economic Journal: Econ. Policy*, 8(3): 258–83.

Lewis, Gregory, and Patrick Bajari. 2011. "Procurement with Time Incentives: Theory and Evidence." *Quarterly Journal of Economics*, 126(3): 1173–1211.

Lewis, Gregory, and Patrick Bajari. 2014. "Moral hazard, Incentive Contracts, and Risk: Evidence from Procurement." *The Review of Economic Studies*, 81(3): 1201–1228.

Liebman, Jeffrey B., and Neale Mahoney. 2017. "Do Expiring Budgets Lead to Wasteful Year-End Spending? Evidence from Federal Procurement." *American Economic Review*, 107(11): 3510–49.

MacKay, Alexander. 2017. "The Structure of Costs and the Duration of Supplier Relationships." *mimeo.*

Manelli, Alejandro M, and Daniel R Vincent. 1995. "Optimal Procurement Mechanisms." *Econometrica*, 63(3): 591–620.

Marion, Justin. 2007. "Are bid preferences benign? The effect of small business subsidies in highway procurement auctions." *Journal of Public Economics*, 91(7-8): 1591–1624.

Mohamed, Khaled A., Shafik S. Khoury, and Sherif M. Hafez. 2011. "Contractors decision for bid profit reduction within opportunistic bidding behavior of claims recovery." *International Journal of Project Management*, 29(1): 93 – 107.

Papke, Leslie E., and Jeffrey M. Wooldridge. 2008. "Panel data methods for fractional response variables with an application to test pass rates." *Journal of Econometrics*, 145(12): 121 – 133.

Pflueger, Carolin, and Jose Luis Montiel Olea. 2013. "A Robust Test for Weak Instruments." *Journal of Business and Economic Statistics*, 31(3): 358–369.

Rasul, Imran, and Daniel Rogger. 2016. "Management of Bureaucrats and Public Service Delivery: Evidence from the Nigerian Civil Service." *Economic Journal*, 44: 1–40.

Saussier, Stephane, and Jean Tirole. 2015. "Strengthening the Efficiency of Public Procurement." *Counseil d'Analyse Economique*, 22.

Spiller, Pablo T. 2008. "An Institutional Theory of Public Contracts: Regulatory Implications." National Bureau of Economic Research Working Paper 14152.

Spulber, Daniel F. 1990. "Auctions and Contract Enforcement." *Journal of Law, Economics and Organization*, 6(2): 325–44.

Tadelis, Steven. 2002. "Complexity, Flexibility, and the Make-or-Buy Decision." *The American Economic Review*, 92(2): 433–437.

Warren, Patrick L. 2014. "Contracting Officer Workload, Incomplete Contracting, and Contractual Terms." *The RAND Journal of Economics*, 45(2): 395–421.

Williamson, Oliver E. 1971. "The Vertical Integration of Production: Market Failure Considerations." *The American Economic Review*, 61(2): 112–123.

Wooldridge, Jeffrey M. 2002. *Econometric analysis of cross section and panel data.* Cambridge and London:MIT Press.

Young, Alwyn. 2017. "Consistency without Inference: Instrumental Variables in Practical Application." *Working Paper.*

For Publication on the Authors' Web Pages

Web Appendix

I   Sample Selection

For the purpose of our analysis, we will focus on the years where the FEVS has an yearly frequency and where the two datasets, FEVS and FPDS, overlap. Thus, we focus on the years between 2010 and 2015. Although the data contain contracts for both supplies, R&D projects, services, and works, the first two are ruled out of the analysis. Supplies typically do not exhibit any ex post variation in price or delivery time, while the outcome of R&D contracts cannot be reasonably assessed in terms of costs and duration. Thus, for our analysis we focus exclusively on the procurement of services and works and refer jointly to these as *services*.[27] We restrict our sample to those contracts awarded through competitive solicitations because the effect of the treatments would otherwise not be observable. We consider as competitive a lot for which the extent of competition is labelled "Full and open"; those whose participation is not set aside to any specific group of firms; those at or below the micro-purchase acquisition threshold - $3,000 - as allocated without soliciting competitive quotations. FPDS contains every base contract that exceeds an individual transaction value of $2,500. We focus on contracts worth more than $25,000.[28] In non-competitive awardings, the participation criteria restrict the competition *ex-ante* to dimensions other than quality.

