Bureaucratic Competence and Procurement Outcomes

Bureaucratic Competence and Procurement Outcomes*

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

January 30, 2020

Abstract

To what extent 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. We use the death occurrences of specific types of employees as instruments and find that an increase in bureau competence causes a significant and economically important reduction in: i) time delays, ii) cost overruns, and iii) number of renegotiations. Cooperation within the office appears to be a key driver of the findings.

Keywords: Buyer quality, competence, procurement, public management, state capacity.

JEL Classification: D44, H11, H57.

*Francesco Decarolis, Bocconi University and IGIER, email: *francesco.decarolis@unibocconi.it*; Leonardo M. Giuffrida, ZEW Mannheim and MaCCI, email: *leonardo.giuffrida@zew.de*; Elisabetta Iossa, University of Rome Tor Vergata, GREEN-Bocconi and CEPR, email: *elisabetta.iossa@uniroma2.it*; Vincenzo Mollisi, University of Mannheim, email: *vmollisi@mail.uni-mannheim.de*; Giancarlo Spagnolo, SITE-Stockholm School of Economics, EIEF, University of Rome Tor Vergata and CEPR, email: *spagnologianca@gmail.com*. For useful comments, we are grateful to Oriana Bandiera, Klenio Barbosa, Erich Battistin, Michael Best, Sascha Becker, Stephane Bonhomme, Gaétan de Rassenfosse, Mansi Deshpande, Frederico Finan, Philippe Gaignepain, Luca Gnan, Josh Gottlieb, Ari Hyytinen, Vitalijs Jascisens, Camilo Garcia Jimeno, Alex MacKay, Marco Manacorda, Claudio Michelacci, Magne Mongstad, Andras Niedermayer, Benjamin Olken, Lars Persson, Gustavo Piga, Andrea Prat, Emilio Raiteri, Imran Rasul, Stephane Saussier, Andrei Shleifer, Otto Toivanen and the participants at seminars where preliminary versions of this study were presented. Francesco Decarolis acknowledges research funding from the ERC (Stg-REPCOR-679217). Giancarlo Spagnolo acknowledges research funding from Handelsbanken (P16-0277).

I Introduction

The ability of the state to govern effectively is an important determinant of economic activity (Acemoglu [2005]; Besley and Persson [2009]). To effectively administer its policies the state crucially depends on the qualities and efficiency of the bureaucracy, as widely recognized since Weber [1922]. Among its core functions, the state bureaucracy purchases large amounts of goods, works and services, many of which essential to government activity and society in general. The World Bank has recently began to release its Benchmarking Public Procurement, which examines the procurement process in 180 economies. The report reveals the existence of considerable heterogeneity across states, which raises questions about the capacity of public buyers to make efficient purchases. Concerns on the lack of competence of public procurers have recently been voiced also by Saussier and Tirole [2015] and confirmed by a growing economic literature examining the impact of public buyers on the prices of standardized goods.[1] As public procurement accounts for 12 percent of world GDP [OECD, 2017], it is now acknoledged that an inefficient procurement bureaucracy can have a significant impact on state capacity and ultimately on economic activities.

In this paper, we provide the first quantification of the impact of bureaucratic competence on complex procurements – involving public works and services – in the context of a developed country such as the U.S. Throughout the paper, we use the term "competence" to capture all of those factors which affect the capacity of procurement offices to effectively perform their mission. Our main contribution is to show how extensive survey data on US federal employees can be used to quantify this notion of bureau competence and then how to relate it to procurement performance, by combining it with a large contract-level dataset.

The economic literature has offered a variety of explanations for why more competent, higher-quality procurers should improve procurement outcomes. These are related to the

[1] Bandiera, Prat and Valletti [2009], for example, estimates 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. In a similar vein, Best, Hjort and Szakonyi [2017] estimates 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. Bucciol, Camboni and Valbonesi [2017] also finds that buyers' characteristics are a key driver of the substantially different prices paid for the same medical devices.

buyers' involvement in the ex-ante identification and description of procurement needs in the tender documents, the design of the award procedures and contracts and the ex-post contract management (see e.g. Spulber [1990], Manelli and Vincent [1995] and Bajari and Tadelis [2001]). The role of buyers is even more significant in the purchase of non-standardized, complex goods, works or services, as these procurement tasks require more detailed knowledge in multiple fields and from multiple actors within the organization.[2] These actors must then be coordinated effectively until the final contract is executed, which can happen months, if not years, after it is tendered.

Whilst in the procurement of standardized goods, efficiency can be measured in terms of price dispersion, in the procurement of complex heterogeneous goods, price comparisons are insignificant. This raises the issue of what performance measure is suitable. To overcome this challenge, we consider three main performance measures, all available through the Federal Procurement Data System (FPDS), a system tracking nearly every awarded federal contract, as well as every follow-on action. The first two performance measures are regularly used by governments in their procurement reports as well as by researchers in the field when focussing on more complex procurements than standardized goods, those for which competence is more needed.[3] The third performance measure that we use is the number of renegotiation episodes. It captures Williamson [1971]'s transaction or "haggling" costs which exist whatever the reason behind the renegotiation and which have been shown to be economically sizeable for complex contracts, [Bajari, Houghton and Tadelis, 2014].[4]

There are two main measurement challenges regarding bureaucratic competence. Translating the complex and multifaceted concept of competence into a variable entails some choices. The prevalent approach in the field has been to measure buyers' competence using a fixed effects strategy (Bandiera, Prat and Valletti [2009], Best, Hjort and Szakonyi [2017] and Bucciol, Camboni and Valbonesi [2017]), but this requires adequate variability in the

[2]Complex procurement requires technical knowledge of product and market characteristics, legal knowledge of complex legal rules and strategic abilities in the tender design and contract negotiation. Typically, different experts with different skills are involved in the different phases of the procurement process.

[3]See, for instance, Bajari and Lewis [2011] for delays, and Mohamed, Khoury and Hafez [2011], Iimi [2013], Bajari, Houghton and Tadelis [2014] and Jung et al. [2018] for cost overruns.

[4]The size of cost and time overruns also partly captures these huggling costs as, given the number of renegotiations, larger renegotiated values suggest a greater complexity of the renegotiated contract.

data and leaves open the question of what exactly competence is. An exception is Rasul and Rogger [2016], who study complex procurement in the context of a developing country, using a self administered survey. In our study, we also use a survey tool, but make use of a unique and under-exploited dataset: the Federal Employee Viewpoint Survey (FEVS). This survey has been administered for more than ten years with the same questions to nearly all U.S. 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.[5]

The second 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 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 this source of bias, we use an instrumental variable strategy exploiting exogenous changes in bureaus' competence based on death occurrences of specific types of

[5]It might also seem tautological that procurement outcomes correlate with a survey measure which is itself measuring opinions on an outcome - the *overall quality of work done*. But this would only be true if the respondents were to give prominence in their responses to procurement outcomes. As discussed below, our data structure makes this unlikely as the bureau - the unit of analysis at which we work - is rather large, encompassing hundreds of workers. Their responses to the FEVS are therefore better seen as measures of the overall efficacy of the workflow and processes within the bureau, hence proxying for the ideal measure of competence whose traits we described above. An extensive set of robustness checks will assesses the potential problems of measuring competence through the FEVS data.

employees (in the spirit of Warren [2014]).[6] For this purpose, we use a third dataset (FedScope) which contains detailed characteristics of the public workforce. In particular, we construct two instruments that account for the death of bureaus' managers and other white collar employees who are relatively young and have high wage and, hence, who are likely to be most relevant to bureaus' processes and workflows, in particular if the office is not competently run. The idea is that more competent offices adopt better managerial practices, routines and processes that are more resilient to risks, such that of an unexpected loss of a key employee, and less dependent on specific individuals. More competent offices would therefore incur less disruption when important employees suddenly die, including disruption of procurement performance, than less competent ones. This is precisely what the first stage of our IV strategy documents.

The IV estimation strategy allows us to estimate a causal effect of bureau competence on procurement outcomes that is an order of magnitude larger than the corresponding OLS estimate. A one standard deviation increase in competence reduces the number of days of delay by 23 percent, cost overruns by 29 percent and the number of renegotiations by half. This implies that, if all federal bureaus were to obtain NASA's bureau "John Glenn Research Center at Lewis Field" 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, cost overruns would drop by $14.7 billions over the entire sample analyzed (841 thousand days and $2.6 billions, respectively, on yearly basis). We also observe a consistently negative effect of greater competence on the number of renegotiations: one standard deviation increase in competence causes 0.5 (40%) and 0.8 (71%) fewer cost renegotiations and time renegotiations, respectively, 1.3 (52%) fewer in total.

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

[6]Warren [2014] 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.

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 bureau competence. The prominence of cooperation conforms with the view that successful procurement requires to appropriately handle and coordinate a multiplicity of tasks involving different individuals and offices. The complexity of the environment implies that no one size can fit all: tender and contract design must take into account that the often complex characteristics of each specific 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 and managerial processes ensuring smooth coordination and collaboration among employees with different skills is thus essential.

Finally, following Gibbons and Henderson [2016], on the role of on management practices in organizations, 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. In this sense, we investigate whether what "good managers" do is to adopt management practices that foster cooperation within their organization. To this purpose, 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.

This connects our work to the recent literature on the role of the role of managers and managerial practices in the public sector (Bloom et al. [2014], Bloom et al. [2015] and Janke, Propper and Sadun [2019]). In particular, Rasul and Rogger [2016] show that public project completion in Nigeria correlates positively with management practices increasing bureaucratic autonomy, but negatively with those strengthening incentives/monitoring. In contrast, 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. Our findings

on the importance of cooperation in public offices complements the results in Blader, Gartenberg and Prat [2016] on the benefits of "cooperative" managerial practices in private firms, relative to high powered individual incentives. Of course, our results also contribute more generally to the recent and growing literature on the determinants of public procurement outcomes.[7]

Overall, 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 certifying its competence.[8] 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 bureau competence and on cooperation suggest that, while certification of individual contracting officer's capabilities is 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.

[7] 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 or maximum prices (Chassang and Ortner [2019] and Conley and Decarolis [2016], respectively), 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]).

[8] 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.

II 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. 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. Within the same bureau, we will also exploit the dispersion of local offices across different US states. In fact, the procurement outcomes involving a contract taking place in a certain area might be influenced by the competence of both the overall bureau and of its local offices, with the former mattering more for the initial tender design and the latter more concerned with contract management.[9]

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

[9]Although there does not exist a unique organizational model, the relevance of local offices is clearly explained in the source selection guidelines of a few agencies. As an example see 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.

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 starting with 2010 since the FEVS has been run every other year before 2010. There is a total of 96 bureaus from 23 agencies. The agencies that are invited to participate in the survey 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.[10] 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 bureau. 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.[11] We first transform these answers into numerical values from 1 (very poor) to 5 (very good),[12]

[10] 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 10.

[11] 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).

[12] A strength of the FEVS, relative to most surveys, is a limited risk that different employees associate a

Table 1: Summary statistics

	Mean	Median	S.D.	N
Bureau Characteristics (FEVS Data)				
Competence (Q28)	0.50	0.49	0.14	122533
Ag. Competence (Q29)	0.53	0.53	0.11	122533
Cooperation (Q20)	0.45	0.48	0.18	122533
Skill (Q21)	0.47	0.48	0.12	122533
Incentive (Q23)	0.38	0.35	0.10	122533
Contract Characteristics (FPDS Data)				
Contract Amount (000)	531.7	87.0	3595.7	122533
Expected Duration (days)	244.0	212	208.0	122533
Cost Performance	0.85	1	0.25	122533
Time Performance	0.73	1	0.33	122533
Total Cost (000)	891.6	109.2	7127.4	122533
Total Time (days)	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
No. of Offers	3.84	2	6.17	122533
Works	0.19	0	0.39	122533
Bureau Characteristics (FPDS Data)				
$Bu.Performance^C$	0.84	0.89	0.15	112658
$Bu.Performance^T$	0.73	0.73	0.18	112658
Bureau Experience (00)	2.56	0.4	5.23	122533
Bureau Size (000,000)	1 111	380	1 370	122533

Notes: The top panel presents summary statistics for the FEVS data. The unit of observation is a contract. The relative statistics are rescaled by considering the empirical distribution of our sample: the minimum and maximum values are 0.70 and 0.89 for *competence* (Q28), 0.55 and 0.79 for *Ag.competence* (Q29), 0.61 and 0.81 for *cooperation* (Q20), 0.38 and 0.72 for *skills* (Q21), and 0.28 and 0.68 for *incentives* (Q23). The mid and bottom panels present summary statistics for the FPDS data and the unit of observation is still a contract. *Bureau Experience* is scaled down by hundred of units; *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.

then we aggregate answers at the bureau level,[13] and finally we normalize the resulting

different meaning to the same answer about the competence level. Albeit such risk cannot be fully ruled out, the FEVS is an extremely well known survey among public employees: it has been administered regularly over almost two decades to a large share of federal employees, with a very consistent structure of the survey and wording of the questions over time. It also comes with detailed guidelines on how to interpret questions. These features are crucial in limiting the risk that the observed variability in the data is merely the result of an heterogeneous interpretation across employees of what a question is asking. Despite this advantage, there are well known concerns in using survey data in economics, Bertrand and Mullainathan [2001], and some of these problems are even more pronounced when eliciting expectations via Likert scale questions, Giustinelli and Manski [2018]. In the next question we present the IV strategy we use to deal with some of these concerns.

