NBER WORKING PAPER SERIES

BUREAUCRATIC COMPETENCE AND PROCUREMENT OUTCOMES

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

Working Paper 24201
http://www.nber.org/papers/w24201

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
January 2018

For useful comments, we are grateful to Oriana Bandiera, Klenio Barbosa, Erich Battistin, Michael Best, Sascha Becker, Stephane Bonhomme, Ga´etan de Rassenfosse, Mansi Deshpande, Fred Finan, Luca Gnan, Josh Gottlieb, Camilo Garcia Jimeno, Alex MacKay, Marco Manacorda, Claudio Michelacci, Magne Mongstad, Andras Niedermayer, Benjamin Olken, Gustavo Piga, Andrea Prat, Emilio Raiteri, Imran Rasul, Stephane Saussier, Andrei Shleifer, and seminar participants at the Bank of Italy, Paris School of Economics, University of Lancaster and University of Rome Tor Vergata. We acknowledge for financial support the University of Rome Tor Vergata (Bando Doppia Cattedra, 2014-2016). Elisabetta Iossa also gratefully acknowledges financial support from Bocconi University, Laboratorio Infrastrutture-Autostrade per l'Italia. Giancarlo Spagnolo thanks the Swedish Research Council (Vetenskaprådet) for financial support. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

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

ABSTRACT

To what extent does a more competent public workforce contribute to better economic outcomes? We analyze this question in the context of the US federal procurement by combining data on office-level competencies, federal workforce characteristics, and procurement performance. Using an instrumental variable strategy, we find that the effects of competence heterogeneity across bureaus are quantitatively important: if all federal bureaus were to obtain NASA's high level of competence (corresponding to the top 10 percent of competence), delays in contract execution would decline by 7.2 million days and price renegotiations would drop by $13.5 billion over the 2010-2015 period analyzed. Cooperation within the office appears to be a key driver of the findings.

Francesco Decarolis
Department of Economics
Boston University
270 Bay State Road
Boston, MA 02215
and EIEF
and also NBER
francesco.decarolis@eief.it

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

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

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

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

I Introduction

An inefficient bureaucracy can represent a major obstacle to economic activities. In a path-breaking study, De Soto [1990] documented how excessive government requirements for a business to begin operating can dramatically slow down the entry of new enterprises. Djankov et al. [2002] notoriously expanded this work by measuring, for 85 countries, the number of procedures, the official time and the official cost that a start-up must bear before it can operate legally. These works laid the ground for the World Bank's Doing Business project which, as of 2017, provides objective measures of business regulations and their enforcement across 190 economies and it is widely recognized as a fundamental competitiveness' indicator.[1]

In 2015 the World Bank also began to release its *Benchmarking Public Procurement*. In 2017, this report presented data from 180 countries on the legal and regulatory environments affecting the ability of private companies to do business with governments. The project reveals the existence of great heterogeneity in the quality of the procurement process. Empirical research, reviewed in the next section, further confirms this heterogeneity. With about 15 percent of world GDP spent every year on public procurement, it is of fundamental importance to understand which are the main drivers of procurement outcomes.[2]

Measuring the inner functioning of public and private organizations has recently proved of major importance in explaining heterogeneity in their performance (see Chong et al. [2014], Bloom and Van Reenen [2007], Bloom et al. [2014], Bloom et al. [2015] and the vast strand of literature that followed), but to most of us the functioning of procurement offices remains unknown. Saussier and Tirole [2015] lamented the lack of competencies among French procurers reporting data from the French Public Procurement Grouping Union, according to which 63 percent of French public buyers do not have a legal profile and 61 percent joined a purchasing department following a period of internal mobility, with no prior experience in

[1] The extent and the channels through which bureaucracies affect the economy are areas of active research. See, Finan, Olken and Pande [2015], Bertrand et al. [2016] and the related studies reviewed in section 2.