---

[27]Services included in the sample are: special studies/analysis, not R&D; architect and engineering services; information technology and telecommunications; purchase of structures/facilities; natural resources management; social; quality control, testing, and inspection; maintenance, repair, and rebuilding of equipment; modification of equipment; technical representative; operation of structures/facilities; installation of equipment; salvage; medical; support (professional/administrative/management); utilities and housekeeping; photo/map/print/publication; education/training; transportation/travel/relocation. Works include: construction, maintenance, repair, alteration of structures/facilities.

[28]Above this cutoff it is safe for us to include all contracts awarded by federal bureaus. Indeed, according to the FAR subpart 4.6, each executive agency must establish and maintain for a period of 5 years a computer file, by fiscal year, containing unclassified records of all procurements exceeding $25,000. This file shall be accessible to the public using FPDS. Purchases over $25,000 are also publicized on Federal Business Opportunities website. On this website, you will find Requests for Proposals (RFPs) for practically everything the government purchases.

Figure A.1: Federal Employees by State

Notes: Intersection between bureau (columns) and state (rows) are filled with X when, across our sample, at least a worker within the bureau is settled in the state.

For similar reasons, we focus on contracts whose tasks are such that the vendor can influence the outcome metrics through effort. Supply contracts do not allow for renegotiations. Hence, for these contracts our measure of performance does not proxy for outcome quality and we exclude them from the analysis.[29] The same applies to the subcategory "Lease or Rental of Equipment, Structures, or Facilities". We consider only contracts awarded within the U.S. border. Finally, the sample includes only contracts awarded in states where the awarding bureau has at least one employee. This restriction leads us to drop 4% of the sample, but serves to insure that we can match the locations of the bureaus, local offices and of the contracts that they are likely to supervise. Figure A.1 reports the location of each bureau by indicating with an "X" the state in which they employ at least one white-collar worker.

II   Robustness Checks

This section reports the results for the robustness checks summarized in section VI.

- Table A.1: LIML estimates. As is standard for checking for weak instruments, LIML estimates are provided as a robustness for the 2SLS estimates presented in the main text. All the point estimates are very close to those in Table 8, thus limiting concerns about a weak instruments problem.

- Table A.2: control function estimates. Since $performance_{ijt}^C$ and $performance_{ijt}^T$ are fractional variables on $(0,1]$[30] with major spikes in their density at 1, we follow Wooldridge [2002] by employing the fractional probit regression and specifying conditional means as a probit function $\mathbb{E}\left(y|\mathbf{x}\right) = \Phi\left(\mathbf{x}\gamma\right)$.[31] This fractional probit model handles continuous endogenous explanatory through a *two-step control function approach*. The control function approach relies on similar identification conditions of the

---

[29]The typical supply contract shows a 0 value in *extra time/cost* and a unit value in both performances.

[30]In this case, the outcome variables are not standardized

[31]Papke and Wooldridge [2008] and Wooldridge [2002] show that the population model $\mathbb{E}\left(y|\mathbf{x}\right) = x_1\gamma_1 + x_2\gamma_2 + ... + x_J\gamma_J = \mathbf{x}\gamma$, when y is fractional, rarely provides the best description of $\mathbb{E}\left(y|\mathbf{x}\right)$. Indeed, with $y \in (0,1]$ the effect of any particular explanatory variable is usually not constant throughout the range of $\mathbf{x}$.

linear IV described in the main text.[32] Table A.2 presents the estimates obtained via control function, using the same four instruments used for the main analysis. All the qualitative implications described for our baseline estimates are confirmed. The significance of the first stage residuals leads further support to our endogeneity concerns.[33]

- Table A.3: alternative measurement of procurement performance. The estimates in Table A.3 are the analogous to our baseline estimates, but are obtained with outcome variables calculated with the definition of contract renegotiations adopted in Kang and Miller [2017]. Compared to our definition, a broader set of contract modifications are included to calculate the final duration and cost of the contract. Nevertheless, all the qualitative results from our baseline are robust if compared with the estimates reported in Table A.3.