[13]The disaggregated data (on average around 380,000 answers of civil servants each year), show substantial variability, with a ratio of one standard deviation over the mean equal to 0.81/4.22=0.19. This value is six times bigger than the same statistic calculated for the aggregated data, which is equal to 0.03/0,80=0,0375.

variables to be between zero and one.[14] 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.[15] The data track every transaction between federal contracting bureaus and sellers. The system contains detailed information on contract actions over $3,000. Information is of two 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. Compared to the procurement of standardized goods, these contracts involve ex-post cost uncertainty, multidimensional quality heterogeneity and limited contractibility, thus making a competent management of the procurement process crucial and post-award amendments, with the high haggling cost they imply, a useful proxy of contract performance (Tadelis [2002]).[16] Since not all modifications

This variation in the disaggregated data is what allows us to perform a meaningful aggregation at the bureau-year level through the mean statistic. Also note that, despite the lower variability post-aggregation, the relevant variation is indeed at the bureau-level. Indeed, as discussed below, this allows to link the relationship between procurement outcomes and the average level of "competence" at bureau level, encompassing all the features of the bureau-level competence summarized by the FEVS question.

[14]The normalization uses as boundary values the ones observed in the bureau-level data. This offers a more accurate description of the identifying variation in the data relative to using the theoretical minimum (1) and maximum (5) which are observed in the individual responses, but never at the aggregated bureau level.

[15]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].

[16]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 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

are equally problematic, we split the set of amendments into two broad categories: *in-scope* and *out-of-scope* revisions.[17] In line with other studies that use FPDS data, we consider *in-scope* amendments only for measuring delays and cost overruns.[18]

The quantitative relevance of these contract modifications is evident from the summary statistics reported in the central panel of Table 1. The sample ranges from 2010 to 2015 and consists of 122,533 completed projects, associated with 821 procurement categories (i.e., the types of work or service procured). 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 contractual duration is 244 days, while the final contract duration inclusive of any delay is 486 days. Conversely, the average award per contract is $531,700, while the total cost, inclusive of any cost overrun, is $892,000.[19] In both cases, the medians are lower than the means.

To operationalize the data on time and cost renegotiations into proxy variables 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: $performance^{g}_{ijt} = \frac{expected\ outcome^{g}_{ijt}}{expected\ outcome^{g}_{ijt} + overrun^{g}_{ijt}}$, 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.[20] Each performance measure

see also Guasch, Laffont and Straub [2008], Lewis and Bajari [2011], Decarolis [2014], Bajari, Houghton and Tadelis [2014], De Silva et al. [2017] and Decarolis [2018]. For arguments on why, instead, renegotiations can be beneficial in the face of public contracts are inherently too rigid see Beuve, Moszoro and Saussier [2019].

[17] 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*.

[18] 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.6 in appendix).

[19] 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.

[20] In the appendix, we also report results obtained by replacing the convex outcome measure, *Expected*

ranges between zero, worst performance, and one, perfect performance (i.e., no overrun). In the data, about half of the observations show neither cost nor time overrun. The coefficient of the linear correlation between the two equals 0.52 with a Spearman ρ of 0.57.

The 96 bureaus in our sample manage on average 639 contracts per bureau (s.d. 1823) of an average amount of $583,000 (s.d. $881,000).The data also exhibit 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.

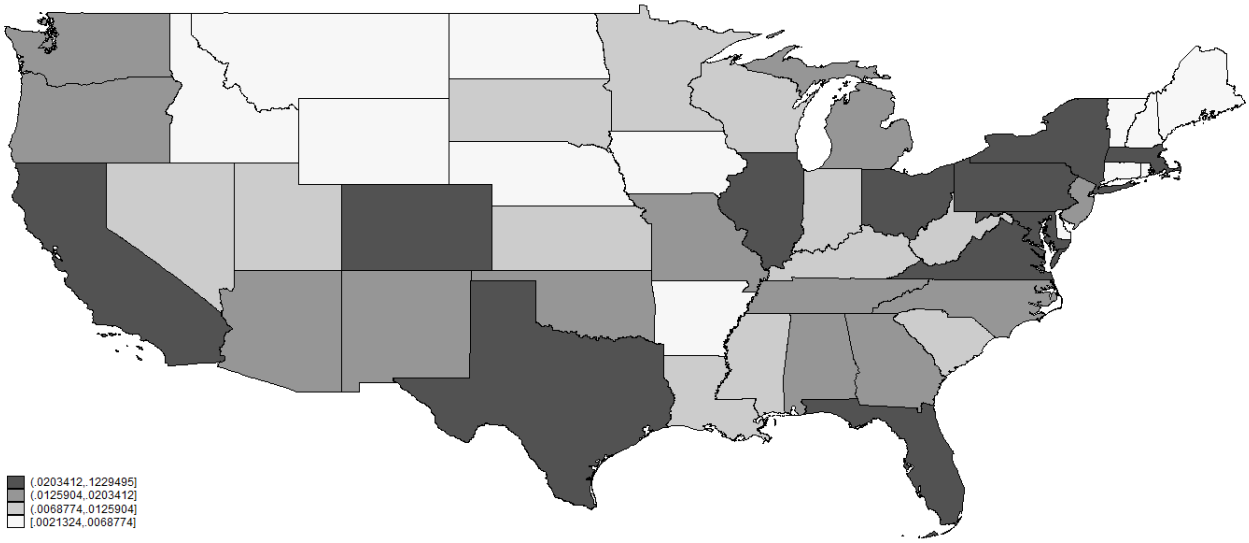
As mentioned in the introduction, an additional outcome measures will also be the number of renegotiations, both overall and separately for cost and time purposes. There are on average 1.29 renegotiation episodes involving costs and 1.17 episodes involving delivery time. Table 1 reports summary statistics 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 procurement category, while the second measures the cumulative value of contracts a bureau has awarded in the same year in the same procurement category.[21] Furthermore, after presenting the baseline results, we summarize the findings of an extensive appendix where we show how competence is associated with other outcomes in terms of the types of awarding procedures (negotiations, auctions, simplified acquisition procedures, etc.) and extent of competition between suppliers (number of participants and instances of single participant).

Two important limitations of the data are that they do not allow observing contract

Outcome /(Expected Outcome + Overrun), with a linear specification such as the one usually employed in literature: *(Expected outcome + Overrun) / Expected Outcome*. The problem with the latter specification is the sensitivity to outliers and misreports of expected and actual outcomes, which can strongly bias our two indexes in the tails of their distributions. Qualitatively, the findings are very similar across the different measures provided that outliers are excluded when using the linear measure. See Table A.10 in the appendix for further details.

[21] Interestingly, their correlation with our measure of competence is low, -0.08 for experience and -0.15 for size. As discussed later, this is likely due to our measure of competence being closely linked to how employees perceive cooperation within their office, and cooperation is unlikely to grow with the bureau experience or size.

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).

penalties and suppliers past performance scores. There is likely some discretion in the amount, design, structure, and enforcement of penalties, within the constrains set by the law (and by how the law is implemented by the relevant courts). Similarly, there is some discretion in how information from the PPIRS (past performance information retrieval system) is used to select contractors. However, there is no available evidence about systematic differences in the use of penalties and past performance across bureaus and agencies. While studying their interaction with competence would certainly be interesting, this would likely require observing a very homogeneous set of contracts and accessing information which is not readily publicly available. The present paper, instead, focuses on a very large and heterogeneous set of complex works and service contracts, exploiting the size of the database to make up for the large unobserved heterogeneity in the object and structure of these different contracts, including that on the size, structure, and enforcement of penalties. Furthermore, an important point is also that while penalties are typically linked to delays, our findings are very robust to the use of alternative outcomes on cost overrnuns and the number of renegotiation episodes.

C. Public Workforce Characteristics: FedScope Data

The Office of Personnel Management (OPM) is an independent federal agency that func-

tions 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.[22] 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.[23] Fedscope is the third data source that we use and we merge it with the FPDS at the bureau level. FedScope data divided into five subject categories (called "cubes"), of which we only consider the "Employment" cube and the "Separations" cube for the years 2010-2015.

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. The adjective "local" refers to the local branches of bureaus,

The Employment cube contains several demographic characteristics along with informa-

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

[23]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.

tion 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 per year, subdivided into 96 bureaus 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 branch of each bureau. Finally, the geographical information in FedScope enables us to match the location (state) of each single federal employee with that of contract performance.[24] More details on the specific ways in which these data are used to construct our instruments are presented in Section IV. Before that, however, in section III we present some relevant descriptive facts about the data that serve to establish the link between the FEVS and FPDS data.

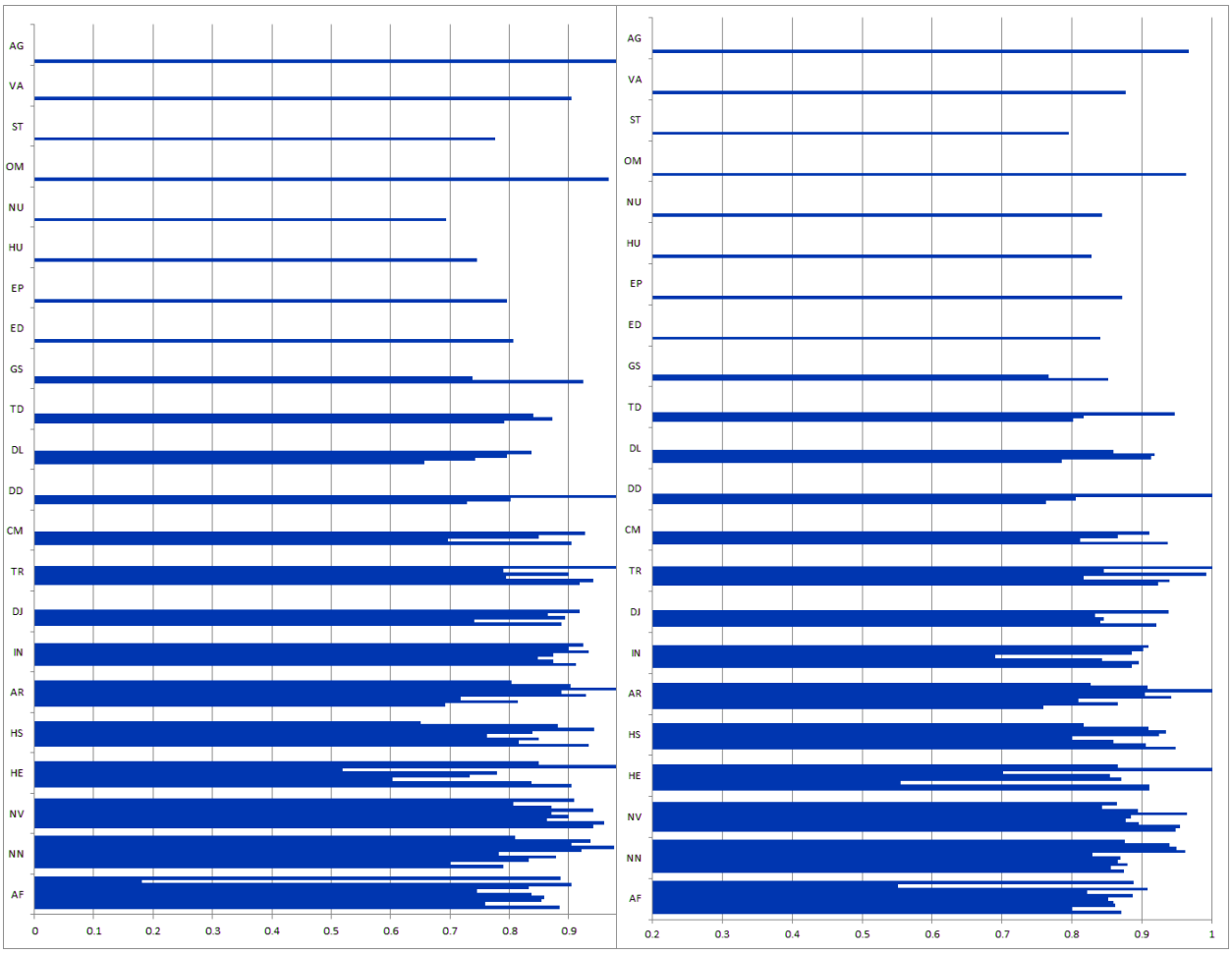
III Descriptive Evidence

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

To illustrate the first point, we begin by constructing a bureau-level performance met-

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

Figure 2: Procurement Performance across Bureaus grouped by Agency



(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 of each agencies. $Bu.Performance^C$ and $Bu.Performance^T$ are moving averages calculated on the entire time series of cost and time performances. We employ Bartlett weights (see Bartlett [1950] to assign more weight to more recent observations). The length of the horizontal lines measure the performance of $Bu.Performance^C$ (left column) and $Bu.Performance^T$ (right column).

ric 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 bureau had previously procured for the same procurement category j: $Bu.Performance_{ijt}^g = \sum_{\{\mathbf{1}|t'<t\}} w_{ijt'} * performance_{ijt'}^g / \sum_{\{\mathbf{1}|t'<t\}} w_{ijt'}$, where w are Bartlett window weights, see Bartlett [1950], which weight more the most 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, by grouping them by agencies. For each agency, we report the performance of all bureaus with which the agency appears in the FPDS. Thus, for instance, the bureau at the very top of the figure is the Rural Housing Service bureau of the United States Department of Agriculture (AG), while at the very bottom of the figure there are the 10 bureaus of the Air Force (AF).[25] 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