[2] In high-income countries, public procurement averages 12 percent of GDP and about 29 percent of total general government expenditure. In 2013 alone, federal procurement spending in the US topped $460 billion, according to the Office of Management and Budget, a number that roughly equates to the GDP of Belgium. In developing countries, the fraction of public procurement to GDP averages 18 percent and public procurement expenditure may reach up to 70 percent of all government expenditure (World Bank [2017]).

the field. Only 39 percent of public buyers undertook specific training, and less than a third felt familiar with the economic and industrial fabric. Best, Hjort and Szakonyi [2017] report that the primary requirements for individuals seeking a job as public procurers in Russia are simply a legal education and knowledge of the existing procurement laws. This study also reveals that 60 percent of within-product price variation in Russia in 2011-2015 was due to the bureaucrats and organizations in charge of procurement. For the same category of goods, Bandiera, Prat and Valletti [2009] estimate that 21 percent of the expenditures by Italian public buyers would be saved if all public bodies paid the same as buyers at the 10th percentile of the estimated procurement price distribution, an amount reaching 1.6-2.1 percent of Italian GDP. They also identify the main sources of this waste: bureaucratic inefficiency accounted for over 83 percent of total estimated waste, and corruption for only 17 percent. The social costs of inefficient procurement thus largely exceed those of corruption even in a country where the latter is perceived as a major problem.

In this paper, we adopt an organizational perspective and focus on the competence of procurers, its drivers and its impacts on procurement outcomes. We analyze US federal contracts execution during the 2010-2015 period. We rely on a major survey, the Federal Employee Viewpoint Survey (FEVS), that for more than ten years has been administered with the same questions to nearly all government agencies, drawing responses from about one fourth of all federal employees every year.[3] From this dataset, we build measures of federal bureaus and agencies competence, and investigate the features of the offices that underlie this competence, focusing on three different competence components: cooperation among employees, incentives and skills.[4]

We evaluate the effects of these organizational characteristics on the performance of federal contracts. Using the Federal Procurement Data System (FPDS), a system tracking nearly every federal contract, as well as every follow-on action, we construct two proxies of

[3]As discussed below, this survey was developed with the objective of monitoring the strategic management of the federal agencies' human capital and became a pillar of the public sector managerialization effort pursued by all US administrations since 2002. Although little known in the economics literature, it has been repeatedly analyzed by political scientists and public administration scholars, see Fernandez et al. [2015].

[4]The term competence is typically used referring to a person's underlying characteristics that are causally related to job performance. In the context of the procurement debate, however, we are talking about "organizational competence," i.e. the competence of a procurement unit within a larger office.

performance based on time delays and cost overruns that represent our main outcome variables. While we would ideally like to find out the extent to which the bureau-level competencies, as measured by the FEVS data, determine the procurement outcomes, the association between more complex contracts and more competent bureaus implies a downward bias in any straightforward regression of performance on competence. Thus, from a third dataset (FedScope) containing characteristics of the public workforce, we construct instruments for bureaus' competencies based on death occurrences of specific types of employees.

Before trying to assess the effect of bureaus' competencies on procurement outcomes, we present detailed descriptive evidence to establish three facts. First, contractual performance is a persistent characteristic of a bureau. Second, the relevant variation in performance occurs at the bureau level and not only at the agency level. Third, the simple association between our measures from the FEVS and the procurement performance proxies is likely to underestimate the benefits of greater competence on procurement.

Our IV estimates indicate a strong, positive effect of competence on both cost and time performance. The magnitude of these effects is substantially higher than the corresponding OLS estimates. Under our preferred IV estimates, the effect of lifting the level of *competence* of all bureaus to that of the bureau at the 90th percentile of the *competence* distribution - which happens to be NASA Glenn Research Center[5] - would imply a reduction in cost overrun of $102,619 on average per contract or around $13.5 billions in total across all contracts in the dataset. Moreover, this implies a saving of 54.4 days in the effective execution time, corresponding to 7.2 million days across all the contracts in the dataset. We assess the robustness of these estimates to a broad set of robustness checks involving the econometric model used, sample selection and measurement of both competencies and performance.