- Recent research indicates that there may be considerable problems with the conventional IV regression technique particularly in its finite sample performance, and that approximations based on the asymptotic theory may yield poor results. A common way to refine the approximations for the distributions of the IV regression estimators and related test statistics is to employ a bootstrap method (see Young, Alwyn [2017]). In table A.4 we replicate our IV analysis by drawing 500 bootstrap samples in a fashion consistent with the error dependence within our cluster of observations (bureau and service category) and independence across observations. This method produces estimates that identify our parameters of interest as accurately as the baseline IV. Indeed,

---

[32]To represent endogeneity in the model, We assume the continuous explanatory variable *competence* to be endogenous, and that it is correlated with an unobserved omitted variable o_{ij}. Then, we assume: $\mathbb{E}(performance_{ijt}|Competence_{jt}, o_{jt}, \mathbf{X}) = \Phi(Competence_{jt}, o_{jt}, \mathbf{X}; \beta)$. By evaluating the impact of an instrument ($instr$) on *competence*, we further assume that $competence_{jt} = f(\mathbf{X}; o_{jt})$, $o_{jt} = \rho\, instr_{jt} + \epsilon_{jt}$ and $(o_{jt}, cf_{jt}) \perp\!\!\!\perp \mathbf{X}$. Then, we estimate a first stage of the endogenous explanatory variable on all the exogenous variables (including fixed effects) plus the extra regressor $instr_{jt}$: $competence_{jt} = \gamma\, instr_{jt} + \rho\, X_{jt} + \psi_j + \delta_t + \eta_{jt}$ and obtain the OLS residuals $res_{jt} = competence_{jt} - compet\hat{e}nce_{jt}$. In the second stage we use the fractional probit of $performance^g_{ijt}$ on $competence_{ijt}$, exogenous explanatory variables and $\hat{cf}_{jt}$ to estimate the scaled coefficient β. We thus include the extra regressors $\hat{cf}_{jt}$ in the estimating equation so that the remaining variation in the endogenous explanatory variable would not be correlated with the unobservables. $\mathbb{E}(performance_{ijt}|competence_{jt}, \hat{cf}_{jt}, \mathbf{X}) = \Phi(\beta\, competence_{jt} + \zeta\, \hat{cf}_{jt} + \theta\, X_{ijt} + \iota_j + \kappa_t)$.

[33]In control function estimates, bureau characteristics only are replaced by their standard scores. The outcome variables enter the regression in their non-standardized version. This is due to the need for non-negative values for the dependent variable when the dependent variable is assumed to be distributed as a binomial and, accordingly, the canonical link function, providing the relationship between the linear predictor and the mean of the distribution function, is a logit.

the bootstrap shows that our baseline analysis does not understate confidence intervals so that the significance of our baseline IV point estimates appears to be robust.

Table A.1: IV-LIML regressions

	Cost Performance			Time Performance		
	(1)	(2)	(3)	(4)	(5)	(6)
Competence	0.49***	0.26*	0.38***	0.41***	0.32**	0.37***
	(0.11)	(0.15)	(0.10)	(0.11)	(0.14)	(0.09)
Bureau Experience	-0.05	-0.03	-0.04	-0.07*	-0.07	-0.07
	(0.03)	(0.03)	(0.03)	(0.04)	(0.04)	(0.04)
Bureau Size	-0.01	0.02	0.01	-0.00	0.01	0.00
	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
Centered R-squared	.07	.13	.1	.07	.09	.08
Observations	122526	122526	122526	122526	122526	122526
Weak Id. F-Test	39.64	21.76	29.38	39.64	21.76	29.38
Underid. F-Test	40.21	18.65	54.2	40.21	18.65	54.2
Overid. F-Test	0	0	1.71	0	0	.29

Notes: Instruments are: *Relevant Deaths* and *Proximal Deaths*. Both contract outcomes and bureau characteristics are replaced by their standard scores. Standard errors are clustered by bureau and service category and are in parentheses. All models include controls for contract features (cost plus format and solicitation procedure), buyer characteristics (experience and yearly procurement budget), fixed effects for service category, agency, deciles for contract value and duration, year, and U.S. state of performance. * Significant at the 10 percent level; ** Significant at the 5 percent level; *** significant at the 1 percent level.