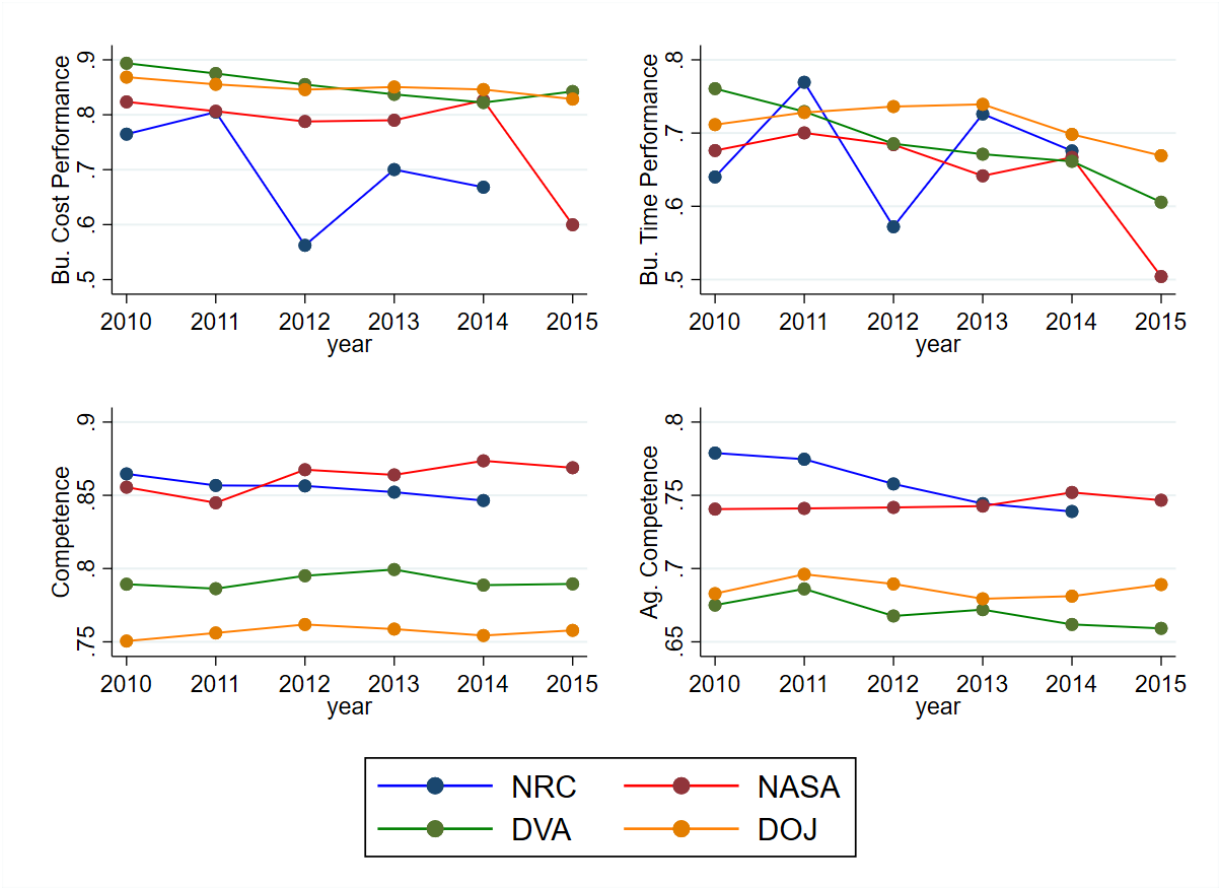
Notes: Average agency scores throughout our whole time-span of data (2010-2015) for *Competence*, *Ag.Competence*, $Bu.Performance^C$, and $Bu.Performance^T$ reported for the two best agencies in terms of average Competence - Nuclear Regulatory Commission (NRC) and National Aeronautics and Space Administration (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

[25]The full list of bureaus is reported in the appendix. See Table A.1.

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)). There are no records for contracts awarded and completed for NRC in the working sample in 2015 and relative scores of $Bu.Performance^C$ and $Bu.Performance^T$ are therefore not computed. For the sake of consistency, also relative agency-level averages for *Competence* and *Ag. Competence* are excluded.

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.

IV Empirical Analysis

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

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

where $g = \{C, T\}$ indicates whether the outcome variable is cost or time performance; i, j, k, c, and t indicate the contract, bureau, agency, procurement category and year, respectively; $\mathbf{X_{ij}}$ is a matrix of contract- and bureau-level covariates, and κ_k, ζ_c, and τ_t indicate agency, procurement category and year fixed effects, respectively. In the estimates we also include state fixed effects and we control for the contract initial amount and duration to proxy for contract complexity. Bureau fixed effects, instead, are not included as the high degree of persistency of competence over time, coupled with the short length of our time span, makes it unfeasible to identify competence when these fixed effects are included.[26] 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 across bureau within the same agency.

There are several challenges in interpreting the OLS estimate in a causal fashion. First,

[26]However, although we do not pursue this strategy in the paper, it would be possible to extend the panel of bureau features sourced by FEVS for a subset of bureaus back to 2002 which can potentially allow us to include bureau fixed effects. The persistency of the competence measures is evident for the case of the four agencies shown in Figure 3. An alternative strategy exploiting long run changes is described when introducing the robustness checks.

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 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.[27]

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 contract performance. In particular, we always include agency and procurement 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 in which the contract is performed. 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 bureaus can have on the procurement processes.

First, inspired by the vast literature on CEO deaths, we focus on deaths of those em-

[27]One might also worry about reverse causality, 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 causality bias is addressed by our IV strategy presented next.

Table 4: Instruments Summary Statistics

	Mean	Median	S.D.	N
Proximal Deaths	0.62	1	0.48	122533
Relevant Deaths	0.91	1	0.29	122533

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

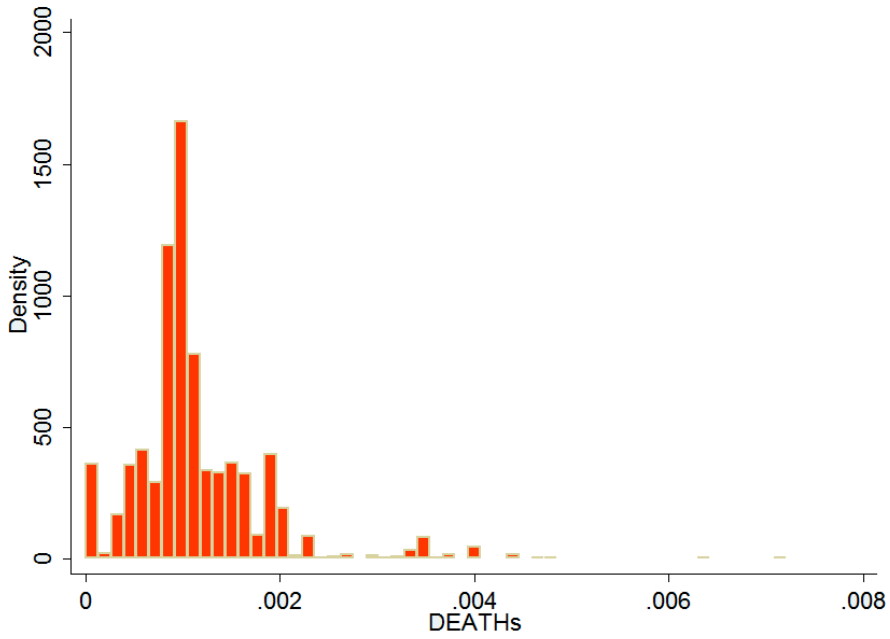
ployees more likely to have positive roles for the productivity of their office. We thus look at white-collar employees of an age no higher than the median and with a salary no lower than the median, relative to the distributions of these variables for other white-collar employees. According to Table 2, this implies looking at employees with a salary of $50,000 or more and an age of 49 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:

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

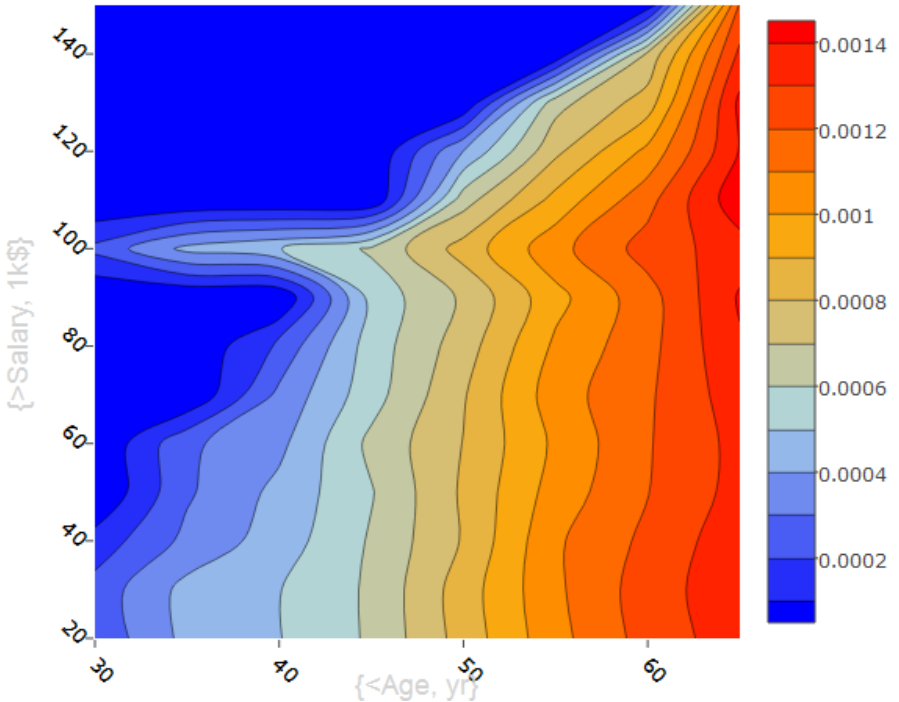
where j is the bureau and t the year. Table 4 reports the summary statistics for this instruments which are most easily understood through Figure 4a. This figure illustrates for all the bureaus-years in the sample, the distribution of the share of deaths within the relevant age/salary population. It 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 exogeneity 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 unexpected and likely disruptive. We return to this aspect after having introduced the other instrument.

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

(a) Histograms at contract level



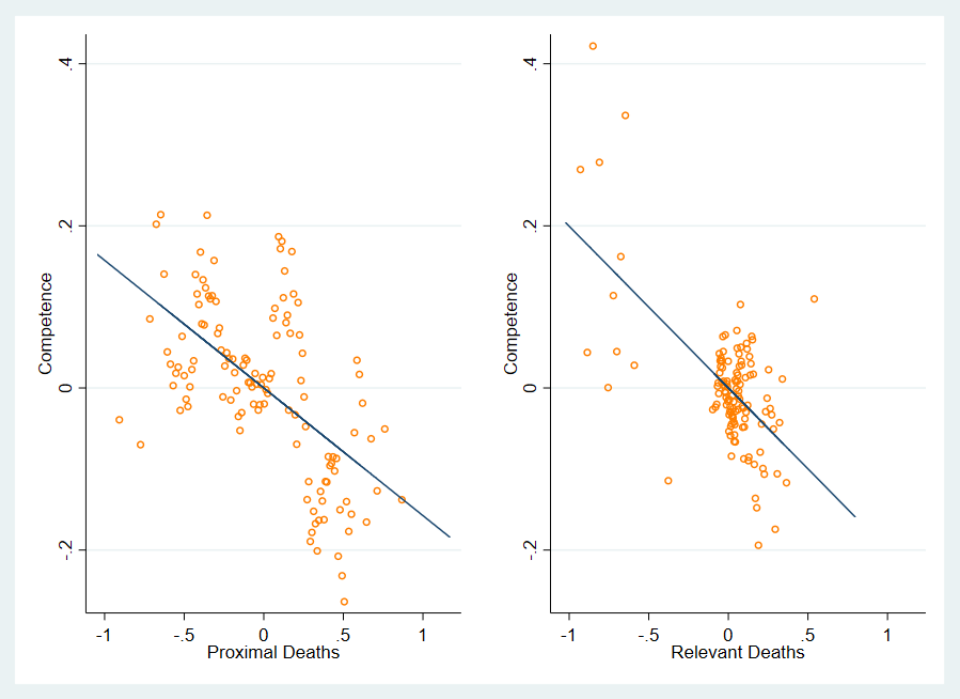
(b) Median frequency by Age and Salary



Notes: In panel (a), we report the histogram of the ratio between the count of death events and the workforce population for each bureau throughout our whole time-span of data (2010-2015). In panel (b), we report the median value of the ratio for each combination of age and salary in the same period.

For a second instrument, we follow Bruce, de Figueiredo and Silverman [2019] who suggest that the spatial proximity of a death event 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 (irrespective of their wage and salary) 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 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).

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. The variables are residualized including as controls: Bureau Experience, Bureau Size, a set of dummies for the deciles of contract value and duration, agency fixed effect, procurement category fixed effect, year fixed effect, and State fixed effects. Each graph is a binned scatterplot. This means that each point represents the mean statistic of the residualized Proximal Deaths and Relevant Deaths variables inside each bin. The selected number of bins is 122 and it is optimal in minimizing the (asymptotic) integrated mean squared error (IMSE) following Cattaneo et al. [2019].

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, *rele-*

vant 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.

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 has a long tradition in economics.[28] The validity of the instrument is supported by the fact that as-good-as-random separations of office managers negatively affect 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, or even simply labor shortage. Since the FEVS data covers a large share of employees, it would be highly unlikely that the occurrence of a relevant death does not trigger any worsening in how the affected employees respond to the survey questions. 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. In the federal workforce, 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. Both these effects will be smaller the higher the competence of the bureau, as a more competent bureau will have more effective procedures to manage such shocks. In this sense, the broad definition of competence captured by the FEVS question used ensures that the impact of death events should influence procurement outcomes only through variation in competence.[29]

[28]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.