We find that, out of the competence components, cooperation among employees plays the most important role. The IV estimates for cooperation indicate that improvements in this dimension have effects comparable to those of improving the overall competence. Although the validity of our proposed instruments is stronger when we look at the aggregate

[5]This is one of the 98 bureaus from 23 different agencies that we observe conducting procurement of federal contracts in the 2010-2015 period analyzed. NASA Glenn Research Center, located in Ohio, designs and develops innovative technology to advance NASA's missions in aeronautics and space exploration.

competence relative to the case of the individual competence components, the fact that cooperation appears so relevant is interesting. First, heterogeneity in reported cooperation is a more important driver of differences in competence than skills and incentives, regardless of the connection of all these variables with the procurement outcomes. Second, the prominence of cooperation conforms with the view that successful procurement requires procurers to appropriately handle and coordinate a multiplicity of tasks involving different individuals and offices. These tasks include: choosing the tender format, the award criteria, the tender specifications, and handling all ex-post contract modification requests. The complexity of the environment implies that no one size can fit all: the tender and contract design must take into account the type and complexity of the good, work or service being acquired, the existing competition in the market and the characteristics of the pool of potential suppliers, besides the legal principles and available contract management ability and resources. A multidisciplinary approach requiring collaboration among employees with different skills is thus essential to choose the appropriate tender and contract design. Hence, incentives and skills in procurement may matter only to the extent that they promote cooperation.

We explore the extent to which the key role played by cooperation that our analysis has uncovered can be ascribed to the presence of effective managers, able to lead a group to effective cooperation. While we lack direct measures to draw definitive answers, in the spirit of the recent work by Jäger [2017], we explore the heterogenous effects obtained through instruments considering the deaths of different subgroups of employees. 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 the bureau competencies.

Our quantification of the impact of competence on procurement outcomes confirms the importance of improving decision making within procurement organizations. Therefore it adds to the small, but growing analysis of the demand side in procurement (Bandiera, Prat and Valletti [2009], Bajari, McMillan and Tadelis [2009], Decarolis [2014] and Best, Hjort and Szakonyi [2017]). 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 federal acquisition workforce and certify their competence.[6] 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 crucial role of cooperation suggest that, while certification of individual contracting officer's capabilities is certainly welcome and important, it may not be sufficient. Certification programs would 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 is often done for private firms.

Our results also have implications for the internal design of the public sector. Despite the renewed interest in this area (Dal Bo, Finan and Rossi [2013], Finan, Olken and Pande [2015] and Bertrand et al. [2016]), still much remains to be understood about the inner working of procurement offices. In this respect, our results reinforce the insights from the management literature that technical skills are necessary, but definitely not sufficient. Interpersonal skills are also needed to build and manage the procurement team, a result in line with our "work cooperation" effect. In particular, we consider our findings the first strong empirical indication that improving the effectiveness of a public purchasing unit requires not only selecting employees with the right technical skills, but also, and more importantly, ensuring that the unit is run by highly skilled procurement managers, able to adopt appropriate procurement management practices that ensure strong cooperation among different specialties within the unit. This result thus complements the recent findings surveyed in Bloom et al. [2015] on the importance of management practices in the private sector and in other branches of the public sector, as we discuss below. Furthermore, the central role of cooperation that we

[6]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.

uncover may serve as a useful check on the "New Public Management" agenda developed in the 1980s as part of an effort to make the public sector more "businesslike" and to improve its efficiency via performance incentives. It suggests the importance of choosing managers favoring collaborative environments, and crafting incentives that are team-based rather than individual, so as to reward collaboration and strengthen group identity.

II Related literature

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

Regarding procurement, our paper contributes to the recent and growing literature on the determinants of public procurement outcomes. 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 auctions with time incentives (Lewis and Bajari [2011], Lewis and Bajari [2014]), minimum prices (Chassang and Ortner [2017]), contract duration (MacKay [2017]), electronic procurement (Lewis-Faupel et al. [2016]), transparency (Coviello and Mariniello [2014]), discretion (Coviello, Guglielmo and Spagnolo [2017]), contract renewal (Chong, Saussier and Silverman [2015]), and past performance (Banerjee and Duflo [2000]).