Table A.2: Control Function Estimates

	Cost Performance			Time Performance		
	(1)	(2)	(3)	(6)	(7)	(8)
Competence	0.10***	0.07	0.09***	0.13***	0.09*	0.12***
	(0.02)	(0.04)	(0.02)	(0.03)	(0.06)	(0.03)
FS Residual	-0.09***	-0.06	-0.08***	-0.11***	-0.08	-0.10***
	(0.02)	(0.04)	(0.02)	(0.03)	(0.06)	(0.03)
Observations	131686	131686	131686	131686	131686	131686

Notes: Table 8 is replicated by using the two-step fractional probit approach proposed in Wooldridge [2002]. For more details, see notes from Table 8.

Table A.3: Alternative Performance Measures: IV Estimates

	Cost Performance			Time Performance		
	(1)	(2)	(3)	(4)	(5)	(6)
Q28	0.49***	0.23	0.36***	0.40***	0.31**	0.35***
	(0.11)	(0.15)	(0.10)	(0.11)	(0.13)	(0.09)
Observations	122326	122326	122326	122326	122326	122326
Centered R-squared	.07	.13	.11	.07		.08
Weak Id. F-Test	39.48	21.76	29.33	39.48	21.76	29.33
Underid. F-Test	40.06	18.65	54.11	40.06	18.65	54.11
Overid. F-Test	0	0	2.02	0	0	.28

Notes: Results from Table 8 are replicated by recomputing *Cost Performance* and *Time performance* according to the definition of contract renegotiation proposed by Karam and Miller (2017). Instruments are: *Relevant Deaths* and *Proximal Deaths*. Both contract outcomes and bureau characteristics are replaced by their standard scores. Standard errors are clustered by bureau and service category and are in parentheses. All models include controls for contract features (cost plus format and solicitation procedure), buyer characteristics (experience and yearly procurement budget), fixed effects for service category, agency, deciles for contract value and duration, year, and U.S. state of performance. * Significant at the 10 percent level; ** Significant at the 5 percent level; *** significant at the 1 percent level.

Table A.4: IV regressions - Cluster Bootstrap

	Cost Performance			Time Performance		
	(1)	(2)	(3)	(4)	(5)	(6)
Competence	0.75***	0.12*	0.23***	0.47***	0.24***	0.28***
	(0.13)	(0.07)	(0.06)	(0.14)	(0.08)	(0.07)
Observations	122533	122533	122533	122533	122533	122533

Notes: Results from Table 8 are replicated with standard errors - in parentheses - clustered by bureau and service category and bootstrapped with 500 replications. Instruments are: *Relevant Deaths* and *Proximal Deaths*. Both contract outcomes and bureau characteristics are replaced by their standard scores. All models include controls for contract features (cost plus format and solicitation procedure), buyer characteristics (experience and yearly procurement budget), fixed effects for service category, agency, deciles for contract value and duration, year, and U.S. state of performance. * Significant at the 10 percent level; ** Significant at the 5 percent level; *** significant at the 1 percent level.

Table A.5: Cooperation, Incentives, and Skills.

	Competence			Cost Performance		Time Performance	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Cooperation Q20			0.97***				
			(0.05)				
Cooperation Q26	0.83***	0.59***	-0.03				
	(0.03)	(0.04)	(0.04)				
Incentives		0.22***	0.10***				
		(0.04)	(0.03)				
Skills		0.09**	0.06**				
		(0.04)	(0.03)				
PCA - Cooperation				0.09***	0.09***	0.04**	0.04**
				(0.02)	(0.02)	(0.02)	(0.02)
PCA - Skill/Incentive					-0.00		0.00
					(0.01)		(0.01)
Bureau Experience				-0.02	-0.02	-0.05	-0.05
				(0.03)	(0.03)	(0.04)	(0.04)
Bureau Size				0.04***	0.04***	0.04**	0.04**
				(0.02)	(0.02)	(0.02)	(0.02)
Observations	441	441	441	122526	122526	122526	122526
R-squared	.67	.72	.84	.15	.15	.12	.12
Amount FEs	No	No	No	Yes	Yes	Yes	Yes
Duration FEs	No	No	No	Yes	Yes	Yes	Yes
Agency FEs	No	No	No	Yes	Yes	Yes	Yes
Year FEs	No	No	No	Yes	Yes	Yes	Yes
State FEs	No	No	No	Yes	Yes	Yes	Yes

Notes: OLS regressions of *competence* on *cooperation, incentives, and skills* collapsed at bureau-year level (columns 1-3) and of cost performance (columns 5-6) and time performance (columns 7-8) on PCAs for cooperation and incentives/skills.