[29]In the appendix, we explore the soundness of the exclusion restriction by constructing alternative measures of competence through a principal component analysis of various FEVS questions.

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	-0.08***	-0.07**	-0.07**	-0.06*	-0.13***	-0.15***	-0.11**	-0.12**
	(0.03)	(0.03)	(0.03)	(0.03)	(0.04)	(0.04)	(0.05)	(0.05)
Median LOS	0.02	0.02	0.01	0.00	0.01	0.02	0.01	0.00
	(0.03)	(0.03)	(0.03)	(0.03)	(0.04)	(0.04)	(0.04)	(0.04)
Median Salary	-0.06***	-0.06***	-0.06***	-0.07***	-0.03	-0.05**	-0.05*	-0.04
	(0.01)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.03)	(0.03)
Median WF Composition	0.11**	0.11**	0.10**	0.06	0.24*	0.18	0.10	0.08
	(0.05)	(0.04)	(0.05)	(0.05)	(0.13)	(0.13)	(0.14)	(0.14)
Accomplishment		-0.49	-0.52	0.84		2.24	-1.52	-0.78
		(0.88)	(0.98)	(1.06)		(1.39)	(1.86)	(2.17)
Appreciation		-0.32	-0.87	-1.76		-2.69	-0.88	-2.12
		(0.75)	(1.04)	(1.27)		(1.82)	(2.22)	(2.74)
Level of Workload		-1.06**	-1.16**	-1.09*		-1.83	-2.53*	-2.28*
		(0.47)	(0.53)	(0.55)		(1.19)	(1.31)	(1.35)
Physical condition workplace		-0.16	-0.36	-0.20		-0.12	-0.69	-0.61
		(0.44)	(0.45)	(0.46)		(0.91)	(1.14)	(1.10)
Integration policy			-0.37	-0.10			-1.30	-1.08
			(0.62)	(0.58)			(1.23)	(1.28)
Health Security			0.44	0.35			0.82	0.96
			(0.61)	(0.57)			(1.30)	(1.24)
Good Place to work			0.67	3.02***			2.16	3.81
			(0.71)	(1.05)			(1.38)	(2.32)
Balance work/life			-0.53	-0.17			-2.38	-1.50
			(1.15)	(1.12)			(2.03)	(2.17)
Respect and Self esteem			-0.21	-0.34			5.28*	4.77
			(1.51)	(1.42)			(2.92)	(3.08)
Job Satisfaction				0.29				2.79
				(1.25)				(2.93)
Pay Satisfaction				0.08				-0.57
				(0.38)				(0.93)
Organization Satisfaction				-3.55***				-4.22
				(1.20)				(3.17)
Healthcare Program				0.17				0.12
				(0.11)				(0.38)
R-squared	.25	.26	.26	.26	.22	.24	.26	.27
N	6711	6711	6711	6711	440	440	440	440

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

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 perform a regression analysis (see Table 5) to identify the determinants of our two instruments. Based on the way we construct them, we find that these deaths are associated in a mechanical way with the median age and the median salary of the bureau. *Relevant Deaths* are not associated with any of the potential death-event predictors at the bureau level appearing in Table 5: none reaches statistical significance. A similar scenario applies for *Proximal Deaths*, yet in this case, three predictors reach 95% statistical significance: *No. of Contracts*, *Good Place to Work*, and *Organization Satisfaction*. In Table A.8 in the appendix, we show that including these three bureau-level variables in our baseline analysis do not alter the results.

V Results

We begin the presentation of our results from Table 6 where we show the OLS estimates corresponding to equation (3). 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.[30] To

[30]The standard errors are two-way clustered at bureau and procurement category level. The idea is that employees with similar skills are likely to be involved, within the same bureau, in the purchasing process of the same categories of procurements. As we do not observe who is involved in what, we assume that these unobserved components in outcomes for subgroups of employees are likely to be correlated within the same category of purchase in the bureau. The number of clusters is 2,073.

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 Competence

	Cost Performance					Time Performance				
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Competence (Q28)	-0.02	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			-0.06	-0.06	-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	.13	.14	.14	.14	.15	.11	.12	.12	.12	.12
Observations	122526	122526	122526	122526	122526	122526	122526	122526	122526	122526
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 procurement category and are in parentheses. Amount FEs and Duration FEs represent deciles for contract value and duration. All models include procurement category fixed effects. * 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 contracts that different agencies procure. Adding *Bureau Experience* and *Bureau Size* has, instead, no impact 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).

Despite the inclusion of these controls, a concern with the potential downward bias in the

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	-0.06***		-0.06***	-0.05***		-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 procurement 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 procurement 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.

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.

Table 8: IV regressions

	Cost Performance			Time Performance		
	(1)	(2)	(3)	(4)	(5)	(6)
Competence (Q28)	0.47***	0.25*	0.37***	0.39***	0.32**	0.36***
	(0.11)	(0.15)	(0.10)	(0.10)	(0.14)	(0.09)
Bureau Experience	-0.04	-0.03	-0.04	-0.07	-0.06	-0.06
	(0.03)	(0.03)	(0.03)	(0.04)	(0.04)	(0.04)
Bureau Size	0.00	0.02	0.01	0.01	0.01	0.01
	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
Observations	122526	122526	122526	122526	122526	122526
Weak Id. F-Test	42.73	21.84	30.79	42.73	21.84	30.79
Underid. F-Test	42.37	18.45	55.61	42.37	18.45	55.61
Overid. F-Test	0	0	1.44	0	0	.14
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 procurement category and are in parentheses. All models include controls for contract features (cost plus format and solicitation procedure), buyer characteristics (*Bureau Controls*, i.e. experience and yearly procurement budget), fixed effects for procurement 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 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.

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 0.37 and 0.36 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.[31] 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 (0.15), the estimates are close. 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.

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.[32] This implies a reduction in cost overruns of $120,126 on average per contract, or around $14.7 billions in total across all contracts in the dataset ($2.6 billions on yearly basis). Moreover, this would imply a saving of 39.5 days in effective execution time, corresponding to 4.8 million days across all the contracts in the dataset (841 thousand days on yearly basis). The amounts are economically sizable and compare well to what the literature has indicated

[31] 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.

[32] The bureau at the 90th percentile of the cross-sectional distribution of *competence* is the John Glenn Research Center at Lewis Field.

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 direct 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 (Q28)	-0.63**	-0.29	-0.47**	-0.81**	-0.86**	-0.83***	-1.44**	-1.15*	-1.30***
	(0.30)	(0.27)	(0.23)	(0.32)	(0.39)	(0.29)	(0.58)	(0.63)	(0.49)
Observations	122526	122526	122526	122526	122526	122526	122526	122526	122526
Weak Id. F-Test	42.73	21.84	30.79	42.73	21.84	30.79	42.73	21.84	30.79
Underid. F-Test	42.37	18.45	55.61	42.37	18.45	55.61	42.37	18.45	55.61
Overid. F-Test	0	0	1.02	0	0	.01	0	0	.17

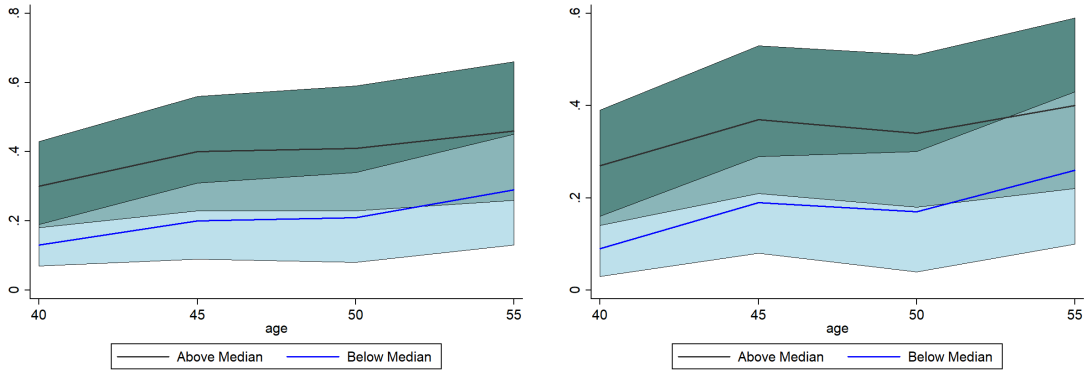
Notes: IV models of columns (1) to (3) of Table 8 are replicated with the number of time renegotiations (columns 1 to 3), the number of cost renegotiations (columns 4 to 6), and the total number of renegotiations (columns 7 to 9) as substitutes for cost performance and time performance. No of time renegotiations 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 procurement 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 procurement 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.

Given the crucial role that 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. Figure 6 plots of how our baseline estimates would differ altering the definition of the relevant deaths instrument.[33] In the baseline estimates, the median values of age and salary are the cutoffs 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 white-collar 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

[33] A more thorough heterogenety analysis was performed in previous versions and is available upon request.

50, the estimates become statistically identical. This evidence is indicative of interesting heterogeneous 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. 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.

Before concluding tis section with an extensive list of robustness checks, we present one final set of results concerning three different outcomes. They are all measures of 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 (40%) and 0.8 (71%) fewer cost renegotiations and time renegotiations, respectively, 1.3 (52%) - fewer in total. Since each negotiation episode is likely to be associated with some waste - i.e.,

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 summary of insights from the main robustness checks among those presented in the appendix. In essence, these additional results exploit further the richness of the three datasets: to address possible concerns on the outcome measure, we calculated additional outcome measures from the FPDS, to confront challenges in measurement and meaning of competence, we exploit additional variables from the FEVS and, finally, to assess the soundness of the identification strategy, we explore alternative definitions of the instrument by relying on the personnel data in the FedScope dataset. Overall, while the qualitative findings prove robust, this additional evidence plays an important role to strengthen the quality and depth of the analysis has certainly improved further. We refer to the appendix for a more exhaustive description of both these robustness checks and the additional ones presented there. To simplify the exposition, we present the findings by categorizing them in three groups.

In an heterogeneity analysis fashion, the first one revolves around whether the findings are driven by specific subset of the data. The two main results are that the effect of competence is stronger on bureaus awarding more complex contracts and when the awarding procedure involves more discretion for the public buyer. Regarding the latter, both the magnitude and the significance of the estimates is indeed larger when the awarding procedure involved is either a Simplified Acquisition Procedure (SAP) or a negotiation. Finally, while one might have suspected an heterogeneous effect of competence depending on the frequency with which a bureau awards contracts, this turns out not to be the case. This finding is in line with the evidence discussed earlier of a low correlation between competence and bureau experience and size.

The second group of robustness checks involves threats to the causal identification of the estimates. First, we assess the reliability of the inference conducted on the estimated parameters by applying the recently proposed method by Young, Alwyn [2017]. Having confirmed that the IV estimates do indeed imply a positive effect of competence, we engage in

a series of robustness checks on the IV strategy itself. Relevance of the instrument, although supported by the statistical evidence presented earlier, might be problematic if timing of the FEVS responses and the incidence of death events are not aligned. We thus consider different time windows and, in addition to confirming the baseline estimates, also discover that the impact of death episodes is greater when they occur in the tender design phase, leading up to the contract award, as opposed to death episodes occurring during the contract execution. This feature also speaks to the analysis of the channels presented in the next section. In terms of relevance, an additional concern is that greater relevance might mean a greater risk of a tautological association between competence and the outcome. This would be the case, for instance, if the bureaus where death events are particularly relevant for competence are also those most specialized in procurement: in this case, the FEVS response measuring competence might reflect perceptions about the procurement outcomes. To address this risk, we repeat the analysis without all contracts of the DOD, DVA, and GSA, finding, however, little changes relative to the baselines estimates. Regarding the exclusion restriction, it is crucial that the definition of competence used is broad enough to encompass all the possible channels through which death-induced shocks might impact procurement outcomes. In this regard, we consider an alternative definition of competence based on a principal component analysis of many FEVS questions capturing an even broad measure of competence relative to that of our baseline measure. Overall, the results hold qualitatively unchanged.

For the second group of robustness checks, we also consider the possibility of alternative strategies to an IV. Given the source of potential omitted variable bias discussed earlier, including bureau fixed effects could go a long way in allowing interpreting OLS estimates as causal. Due to the low within-bureau variability in competence, however, we consider this approach to be not ideal. Thus, while we do report for completeness these estimates - all not significant, - we emphasize an alternative approach based on long run changes:[34] we estimate the effect on end of period performance of beginning of period competence and the within-bureau change in competence between the two sample end-periods. This incorporates the

[34]For both cost and time performance, the fixed effects estimates show a drop in magnitude and a loss of significance relative to the baselines, although for time performance the estimates are borderline weakly significant.

logic of the fixed effects, while accounting for the persistency in competence. The findings from this strategy are in line with the baseline IV estimates in this section.[35]

The third and last group of robustness checks involves alternative outcome measures. Here we gradually expand the analysis, from basic modifications of the performance measures analyzed above, to fully alternative outcome measures. Regarding the former, we present results based on linear time and cost performance indexes, as opposed to the convex ones used above. The findings clearly show the weakness of these alternative measures to outliers, thus supporting the choice of our preferred convex measures. We also consider the problem of what kind of renegotiations enter our measures. In line with other studies, we have considered *in-scope* amendments only.[36] 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.6), we find similar results to those in our baseline estimates.