An aspect which has not received due attention, however, is the role of buyers. Aside from the obvious concerns about corruption risks, the theoretical literature has offered a variety of explanations for why more competent, higher quality procurers should improve procurement outcomes which hinge on the buyers' involvement with the various stages characterizing contract procurement: the ex ante design of an adequate award procedure and contract, the selection of participants and winner(s) at the award stage and, finally, the ex post contract management (see Spulber [1990], Manelli and Vincent [1995], Bajari and Tadelis [2001]). Nevertheless, a systematic analysis of the hypothesis that buyers' characteristic matter has not yet been undertaken. Among the few studies in this area, the closest papers to ours that we are aware of are Bandiera, Prat and Valletti [2009], described in the introduction, Bajari, McMillan and Tadelis [2009], Decarolis [2014] and Best, Hjort and Szakonyi [2017]. Bajari, McMillan and Tadelis [2009] analyzes auctions versus negotiations. Employing a dataset of private sector building contracts awarded in Northern California during the years 1995-2000, they find that project characteristics affect the choice of the award mechanism and that auctions are used more often by more experienced buyers (i.e., those in organizations that are larger and procure contracts more frequently). Decarolis [2014] studies procurement outcomes in terms of ex post contract renegotiations and shows that they depend on the choice of the procurement mechanism and on the level of bid screening undertaken by the buyer. Large buyers, who are the most experienced, are better able to screen offers, as shown by the better outcomes in terms of time and cost renegotiations for given contract choices. Our paper complements these studies by analyzing procurer quality features that go beyond the mere frequency of tendering and organizational size; we measure their impact and investigate the specific channels through which these features affect procurement outcomes.

Closest to us is probably the study of Best, Hjort and Szakonyi [2017] mentioned above. Like them, we are interested in the extent to which public procurement is affected by the effectiveness of the bureaucracy. We both find that improving the effectiveness of the public workforce would have sizable effects on procurement outcomes. Nevertheless, there are several major differences between our studies. First, the main quantitative contribution of Best, Hjort and Szakonyi [2017] is to estimate the causal effect of specific individuals and organizations on prices paid. This is made possible by the specific legal arrangement

of the Russian procurement which involves the same individuals purchasing on behalf of different organizations. Instead, our main contributions are to analyze which characteristics of the purchasing organizations drive their greater procurement effectiveness and to offer a clear indication of the relative merits played by cooperation, skills and incentives in the office. Second, our setting is substantially different not just because Russia and the US have obvious economic and institutional differences, but also because the contract procurement on which we focus entails major differences to the procurement of standardized goods analyzed in Best, Hjort and Szakonyi [2017] (and Bandiera, Prat and Valletti [2009]). This is reflected in the different economic outcomes that we study (contract renegotiations instead of purchase price) and, hence, in the underlying mechanisms and policy prescriptions. Indeed, while their focus is on what happens at the stage of contract design and award, our focus is on the follow-up phase of contract management. Clearly, all phases are interlinked and, as discussed in the conclusions, some of their policy prescriptions match well with ours.

Finally, our focus on different types of bureau competence connects our work to the recent literature on managerial practices. In particular, Bloom et al. [2015] extend their survey-based methodology developed for manufacturing firms in Bloom and Van Reenen [2007] to investigate the role of managerial practices in the public sector. They collect data on management practices in over 1,800 high schools in eight countries, and show that school management quality is strongly associated with better educational outcomes. Analogously, Bloom et al. [2014] document large variations in hospitals performance across countries which they link to better management practice. Closer to our theme, Rasul and Rogger [2016] show that management practices affecting autonomy correlate robustly with public project completion in Nigeria, while practices related to incentives/monitoring of bureaucrats are negatively associated with completion rates. Consistent with the findings in this literature, we document a substantial variation in the quality of US procurement agencies, but in contrast to Rasul and Rogger [2016] we do not find a clear negative effect of incentives. Our estimates indicate either a small, positive effect (for time performance) or no effect (for cost performance). This suggests that incentives in the public sector play a different role in strong and weak institutional environments. Moreover, methodologically our approach differs in that we can exploit both the time variation present in our data and an instrumental variable

strategy for causal identification, while they adopt a variance decomposition approach based on observing the same procurers working for different organizations. Finally, our finding of a dominant effect of cooperation also squares well with the recent findings in Blader, Gartenberg and Prat [2016] on the benefits of "cooperative" managerial practices relative to high powered individual incentives.

III Institutional Background

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

The bureaus in charge of contract procurement typically have an organizational structure that we will exploit to devise our instrumental strategy. As source selection guides reflect, the planning and competition process often calls for input and involvement by both central and regional bureau officials, the latter being those located in the area of contract execution.[7] Whilst there are exceptions, the general practice in U.S. federal procurement is to have the tendering and award processes overseen by a "procurement contracting officer" (PCO) in a regional office. The local PCO can receive extensive input from attorneys, program personnel and other experts in the central headquarters, who can even overrule the regional officer's decision. Thus, the training and judgment of both the regional officer and the headquarters personnel may have a material impact on the quality of the contract award. Instead, the actual management of the contract is typically done only by a regionally located contracting officer. The PCO decides on contract renegotiations by handling the claims for additional

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

costs or time from contractors during the course of contract execution.[8] We will return to the distinction between central and regional bureaus when discussing our instruments.

IV Data

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

A. Federal Bureaus' Competencies: FEVS Data

The principal explanatory variables that we use to measure bureaus' competencies come from the Federal Employee Viewpoint Survey (FEVS). Since the early 2000s, the Office of Personnel Management has called on federal employees to provide their opinions on all aspects of their employment, including evaluations of their supervisors, bureaus, agencies and, more generally, of their work experience. The goal is to measure government employees' perceptions of whether, and to what extent, conditions characterizing successful organizations are present in their bureaus and agencies and, ultimately, to influence change in their workplace. The beginning of this survey dates back to 2002 when it was first administered under the name "Federal Human Capital Survey" as an essential tool of the George W. Bush administration's agenda for a managerialization of the public administration. Since then, the survey has been mainly used for internal human resources management recommendations from the Office of Personnel Management to the agencies. This office uses the FEVS to monitor human capital management initiatives and outcomes and to provide guidance, resources, and technical assistance to the entire federal government.[9]

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

[9] Despite the proliferation of published works based on FEVS data (see Fernandez et al. [2015]), our 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.[10] There is a total of 98 bureaus from 23 agencies. The agencies that are invited to participate account for 97 percent of the executive branch workforce. 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.

Table 1: 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 % contributions that each variable assumes through the weights calculated by the factor analysis.

Table 1 reports the entire set of nine questions composing the "My work unit" section, running from question 20 to 29 of the survey. The first eight questions span different features of the workplace that we classify into three broad categories: cooperation, incentives and skills, respectively. The ninth question (Q28) is a summary measure that we consider as a measure of overall bureau's competence. Based on the wording of the questions, we select the answers to Q20, Q21 and Q23 to form the measures of *cooperation*, *skills*, and *incentives*, respectively, used in the analysis. To limit the arbitrariness of this choice, we also use the whole set of eight questions to perform a principal components analysis (PCA). The last two columns of Table 1 report the weights that the PCA analysis associates to each question. Only two factors are specified as the analysis reveals that two factors are sufficient to explain 84 percent of the common variance among cooperation, skills and incentives. 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.[11] The second factor,

[10] We consider this time frame as before 2010, the FEVS was run every other year, before becoming an annual survey starting in 2010.

[11] In section 6, we discuss the construction of these weights as well as the robustness to other alternatives.

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). Below we refer to the first component as *PCA Skill/Incentive* and to the second as *PCA Cooperation*.

The summary measure Q28 is particularly interesting as this is the only question in the survey that can proxy for a self-evaluation of the overall work conducted by individual work units within each agency. Therefore, we use this variable as our main measure of overall bureau competence. To distinguish bureau features from agency features, we will also use Q29 from the section "My agency" which asks wheter "The workforce has the job-relevant knowledge and skills necessary to accomplish organizational goals". We label this variable *Ag.competence*.