Finally, we consider expanding the set of outcomes to features of the procurement process that have received attention in the procurement literature. We consider a series of outcomes relative to both the extent of bidders' competition in the procedure and the choice of using different selection procedures. Nevertheless, the estimates do not reveal any systematic association of these outcomes with competence.

[35] Among other robustness checks on the IV strategy presente in the appendix, we also consider two alternative to our 2SLS estimation approach: a limited information maximum likelihood estimator (to account for weak instruments) and the Wooldridge [2002]'s fractional probit model within control function. The latter, is particularly relevant as the particular shape of the distribution of the performance measures (bounded between zero and one and with a mass point at one) might affect our results. In both cases, however, the estimates obtained are very close to our baselines (see Table A.4 and A.5 in the appendix).

[36] 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.

VI 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 10 reports the full list of questions composing the "my work unit" section in the FEVS. 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.

Table 10: 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.

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 illustrate 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.[37] The first four columns of Table 11 show OLS estimates obtained by regressing

[37]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

bureau competence on the three components, first separately and then jointly after collapsing the observations at the 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 appears to be a key driver of bureu 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.

Table 11: 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)
Skills		0.68***		0.09***		0.00		0.05***
		(0.03)		(0.03)		(0.02)		(0.02)
Incentives			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 *cost performance* and *time performance*. The model specification is identical to that in column 5 Table 6, but for the substitution of *competence* with its three components, as detailed in the table.

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 (and the other covariates as in column 5 Table 6). The results are reported in the latter columns of Table 11. 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

and incentives. The last two columns of Table 10 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).

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.

These results on cooperation are well aligned with what is known about, for instance, DoD procurement. Apte, Apte and Rene [2011] run a survey to collect data on management practices in services acquisition in the U. S. Navy. They find that while the organizational structure of the procurement office, being it an individual installation or a larger regional office plays a limited role, management practices do matter, a result also shared by Hyväri [2006]. Among best practices, there is the use of project teams - specifically cross-functional teams - coordinated by a formal project manager to facilitate the proper integration and control of the various functional disciplines involved in the project effort [Rendon et al., 2012].

We fully acknowledge that the result on the role of cooperation is correlational as a credible identification strategy is not feasible given our data. Nevertheless, given its potential novelty and interest, we conclude the analysis with two final considerations based on our data. The first, not surprising one, is that if we substitute the measure of cooperation used above with that from with Q26 ("Employees in my work unit share job knowledge with each other"), the results obtained are qualitatively identical. The second consideration is that, if we follow the public sector management literature by constructing a measure for cooperation based on answers to the FEVS questions on management practices, the resulting measure is strongly associated with both cooperation and procurement outcomes.[38]

[38]The correlation between this new variable and Q20 is 0.9. See Table A.15 in the appendix for details.

VII 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 could be seen as rather surprising for a developed country like the US. They could however be rationalized if we think of what Syverson [2011] finds relative to the intra-industry productivity dispersion in US firms and the fact we can expect even more dispersion when the market is not at play. In this regard, the magnitude of the effect might point at the fact that it is indeed difficult to pass through the best practices to the lesser efficient bureaus and that even in advanced countries there is considerable scope for improving public service provision by investing in the competence of public bureaucracies.

Our second main result, to be taken 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 bureaus' employees appears to be therefore a crucial ingredient for effective procurement. This finding clearly drives policy recommendations on the importance of identifying instruments - including the selection of adequate managers - that can foster cooperative work practices.

References

Acemoglu, Daron. 2005. "Politics and Economics in Weak and Strong States." *Journal of Monetary Economics*, 52: 1199–226.

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.

Apte, A., U. M. Apte, and G. R. Rene. 2011. "Services Supply Chain in the United States Navy: An Empirical Study of Current Management Practices." *Journal of Public Procurement*, 11(3): 538–562.

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.

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, 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.

Bartlett, MS. 1950. "Periodogram analysis and continuous spectra." *Biometrika.*

Becker, Sascha O., and Hans K. Hvide. 2013. "Do Entrepreneurs Matter?" *CESifo Working Paper Series N. 4088.*

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

Bertrand, Marianne, and Sendhil Mullainathan. 2001. "Do People Mean What They Say? Implications for Subjective Survey Data." *American Economic Review*, 91(2): 67–72.

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.

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.*

Beuve, Jean, Marian W. Moszoro, and Stéphane Saussier. 2019. "Political contestability and public contract rigidity: An analysis of procurement contracts." *Journal of Economics & Management Strategy*, 28(2): 316–335.

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.

Branzoli, Nicola, and Francesco Decarolis. 2015. "Entry and Subcontracting in Public Procurement Auctions." *Management Science*, 61(12): 2945–2962.

Bruce, Joshua R., John M. de Figueiredo, and Brian S. Silverman. 2019. "Public contracting for private innovation: Government capabilities, decision rights, and performance outcomes." *Strategic Management Journal*, 40(4): 533–555.

Bucciol, A., R. Camboni, and P. Valbonesi. 2017. "Buyers' Ability in Public Procurement: A Structural Analysis of Italian Medical Devices." *Working Paper n.17, Department of Economics, University of Verona.*

Cattaneo, M. D., R. K. Crump, M. H. Farrell, and Y. Feng. 2019. "On Binscatter." *arXiv.*

Chassang, Sylvain, and Juan Ortner. 2019. "Collusion in Auctions with Constrained Bids: Theory and Evidence from Public Procurement." *Journal of Political Economy*, 5(127).

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.

Conley, Timothy G., and Francesco Decarolis. 2016. "Detecting bidders groups in collusive auctions." *American Economic Journal: Microeconomics*, 8(2): 1–38.

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).

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. 2018. "Comparing Public Procurement Auctions." *International Economic Review*, 59(2): 391–419.

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.*

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.

Gibbons, Robert, and Rebecca Henderson. 2016. "What do managers do? Exploring persistent performance differences among seemingly similar enterprises." In *The Handbook of Organizational Economics*. , ed. R. Gibbons and J. Roberts, Chapter 28, 391–397. Princeton:Princeton University Press.

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

Giustinelli, Pamela, and Charles F. Manski. 2018. "Survey Measures of Family Decision Processes for Econometric Analysis of Schooling Decisions." *Economic Inquiry*, 56(1): 81–99.

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.

Hyväri, I. 2006. "Success of Projects in Different Organizational Conditions." *Project Management Journal*, 37(4): 31–41.

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*.

Janke, Katharina, Carol Propper, and Raffaella Sadun. 2019. "The Impact of CEOs in the Public Sector: Evidence from the English NHS." National Bureau of Economic Research Working Paper 25853.

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).

Jung, Hojin, Georgia Kosmopoulou, Carlos Lamarche, and Richard Sicotte. 2018. "Strategic Bidding and Contract Renegotiation." *International Economic Review*, forthcoming.

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. "Contractor's decision for bid profit reduction within opportunistic bidding behavior of claims recovery." *International Journal of Project Management*, 29(1): 93 – 107.

OECD. 2017. "Government at a Glance 2017." OECD.

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(1–2): 121 – 133.

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

Rendon, Rene G., Uday M. Apte, , and Aruna Apte. 2012. "Services Acquisition in the DoD: A Comparison of Management Practices in the Army, Navy, and Air Force." *ARJ*, 19(1): 3–32.

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.

Syverson, Chad. 2011. "What Determines Productivity?" *Journal of Economic Literature*, 49: 326–365.

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.

Weber, Max. 1922. *Economy and Society: An Outline of Interpretive Sociology.* University of California Press.

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

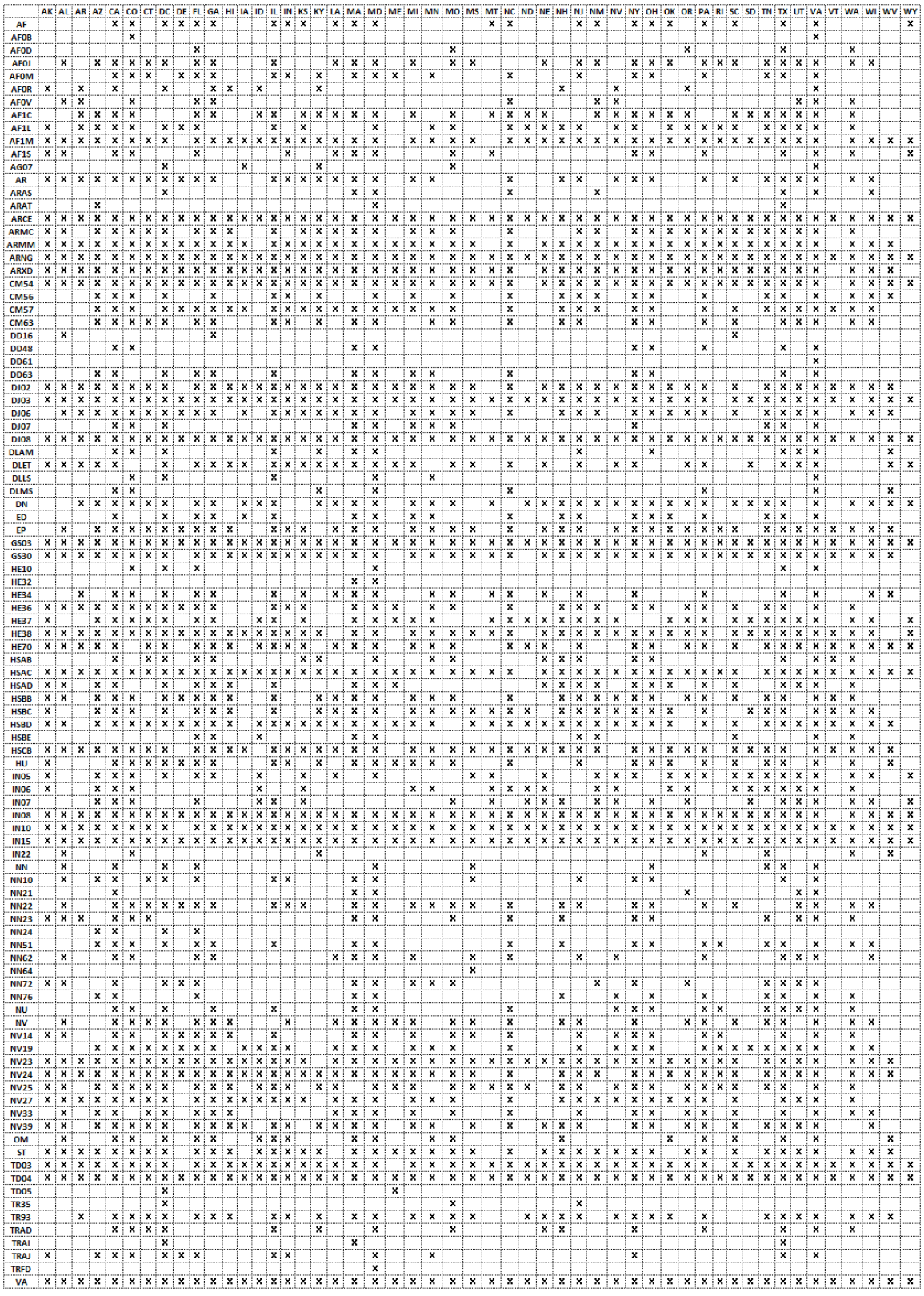
For the purpose of our analysis, we will focus on the years where the FEVS has an yearly frequency and where the two datasets, FEVS and FPDS, overlap. Thus, we focus on the years 2010-2015. Although the data contain contracts for supplies, R&D projects, services, and works, the first two are ruled out 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.[39] The same applies to the subcategory "Lease or Rental of Equipment, Structures, or Facilities". Thus, for our analysis we focus exclusively on the procurement of services and works.[40] We restrict our sample to those contracts awarded via competitive solicitations as the effect of the treatments would otherwise not be observable. We consider as competitive those for which the extent of competition is labelled "Full and open"; those whose participation is not set aside to any specific set 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 a transaction value of $3,000. We focus on contracts worth more than $25,000.[41] In non-competitive awardings, the participation criteria restrict the competition *ex-ante* to dimensions other than quality.

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

[40] 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.

[41] 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. 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. Table 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. The full name of bureaus present in the resulting dataset is presented in Table A.1 below.

Table A.1: List of Bureaus

COD	BUREAU	COD	BUREAU
AF	DEPT OF THE AIR FORCE	HSAC	U.S. COAST GUARD
AF0B	U.S. AIR FORCE ACADEMY	HSAD	U.S. SECRET SERVICE
AF0J	AIR EDUCATION AND TRAINING COMMAND	HSBB	U.S. IMMIGRATION AND CUSTOMS ENFORCEMENT
AF0M	HEADQUARTERS, AIR FORCE RESERVE	HSBC	TRANSPORTATION SECURITY ADMINISTRATION
AF0R	PACIFIC AIR FORCES	HSBD	U.S. CUSTOMS AND BORDER PROTECTION
AF0V	AIR FORCE SPECIAL OPERATIONS COMMAND	HSBE	FEDERAL LAW ENFORCEMENT TRAINING CENTER
AF1C	AIR COMBAT COMMAND	HSCB	FEDERAL EMERGENCY MANAGEMENT AGENCY
AF1L	AIR MOBILITY COMMAND	HU	HOUSING AND URBAN DEVELOPMENT, DEPARTMENT OF
AF1M	AIR FORCE MATERIEL COMMAND	IN05	BUREAU OF LAND MANAGEMENT
AF1S	SPACE COMMAND	IN06	BUREAU OF INDIAN AFFAIRS & EDUCATION
AG07	RURAL HOUSING SERVICE	IN07	BUREAU OF RECLAMATION
AR	DEPT OF THE ARMY	IN08	US GEOLOGICAL SURVEY
ARAS	U.S. ARMY INTELLIGENCE AND SECURITY COMMAND	IN10	NATIONAL PARK SERVICE
ARAT	US ARMY TEST AND EVALUATION COMMAND	IN15	U.S. FISH AND WILDLIFE SERVICE
ARCE	U.S. ARMY CORPS OF ENGINEERS	IN22	OFFICE OF SURFACE MINING, RECLAMATION & ENFORCEMENT
ARMC	U.S. ARMY MEDICAL COMMAND	NN	NATIONAL AERONAUTICS AND SPACE ADMINISTRATION
ARMM	U.S. ARMY MATERIAL COMMAND	NN10	NASA HEADQUARTERS
ARNG	ARMY NATIONAL GUARD UNITS	NN21	AMES RESEARCH CENTER
ARXD	U. S. ARMY CONTRACTING COMMAND	NN22	JOHN GLENN RESEARCH CENTER
CM54	NATIONAL OCEANIC AND ATMOSPHERIC ADMINISTRATION	NN23	LANGLEY RESEARCH CENTER
CM56	US PATENT AND TRADEMARK OFFICE	NN24	DRYDEN FLIGHT CENTER
CM57	NATIONAL INSTITUTE OF STANDARDS AND TECHNOLOGY	NN51	GODDARD SPACE FLIGHT CENTER
CM63	US CENSUS BUREAU	NN62	GEORGE C. MARSHALL SPACE FLIGHT CENTER
DD16	DEPARTMENT OF DEFENSE EDUCATION ACTIVITY	NN64	JOHN C. STENNIS SPACE CENTER
DD48	DEFENSE HUMAN RESOURCES ACTIVITY	NN72	LYNDON B. JOHNSON SPACE CENTER
DD61	DEFENSE THREAT REDUCTION AGENCY	NN76	JOHN F. KENNEDY SPACE CENTER
DD63	DEFENSE CONTRACT MANAGEMENT AGENCY	NU	NUCLEAR REGULATORY COMMISSION
DJ02	FEDERAL BUREAU OF INVESTIGATION	NV	DEPT OF THE NAVY
DJ03	FEDERAL PRISON SYSTEM / BUREAU OF PRISONS	NV14	OFFICE OF NAVAL RESEARCH
DJ06	DRUG ENFORCEMENT ADMINISTRATION	NV19	NAVAL AIR SYSTEMS COMMAND
DJ07	OFFICE OF JUSTICE PROGRAMS	NV23	NAVAL SUPPLY SYSTEMS COMMAND
DJ08	U.S. MARSHALS SERVICE	NV24	NAVAL SEA SYSTEMS COMMAND
DLAM	OFFICE OF THE ASSISTANT SECRETARY FOR ADMIN & MGMNT	NV25	NAVAL FACILITIES ENGINEERING COMMAND
DLET	EMPLOYMENT AND TRAINING ADMINISTRATION	NV27	U.S. MARINE CORPS
DLLS	BUREAU OF LABOR STATISTICS	NV33	MILITARY SEALIFT COMMAND
DLMS	MINE SAFETY AND HEALTH ADMINISTRATION	NV39	SPACE AND NAVAL WARFARE SYSTEMS COMMAND
DN	ENERGY, DEPARTMENT OF	OM	OFFICE OF PERSONNEL MANAGEMENT
ED	EDUCATION, DEPARTMENT OF	ST	STATE, DEPARTMENT OF
EP	ENVIRONMENTAL PROTECTION AGENCY	TD03	FEDERAL AVIATION ADMINISTRATION
GS03	PUBLIC BUILDINGS SERVICE	TD04	FEDERAL HIGHWAY ADMINISTRATION
GS30	FEDERAL ACQUISITION SERVICE	TD05	FEDERAL RAILROAD ADMINISTRATION
HE10	OFFICE OF THE SECRETARY OF HEALTH AND HUMAN SERVICES	TR35	OFFICE OF THRIFT SUPERVISION
HE34	HEALTH RESOURCES AND SERVICES ADMINISTRATION	TR93	INTERNAL REVENUE SERVICE
HE36	FOOD AND DRUG ADMINISTRATION	TRAD	UNITED STATES MINT
HE37	INDIAN HEALTH SERVICE	TRAI	BUREAU OF ENGRAVING AND PRINTING
HE38	NATIONAL INSTITUTES OF HEALTH	TRAJ	OFFICE OF THE COMPTROLLER OF THE CURRENCY
HE70	CENTERS FOR MEDICARE AND MEDICAID SERVICES	TRFD	BUREAU OF THE FISCAL SERVICE
HSAB	U.S. CITIZENSHIP AND IMMIGRATION SERVICES	VA	VETERANS AFFAIRS, DEPARTMENT OF

II Robustness Checks

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

- In Table A.2, we show how estimates change relative to our baseline from Table 8 when we modify a few elements of the analysis. First, in column 1 we verify that the findings are not driven by outliers by repeating the analysis after dropping the most extreme observations either in terms of cost or time performance (i.e., those exceeding the contractually agreed duration or cost by four times). The following column considers a sample of "competitive tenders", i.e. those that receive at least two bids. This is not surprising as the effect of competence should matter more when the buyer can select among multiple bidders, but the channel could also be that more competent buyers are more effective in inducing participation.[42] Column 3, implements a specification where bureau-level fixed effects replace the agency-level ones used in our baselines. This is the specification that we had ideally liked to implement if the variability over time had allowed us to do so, as a within-bureau strategy would have avoided altogether the need for instruments. Yet, the competence measures are very persistent within bureaus over time, as discussed earlier. Therefore, it is not surprising to see in column 3 that for both cost and time performance, the estimates show a drop in magnitude and a loss of significance relative to the baselines, although for time performance the estimates are borderline weakly significant. Column 4 introduces an important sample restriction to assess the concern that our estimates are mechanically showing the relationship between two proxies of procurement outcomes. This would happen if the FEVS respondents were basing their answers on the same procurement outcomes that we look at. Although the broad dimension and composition of the FEVS respondents should make this risk minimal, there are some bureaus where the share of employees involved with contracting is quite large. Since these are the bureaus for which the concerns of a mechanical effect is larger - most of their budget is spent on procurement activities -, we repeat the analysis having dropped them. We thus rule out all contracts of the DOD, DVA, and GSA. But the estimates in column 4 show that qualitatively little changes relative to our baselines. Nevertheless, to further investigate the same concern, we also implement a different strategy whose results are reported in

[42]This is in line with the model of Kang and Miller [2017]. In practice, this could take place through the choice of the tendering structure, as in Branzoli and Decarolis [2015].

column 5. There we replace our measure of competence with its lagged value. But once again the results are qualitatively similar and, if anything, stronger in terms of both magnitude and significance. Finally, in column 6 and 7, we present weighted versions for our regressions where we try to address the issue of the heterogeneity between bureaus in a different way relative to the baselines. Here we use weights that use the propensity score of our instruments, separately, on procurement-related characteristics of the bureau, that is the percentage of number of procurers over the total bureau workforce and the number of contracting offices within the bureau. With different inverse probability weights associated to our instruments, in column 6 we replicate columns 1 and 4 of Table 8 in the top panel and the bottom panel, respectively; in column 7 we replicate columns 2 and 5 of Table 8 in the top panel and the bottom panel, respectively. All estimates confirm the qualitative results of the baselines and also indicate that giving a greater weight to larger offices produces larger estimates for the role of competence. We take this as an indication that our baseline estimates are likely to be a conservative measure of the effect of competence that in these estimates is watered down by the substantial heterogeneity across bureaus.

- Table A.4: 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. LIML is a linear combination of the OLS and 2SLS estimate with the weights designed to approximately eliminate the 2SLS bias. All the point estimates are identical to those in Table 8, thus limiting concerns about a weak instruments problem. In our data the weights happen to be 0 to OLS and 1 to 2SLS, i.e. there is no 2SLS bias.

- Table A.5: control function estimates. Since $performance^{C}_{ijt}$ and $performance^{T}_{ijt}$ are fractional variables on (0,1][43] 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 $\mathbb{E}(y|\mathbf{x}) = \Phi(\mathbf{x}\gamma)$.[44] This fractional probit model

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

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

handles continuous endogenous explanatory through a *two-step control function approach*. The control function approach relies on similar identification conditions of the linear IV described in the main text.[45] Table A.5 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.[46]

- Table A.6: alternative measurement of procurement performance. The estimates in Table A.6 are analogous to our baseline estimates, but 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.6.

- 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.7 we replicate our IV analysis by drawing 500 bootstrap samples in a fashion consistent with the error dependence within our cluster of observations (bureau

[45] 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 $competence_{jt} = f(\mathbf{X}; o_{jt})$, $o_{jt} = \rho\, instrjt + \epsilon_{jt}$ 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 $res_{jt} = competence_{jt} - \hat{competence}_{jt}$. 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. $\mathbb{E}(performance_{ijt}|competence_{jt}, \hat{cf}_{jt}, \mathbf{X}) = \Phi(\beta\, competence_{jt} + \zeta\, \hat{cf}_{jt} + \theta\, X_{ijt} + \iota_j + \kappa_t)$.

[46] 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.

and procurement category) and independence across observations. This method produces estimates that identify our parameters of interest as accurately as the baseline IV. Indeed, 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.

- In our empirical strategy, the identification of the effect of bureau competence relies on our endogenous independent variable to be an overall and comprehensive measure of quality/capacity of the work of those offices. In other words, Q28 should consider and be affected by any possible mechanism that an exogenous shock to the labor force (i.e. our two instruments) can trigger. To show further evidence on the reliability of our instrumental variable approach, we constructed a synthetic measure of competence, by using all the available information in the FEVS survey, although excluding Q28. The questions in the FEVS survey are undoubtedly comprehensive and cover extensively any organizational aspect or feature of the labor force in an office. We drop all the questions with an excessive number of missing data and the questions without a direct connection with the organizational aspects or features of the labor force in an office. We also drop question 27 ("The skill level in my work unit has improved in the past year") due to its reference to past levels of a feature in an office. Similarly, we drop questions that relate to the agency levels (questions from 29 to 41). In Table A.3, we report the list of the 39 questions employed to construct the synthetic measure. We perform a PCA analysis to identify a vector of weights to construct the unique synthetic measure. The PCA analysis shows that only two factors have eigenvalue above the usual rule of thumb of 1 (the second is slightly above 1). Moreover, the first factor has an eigenvalue exceeding the second by 27.71 and explains 88% of the entire variance. This strongly suggests the existence of a unique factor that we can interpret as a measure of quality/capability. The synthetic measure is strongly correlated with q28 (i.e. a Pearson correlation of 71%). The IV regressions using this synthetic measure of competence are reported in Table A.9. We found results consistent with the IVs using Q28.

- Given the strong persistence of our measure of competence, a specification with bureau fixed effects yields statistically insignificant results, see Column 3 of panel a in Table A.2. To address this issue, we implement a long-run differences specification, as an alternative to the bureau fixed-effect model, by regressing the contract performance measures in the last available years on their 3-years lags and the 3-years changes in competence [47]. We also instrument the changes in competence between the two periods. In principle, all the death events between the first and last periods are feasible, but to maximize the statistical power we use cross-validation to select the set of instruments that perform better in terms of weak instruments statistics. The model is a cross-bureau specification on the changes rather than the levels:

$$performance^{g}_{jkct} = \rho performance^{g}_{jkc(t-3)} + \beta \Delta_3 competence_{jt} + \theta \mathbf{X_j} + \kappa_k + \zeta_z + \tau_t + \varepsilon_{jkct} \tag{3}$$

  In order to match our most detailed level of office-level employment variation, the model can be estimated only after collapsing the performance measures at the bureau, procurement category, U.S. state of performance, and year level. The data are collapsed by using the mean statistic and we set the depth of our procurement category fixed effect at the second digit level, to reduce the missing periods in each panel. In Table A.16, we report the results of this long-run iv regressions. In columns 1 and 3, we report the results respectively for cost and time renegotiations without controlling for the characteristics of contracts. In columns 2 and 4, we use as dependent variables the residuals from a regression of the outcomes variables on the characteristics of contracts We find that a rise of a standard deviation (equal to 1.2) in the difference between the contemporaneous competence and its 3-years lag increases the average (residualized) cost (time) performances by 0.4 standard deviation. Our estimates suggest a persistent effect of our measure of competence on the long run performance of bureaus.

- In Table A.15, we address possible concerns about the timing of a death event. We

[47] A 3-years lag is a good compromise between the temporal depth required by this analysis and the number of available observations.

construct a new instrument that counts if a death event occurred in the same state of performance of a contract in the six months before (and, as a sensitivity check, after) the signature of the contract itself. According to the managerial literature, six months is a sensitive period for newly hired employees to gain full efficiency, the so-called 'onboarding effect, which avoids possible confounding effects due to the newly hired employees. To avoid any confounding effects due to cumulative death events, we follow Jäger [2017] and restrict our instrument to count single death events in a local office. Moreover, we address concerns about possible changes in workload by including as included instrument a measure of the local white-collar workforce. FedScope snapshots are taken in September, while FEVS in June. To account for any variation in the employment stock owing to the death occurrences before September of the same year, for contracts signed up to September, we substitute the employment stock with its lag that is unaffected by those changes. We include the z-score of this measure in our regressions. In Table A.15, we report the IV estimates using the 6 months before (after) instruments. The estimates in columns 1 and 3 for cost and time performances confirm that one standard deviation increase in competence amounts to an improvement in the cost performance of 57 percent of a standard deviation of cost performance (28 percent in the case of time performance). In columns 2 and 4, we find that the 6 months after dummy does not pass the standard test as a good instrument for competence. We believe this represents strong evidence for the important role that bureaus play in the period before the signature of a contract. In columns 5 and 7, we add the workforce control and we find that our instruments generally improve (we find higher values of the RK f tests than the ones in columns 1 and 3). The estimates are more in line with our baseline. In columns 6 and 8, we also add firm fixed effects to check whether the effect is driven by firm selection. Results are qualitatively similar but to a lower level of statistical significance. We conclude that the post awarding phase results not so relevant.

- In Table A.18, we reproduce our baseline estimates including two different measures of variation of the workforce as possible controls. Columns 1 and 2 replicate for cost

and time performances the estimates in Table 8. In column 3 and 4, the first measure is the z-score of the available amount of white-collars in the bureau before the yearly survey is administered. In columns 5 and 6, the second measure is the z-score of the yearly change in the number of white-collars. Across specifications, we do not find any sensible change in the impact of competence.

Table A.2: Robustness Checks: Sample Selection

Panel A: Cost Performance

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Competence (Q28)	0.14**	0.36***	0.03	0.24**		0.33***	0.80*
	(0.06)	(0.11)	(0.13)	(0.10)		(0.12)	(0.41)
Lagged Competence					0.48***		
					(0.15)		
Observations	102061	74711	122526	54427	90127	81818	34978
Weak Id. F-Test	28.86	26.6	11.39	24.61	14.14	35.65	6.14
Underid. F-Test	53.41	52.19	21.08	41.9	27.87	35.5	5.79
Overid. F-Test	2.87	1.04	11.19	.15	0	0	0

Panel B: Time Performance

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Competence (Q28)	0.16***	0.33***	0.22*	0.13		0.25**	0.68*
	(0.06)	(0.11)	(0.14)	(0.09)		(0.12)	(0.36)
Lagged Competence					0.51***		
					(0.14)		
Observations	102061	74711	122526	54427	90127	81818	34978
Weak Id. F-Test	28.86	26.6	11.39	24.61	14.14	35.65	6.14
Underid. F-Test	53.41	52.19	21.08	41.9	27.87	35.5	5.79
Overid. F-Test	.46	.07	3.91	3.95	.70	0	0

Notes: The table presents the results of applying a series of modifications to the baseline estimates of Table 8. In column 1, we exclude contracts with cost and time performance lower than 0.25, respectively. Column 2 restricts the sample to tenders receiving at least two offers. Column 3 presents results with bureau fixed effects, instead of agency fixed effects. In column 4, we discard all contracts held by DOD, DVA, and GSA. Column 5 replaces *Competence* with its lagged value. In column 6 and 7, we present weighted regressions were the weights use the propensity score of our instruments on procurement-related characteristics of the bureau: i) percentage of number of procurers over the total bureau workforce and ii) number of contracting offices within the bureau. * Significant at the 10 percent level; ** Significant at the 5 percent level; *** significant at the 1 percent level.

Table A.3: FEVS Questions Part of the PCA for Cooperation

N	Question	N	Question
1	I am given a real opportunity to improve my skills in my organization.	46	My supervisor provides me with constructive suggestions to improve my job performance.
2	I have enough information to do my job well.	47	Supervisors in my work unit support employee development.
3	I feel encouraged to come up with new and better ways of doing things.	51	I have trust and confidence in my supervisor.
4	My work gives me a feeling of personal accomplishment.	53	In my organization, senior leaders generate high levels of motivation and commitment in the workforce.
5	I like the kind of work I do.	54	My organization's senior leaders maintain high standards of honesty and integrity.
9	I have sufficient resources (for example, people, materials, budget) to get my job done.	55	Supervisors work well with employees of different backgrounds.
10	My workload is reasonable.	56	Managers communicate the goals and priorities of the organization.
11	My talents are used well in the workplace.	57	Managers review and evaluate the organization's progress toward meeting its goals and objectives.
16	I am held accountable for achieving results.	58	Managers promote communication among different work units (for example, about projects, goals, needed resources).
18	My training needs are assessed.	60	Overall, how good a job do you feel is being done by the manager directly above your immediate supervisor?
20	The people I work with cooperate to get the job done.	61	I have a high level of respect for my organization's senior leaders.
21	My work unit is able to recruit people with the right skills.	63	How satisfied are you with your involvement in decisions that affect your work?
22	Promotions in my work unit are based on merit.	64	How satisfied are you with the information you receive from management on what's going on in your organization?
23	In my work unit, steps are taken to deal with a poor performer who cannot or will not improve.	65	How satisfied are you with the recognition you receive for doing a good job?
24	In my work unit, differences in performance are recognized in a meaningful way.	67	How satisfied are you with your opportunity to get a better job in your organization?
25	Awards in my work unit depend on how well employees perform their jobs.	68	How satisfied are you with the training you receive for your present job?
26	Employees in my work unit share job knowledge with each other.	69	Considering everything, how satisfied are you with your job?
42	My supervisor supports my need to balance work and other life issues.	70	Considering everything, how satisfied are you with your pay?
43	My supervisor provides me with opportunities to demonstrate my leadership skills.		
44	Discussions with my supervisor about my performance are worthwhile.		

Table A.4: IV-LIML regressions

	Cost Performance			Time Performance		
	(1)	(2)	(3)	(4)	(5)	(6)
Competence (Q28)	0.47***	0.25*	0.37***	0.39***	0.32**	0.36***
	(0.11)	(0.15)	(0.10)	(0.10)	(0.14)	(0.09)
Bureau Experience	-0.04	-0.03	-0.04	-0.07	-0.06	-0.06
	(0.03)	(0.03)	(0.03)	(0.04)	(0.04)	(0.04)
Bureau Size	0.00	0.02	0.01	0.01	0.01	0.01
	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
Observations	122526	122526	122526	122526	122526	122526
Weak Id. F-Test	42.73	21.84	30.79	42.73	21.84	30.79
Underid. F-Test	42.37	18.45	55.61	42.37	18.45	55.61
Overid. F-Test	0	0	1.43	0	0	.14

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 procurement 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 procurement 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: Control Function Estimates

	Cost Performance			Time Performance		
	(1)	(2)	(3)	(6)	(7)	(8)
Competence (Q28)	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.6: Alternative Performance Measures: IV Estimates

	Cost Performance			Time Performance		
	(1)	(2)	(3)	(4)	(5)	(6)
Competence (Q28)	0.46***	0.23	0.35***	0.37***	0.31**	0.34***
	(0.11)	(0.15)	(0.10)	(0.10)	(0.14)	(0.09)
Observations	122326	122326	122326	122326	122326	122326
Weak Id. F-Test	42.62	21.83	30.76	42.62	21.83	30.76
Underid. F-Test	42.28	18.45	55.56	42.28	18.45	55.56
Overid. F-Test	.	.	.	.	.	.

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 procurement 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 procurement 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.7: IV regressions - Cluster Bootstrap

	Cost Performance			Time Performance		
	(1)	(2)	(3)	(4)	(5)	(6)
Competence (Q28)	0.47***	0.25*	0.37***	0.39***	0.32***	0.36***
	(0.09)	(0.14)	(0.08)	(0.10)	(0.12)	(0.09)
Observations	122526	122526	122526	122526	122526	122526

Notes: Results from Table 8 are replicated with standard errors - in parentheses - clustered by bureau and procurement 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 procurement 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.8: IV regressions with Death Predictors

	Cost Performance			Time Performance		
	(1)	(2)	(3)	(4)	(5)	(6)
Competence (Q28)	0.94***	0.25*	0.62**	0.73***	0.32**	0.42*
	(0.29)	(0.15)	(0.26)	(0.28)	(0.14)	(0.23)
Observations	122526	122526	122526	122526	122526	122526
Weak Id. F-Test	20.18	21.84	11.36	20.18	21.84	11.36
Underid. F-Test	18.06	18.45	19.32	18.06	18.45	19.32
Overid. F-Test	0	0	2.3	0	0	3.58
Bureau Controls	Yes	Yes	Yes	Yes	Yes	Yes
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: We replicate Table 8, by adding as further controls those variable from Table 5 that predict a death event at 95% at least, i.e. *No. of Contracts*, *Good Place to Work*, and *Organization Satisfaction*. Please note that those predictors are only significant for *Proximal Death* and we therefore use them as included instruments only in Columns (1), (3), (4), and (6).

Table A.9: IV regressions - Alternative Competence PCA

	OLS		IV					
	Cost Performance	Time Performance	Cost Performance			Time Performance		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Alternative Competence	0.04***	0.03***	0.57***	0.20*	0.31***	0.47***	0.25**	0.32***
	(0.01)	(0.01)	(0.15)	(0.11)	(0.10)	(0.13)	(0.11)	(0.09)
Bureau Experience	-0.02	-0.05	-0.07	-0.04	-0.05	-0.09	-0.07	-0.08
	(0.03)	(0.04)	(0.07)	(0.04)	(0.05)	(0.07)	(0.06)	(0.06)
Bureau Size	0.04***	0.04**	0.04*	0.04**	0.04**	0.03	0.04*	0.04*
	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
Observations	122526	122526	122526	122526	122526	122526	122526	122526
Weak Id. F-Test	.	.	23.38	14.72	14.88	23.38	14.72	14.88
Underid. F-Test	.	.	24.34	11.42	29.58	24.34	11.42	29.58
Overid. F-Test	.	.	0	0	4.98	0	0	2.02
Amount FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Duration FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Agency FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
State FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Notes: We replicate our baseline empirical analysis using a synthetic measure of *Competence* alternative to the question 28. Columns (1) and (2) report OLS estimates. Columns (3) and (6) report IV with *Proximal Deaths*; columns (4) and (7) report IV with *Relevant Deaths*; columns (5) and (8) 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 procurement category and are in parentheses. Amount FEs and Duration FEs represent deciles for contract value and duration. All models include procurement category fixed effects. * Significant at the 10 percent level; ** Significant at the 5 percent level; *** Significant at the 1 percent level.

Table A.10: Robustness Check: Outcome Specification

Panel A: Alternative Outcome Specification

	Cost Performance			Time Performance		
	(1)	(2)	(3)	(4)	(5)	(6)
Competence (Q28)	-0.08	-0.10	-0.09**	0.03	-0.09	-0.03
	(0.06)	(0.07)	(0.04)	(0.06)	(0.10)	(0.06)
Observations	122526	122526	122526	122526	122526	122526
Weak Id. F-Test	42.73	21.84	30.79	42.73	21.84	30.79
Underid. F-Test	42.37	18.45	55.61	42.37	18.45	55.61
Overid. F-Test	0	0	.02	0	0	1.22

Panel B: Alternative Outcome Specification (Winsored 1st-99th pc.)

	Cost Performance			Time Performance		
	(1)	(2)	(3)	(4)	(5)	(6)
Competence (Q28)	-0.41***	-0.24*	-0.33***	-0.30***	-0.23**	-0.27***
	(0.10)	(0.12)	(0.09)	(0.10)	(0.10)	(0.07)
Observations	121300	121300	121300	121300	121300	121300
Weak Id. F-Test	42.14	22	30.37	42.1	21.68	30.38
Underid. F-Test	41.92	18.62	55.07	41.78	18.36	55.02
Overid. F-Test	0	0	1.37	0	0	.28

Panel C: Baseline Outcome Specification (Winsored 1st-99th pc.)

	Cost Performance			Time Performance		
	(1)	(2)	(3)	(4)	(5)	(6)
Competence (Q28)	0.44***	0.21	0.33***	0.40***	0.34**	0.37***
	(0.11)	(0.14)	(0.09)	(0.10)	(0.14)	(0.09)
Observations	121300	121300	121300	121300	121300	121300
Weak Id. F-Test	42.14	22	30.37	42.1	21.68	30.38
Underid. F-Test	41.92	18.62	55.07	41.78	18.36	55.02
Overid. F-Test	0	0	1.85	0	0	.16
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: In Panel A, We replicate Table 8 and replace the structure of our outcomes with the linear function '(Exp. Outcome + Overrun)/Exp Outcome'. In Panel B, we also replicate Table 8 and replace the structure of our outcomes with the linear function '(Exp. Outcome + Overrun)/Exp Outcome'. We trimmed our sample to exclude observations linked to either delays or extra cost below the 1st percentile and above the 99th percentile of the respective distributions. In Panel C, we replicate Table 8 after trimming our sample to exclude observations linked to either delays or extra cost below the 1st percentile and above the 99th percentile of the respective distributions. 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 procurement 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.

Table A.11: Alternative Outcomes

	Cost Perf.	Time Perf.	Log(# of Offers Received)		At least 2 Offers Received		Competition: Open vs. Restricted		Auction (vs. Negotiation)	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Competence (Q28)	0.35**	0.35**	0.06	-0.03	0.05	0.03***	-0.01	-0.01	-0.10**	-0.01
	(0.16)	(0.15)	(0.46)	(0.08)	(0.04)	(0.01)	(0.05)	(0.01)	(0.05)	(0.01)
Observations	106029	106029	122526	122526	122526	122526	122519	122519	36009	36009
Weak Id. F-Test	18.44	18.44	30.79		30.79		30.8		21.55	
Underid. F-Test	35.94	35.94	55.61		55.61		55.62		32.18	
Overid. F-Test	3.31	.1	0		.01		5.13		2.4	
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	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
State FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Firm FEs	Yes	Yes	No	No	No	No	No	No	No	No

Notes: Column 1 and 2 report our baseline empirical analysis (i.e. IV regression of contract outcomes on *Competence* - instrumented by *Proximal Death* and *Relevant Death*) augmented by firm fixed effects. Starting with that, Odd (even) columns report the results of an IV (OLS) estimation of our baseline model by using the same variables (i.e., *Proximal Deaths* and *Relevant Deaths*) to instrument the same endogenous variable of interest (i.e., *Competence*) yet with alternative tender level outcomes: the log-number of offers received - Columns (1) and (2)-, a dummy for at least two offers received - Columns (3) and (4)-, a dummy indicating whether no sources have been excluded prior to requesting proposals/quotes - Columns (5) and (6)- and a dummy indicating whether the solicitation procedures employed is sealed-bid auction (vs. direct negotiations between buyers/sellers - Columns (7) and (8). In the latter case, the number of observations is lower because the US contracting officers can rely on other solicitation procedures, such as Sole Source (i.e. direct award), Two-step (i.e. a combination of sealed bids and negotiated procedures) or Simplified-Acquisition (i.e. Simplified Acquisition Procedures according FAR Part 13). The outcomes employed are in the spirit of Carril and Duggan (2018) that explore the effect of market concentration on the degree of competition in the procurement process of DOD. Their measure of competition, like us, is measured by the intensity of single-bid tenders and the extent of full and open competition. In addition, we test the effect of *competence* on the competitive-entwined procedure of sealed-bid auctions. Unlike them, we cannot test the impact of competence on the contract-format as we only rely on fixed-price contracts in this study.
Both contract outcomes and bureau characteristics are replaced by their standard scores. Standard errors are clustered by bureau and procurement 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.

Table A.12: Solicitation Procedures

	Solicitation Procedures
Alternative Sources	129
Architect - Engineer	2406
Negotiated Proposal	30970
Simplified Acquisition Procedures	27088
Sealed Bid	5156
Two Step	761
Other	56023
Observations	122533

Notes: Tabulation of solicitation formats employed to award contract. *Alternative Sources*: code if the action resulted from the use of procedures that provide for full and open competition after exclusion of sources. *Architect - Engineer*: code if the action resulted from selection of sources for architect-engineer contracts (FAR 6.102). *Negotiated Procedure*: code for contract award using negotiated procedures (FAR 12, FAR, 13, FAR 15). *Simplified Acquisition Procedures*: code for an acquisition when the Simplified Acquisition Procedures in FAR 13 are used. *Sealed Bid*: code for contract award using seal bidding procedures (FAR 14). *Single Source*: only one source, i.e. no solicitation procedure was used for this action. *Two Step*: code for contract awards using a combination of sealed bids and negotiated procedures (FAR 6.102).

Table A.13: Competence and Solicitation Procedures

	Cost Performance					Time Performance				
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Competence (Q28)	0.36***	0.37***	0.28	0.07	0.13	0.36***	0.34***	0.34**	-0.03	0.05
	(0.10)	(0.12)	(0.17)	(0.12)	(0.12)	(0.09)	(0.11)	(0.16)	(0.30)	(0.31)
Observations	122526	27088	30970	5156	4758	122526	27088	30970	5156	4758
Weak Id. F-Test	31.29	19.54	15.56	4.43	4.33	31.29	19.54	15.56	4.43	4.33
Underid. F-Test	56.25	34.3	25.87	7.02	7.16	56.25	34.3	25.87	7.02	7.16
Overid. F-Test	1.06	.07	.01	.15	.34	.02	3	.87	.21	2.31
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	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
State FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Notes: We replicate our baseline empirical analysis (i.e. IV regression of contract outcomes on *Competence* - instrumented by *Proximal Death* and *Relevant Death*) for subsets of contracts depending on the solicitation procedure employed to award the project. Columns 1 and 6 report our baseline results, i.e. coefficients from columns 3 and 6 of Table 8. Columns 2 and 7 focus on Simplified Acquisition Procedures, 3-8 on Negotiated Proposals, 4 and 9 on sealed bid auctions, 5 and 10 on sealed bid auctions with 2 offers received at least. Both contract outcomes and bureau characteristics are replaced by their standard scores. Standard errors are clustered by bureau and procurement 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.

Table A.14

(a) Project Complexity and Bureau Specialty

	Cost Performance			Time Performance		
	(1)	(2)	(3)	(4)	(5)	(6)
Competence (Q28)	0.02	-0.38**	1.27*	0.05	-0.21	1.81**
	(0.08)	(0.16)	(0.72)	(0.09)	(0.13)	(0.89)
Observations	44673	20422	56641	44673	20422	56641
Weak Id. F-Test	25.35	13.7	3.31	25.35	13.7	3.31
Underid. F-Test	29.36	16.43	4.44	29.36	16.43	4.44
Overid. F-Test	3.21	.3	4.18	.21	.15	1.43
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

(b) Bureau Managing Many Contracts

	Cost Performance		Time Performance	
	(1)	(2)	(3)	(4)
Competence (Q28)	0.33***	0.21**	0.36***	0.27***
	(0.10)	(0.11)	(0.09)	(0.09)
Observations	114431	93270	114431	93270
Weak Id. F-Test	31.05	25.25	31.05	25.25
Underid. F-Test	55	45.29	55	45.29
Overid. F-Test	2.43	5.10	.24	2.71
Amount FEs	Yes	Yes	Yes	Yes
Duration FEs	Yes	Yes	Yes	Yes
Agency FEs	Yes	Yes	Yes	Yes
Year FEs	Yes	Yes	Yes	Yes
State FEs	Yes	Yes	Yes	Yes

Notes: Columns 1 and 3 of Panel (a) report the results of our baseline empirical model (i.e. IV regression of contract outcomes on *Competence* - instrumented by *Proximal Death* and *Relevant Death*) on the subset of bureaus managing at least 100 contracts in our sample. Columns 2 and 4 restrict instead the focus on bureaus managing at least 500 contracts. In Panel (b) Within our most detailed definition of procurement categories (*productorserivecode*), we associate the expected cost of the contract to the respective quantile of the distribution. We call *complexity* the resulting categorical variable (1 to 4). Then, we observe at the bureau level the mode of *complexity*. i.e. what the most recurrent classification of contracts is in terms of their procurement-category-specific size. Consider that a bureau procures many different services and the mode of *complexity*, according to its construction, does not necessarily entail that the bureaus awards contracts of similar amounts. We then split our sample of bureaus in administrations that award relative more complexity 1 contracts (Columns 1-3), complexity 2 or 3 contracts (Column 2-5), complexity 4 contracts (Column 3-6). Both contract outcomes and bureau characteristics are replaced by their standard scores. Standard errors are clustered by bureau and procurement 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.

Table A.15: IV regressions - 6 months ahead (before) instrument + workforce control

	Cost Performance		Time Performance		Cost Performance		Time Performance	
	-6m	+6m	-6m	+6m	-6m	-6m	-6m	-6m
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Competence (Q28)	0.57**	-1.02	0.15	-1.07	0.38**	0.44*	0.16	0.21
	(0.28)	(1.63)	(0.25)	(1.55)	(0.16)	(0.23)	(0.17)	(0.26)
Bureau Experience	-0.05	0.02	-0.06	-0.01	-0.04	-0.03	-0.06	-0.04
	(0.04)	(0.09)	(0.05)	(0.09)	(0.03)	(0.03)	(0.05)	(0.05)
Bureau Size	0.00	0.14	0.03	0.13	0.02	0.00	0.03	0.03
	(0.03)	(0.14)	(0.03)	(0.14)	(0.02)	(0.03)	(0.03)	(0.04)
Total Employment					-0.03	-0.01	0.00	0.00
					(0.02)	(0.02)	(0.02)	(0.02)
Observations	92902	92992	92902	92992	92902	92902	92902	92902
Weak Id. F-Test	8.6	.75	8.63	.75	22.28	19.69	22.28	19.69
Underid. F-Test	9.69	.74	9.69	.74	30.48	28.62	30.48	28.62
Overid. F-Test	0	0	0	0	0	0	0	0
Amount FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Duration FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Agency FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
State FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Firm FEs	No	No	No	No	No	Yes	No	Yes

Notes: Columns 1-4 report the results of our baseline model (i.e. IV regression of contract outcomes on *Competence*) instrumented by the 6 months before and ahead single-death-event instruments. Columns 5 and 6 show the same regression model instrumented by the 6 months before instrument including also a control for the worforce. In Columns 7 and 8, we also add firm fixed effets. Both contract outcomes and bureau characteristics are replaced by their standard scores. Standard errors are clustered by bureau and procurement 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.

Table A.16: IV regressions - Long Run Model

	Cost Performance		Time Performance	
	(1)	(2)	(3)	(4)
ΔCompetence	0.41***	0.41**	0.41***	0.41*
	(0.14)	(0.16)	(0.14)	(0.22)
L3 Cost Performance	0.17***			
	(0.02)			
L3 Time Performance			0.10***	
			(0.02)	
L3 Residualized CP		0.17***		
		(0.02)		
L3 Residualized TP				0.10***
				(0.02)
Observations	4686	4686	4686	4686
Weak Id. F-Test	9.9	9.93	9.94	9.95
Underid. F-Test	36.43	36.56	36.54	36.55
Overid. F-Test	2.03	2.67	7.2	7.96

Notes: Columns 1 and 3 report the results of our long run empirical model (i.e. IV regression of contract outcomes on Lagged outcomes and change in *Competence* - instrumented by the 3-years lag and 2-years lag of *Proximal Death* and 3-years lag and the contemporaneous value of *Relevant Death*). Columns 2 and 4 show the same regression model with the residualized outcome variables. Both contract outcomes and bureau characteristics are replaced by their standard scores. Standard errors are clustered by bureau and procurement category and are in parentheses and in columns 2 and 4 they are block-bootstrapped . 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.

Table A.17: Management Practices and Alternative Measure of Cooperation

	Cost Performance		Time Performance	
	(1)	(2)	(3)	(4)
Cooperation Q26	0.07***		0.07***	
	(0.01)		(0.01)	
PCA Management		0.08***		0.07***
		(0.02)		(0.02)
Observations	122526	122526	122526	122526
Adj. R_Squared	0.14	0.14	0.12	0.12
Amount FEs	Yes	Yes	Yes	Yes
Duration FEs	Yes	Yes	Yes	Yes
Agency FEs	Yes	Yes	Yes	Yes
Year FEs	Yes	Yes	Yes	Yes
State FEs	Yes	Yes	Yes	Yes

Notes: We convert the answers to the below questions/statements in FEVS, which we classify as "Management-practices-related", in one-factor principal component analysis (i.e. PCA Management): Q46 ("My supervisor provides me with constructive suggestions to improve my job performance"), Q47 ("Supervisors in my work unit support employee development"); Q48 ("My supervisor listens to what I have to say"); Q49 ("My supervisor treats me with respect"), Q50 ("In the last six months, my supervisor has talked with me about my performance"); Q51 ("I have trust and confidence in my supervisor"), Q52 ("Overall, how good a job do you feel is being done by your immediate supervisor?"), Q53 ("In my organization, senior leaders generate high levels of motivation and commitment in the workforce"), Q54 ("My organization's senior leaders maintain high standards of honesty and integrity"), Q55 ("Supervisors work well with employees of different backgrounds"), Q56 ("Managers communicate the goals and priorities of the organization"), Q57 ("Managers review and evaluate the organization's progress toward meeting its goals and objectives"), Q58 ("Managers promote communication among different work units, for example, about projects, goals, needed resources"), Q59 ("Managers support collaboration across work units to accomplish work objectives"), Q60 ("Overall, how good a job do you feel is being done by the manager directly above your immediate supervisor?"), Q61 ("I have a high level of respect for my organization's senior leaders"). Then, we perform an OLS regression of contract performance metrics on PCA Management in Column 2 and 4. Instead, Column 1 and 3 report OLS regression of contract outcomes on Q26 (" Employees in my work unit share job knowledge with each other") as answer to FEVS cooperation-related question alternative to Q20 ("The people I work with cooperate to get the job done"). Both contract outcomes and bureau characteristics are replaced by their standard scores. Standard errors are clustered by bureau and procurement 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.

Table A.18: IV regressions - Baseline with Workforce Controls

	Cost Performance	Time Performance	Cost Performance	Time Performance	Cost Performance	Time Performance
	(1)	(2)	(3)	(4)	(5)	(6)
Competence (Q28)	0.37***	0.36***	0.32***	0.34***	0.38**	0.39**
	(0.10)	(0.09)	(0.08)	(0.08)	(0.16)	(0.16)
Bureau Experience	-0.04	-0.06	-0.03	-0.06	-0.01	-0.04
	(0.03)	(0.04)	(0.03)	(0.04)	(0.03)	(0.05)
Bureau Size	0.01	0.01	0.01	0.01	0.03	0.03
	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.03)
Total Employment			-0.03**	-0.02		
			(0.01)	(0.01)		
Total Employment: variation					-0.01	-0.00
					(0.01)	(0.01)
Observations	122526	122526	121649	121649	104401	104401
Weak Id. F-Test	30.79	30.79	38.75	38.75	17.68	17.68
Underid. F-Test	55.61	55.61	72.68	72.68	34.34	34.34
Overid. F-Test	1.44	.14	1.12	.08	2.46	.05
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
firms	No	No	No	No	Yes	Yes

Notes: We replicate our baseline empirical analysis (i.e. IV regression of contract outcomes on *Competence* - instrumented by *Proximal Death* and *Relevant Death*) including controls for the workforce variation. Columns 1 and 2 report our baseline results, i.e. coefficients from columns 3 and 6 of Table 8. Columns 3 and 4 include the adjusted number of employees (relevant and not relevant). Columns 3 and 4 include the adjusted yearly change in the number of employees (relevant and not relevant). 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.