Table 2: Summary statistics

	Mean	Median	S.D.	N
Bureau Characteristics (FEVS Data)				
Competence (Q28)	0.81	0.81	0.03	445
Ag.Competence (Q29)	0.69	0.69	0.04	445
Cooperation (Q20)	0.73	0.73	0.04	445
Skill (Q21)	0.54	0.54	0.05	445
Incentive (Q23)	0.45	0.45	0.05	445
Contract Characteristics (FPDS Data)				
Contract Amount	551	91	3533	131686
Expected Duration	251	226	211	131686
Time Performance	0.72	1	0.33	131686
Cost Performance	0.83	1	0.27	131686
Total Cost	1108	118	10337	131686
Total Time	503	364	720	131686
Negotiation	0.26	0	0.44	131686
No. of Offers	3.92	2	6.47	131686
Cost Plus	0.07	0	0.25	131686
Constructions	0.16	0	0.37	131686
$Bu.Performance^C$	0.83	0.86	0.13	126273
$Bu.Performance^T$	0.72	0.73	0.14	126273
Bureau Experience	512	190	725	131686

Notes: The top panel presents summary statistics for the FEVS data. The unit of observation is a bureau-year. Not all 98 bureaus are observed for all years due to organizational changes within the 23 agencies covered. The bottom panel presents summary statistics for the FPDS data. *Bureau Experience* is scaled down by thousands of US dollars.; *Contract Amount* and *Total Cost* are expressed in thousands of US dollars; *Expected Duration* and *Total Time* are expressed in days; *Cost Performance*, *Time Performance*, *Competence*, $Bu.Performance^C$, and $Bu.Performance^T$ are bounded between 0 and 1; all variables are described in the main text.

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.[12] We first transform these answers into numerical values from one to five, then we aggregate answers at the bureau level (by using the FEVS' representative weights), and finally we normalize the resulting variables to be between zero and one. The top panel of Table 2 reports summary statistics for the FEVS variables: competence, agency-competence, cooperation, incentives and skills. The correlation between the bureau and agency competence is high at 0.6, but far away from 1. Further discussion on their relationship is presented in the next section.

Table 3: Channels of Competence: Competence (Q28) on Bureau Characteristics

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Cooperation (Q20)	1.06***				0.93***		0.96***	0.97***	
	(0.02)				(0.03)		(0.05)	(0.05)	
Cooperation (Q26)		0.83***				0.58***	-0.04	-0.06	
		(0.03)				(0.04)	(0.04)	(0.04)	
Skill (Q21)			0.58***		0.06**	0.09**	0.06**	0.12***	
			(0.03)		(0.03)	(0.04)	(0.03)	(0.03)	
Incentive (Q23)				0.72***	0.10***	0.24***	0.10***	0.04	
				(0.03)	(0.04)	(0.05)	(0.04)	(0.03)	
PCA - Cooperation									0.90***
									(0.03)
PCA - Skill/Incentive									0.48***
									(0.02)
Observations	445	445	445	445	445	445	445	445	445
R-squared	.83	.66	.40	.55	.84	.72	.84	.91	.91
Agency & Year FEs	No	No	No	No	No	No	No	Yes	Yes

Notes: Bureau characteristics are aggregated at bureau and year level and replaced by their standard scores. Last two columns include fixed effects for agency and year. * Significant at the 10 percent level; ** Significant at the 5 percent level; *** Significant at the 1 percent level.

Regarding the competence components, it is noticeable that the average response is much lower for skills and, even more, for incentives, than it is for cooperation or competence. This immediately suggests that while we will refer to cooperation, incentives and skills as the components of competence, these three variables do not account for all the variability

[12]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).

in competence across bureaus and years. In Table 3 we explore conditional correlations by reporting OLS regressions of competence on its components. The two clear messages that emerge by comparing the coefficients when each component is entered individually (columns 1-4) to when they are all entered jointly (columns 6-8) are that: *i)* cooperation (Q20) is the main driver of competence (with an estimated coefficient near one, highly statistically significant and an R-squared absent any other control of 83%), and that *ii)* cooperation as measured by Q20 is more strongly associated with competence than the alternative cooperation measure (Q26). This latter result is also confirmed by the last two columns where we report the estimates based on the PCA. This prominent role of cooperation will also drive our emphasis on this component in the later IV analysis.

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.[13] The data track every transaction between federal contracting bureaus and sellers. The system contains detailed information on contract actions over $3,000 (fiscal year 2004 and later data).[14] 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 constructions where, compared to goods' procurement, the extent of ex post cost uncertainty makes post award amendments, with the high haggling cost they imply, a useful proxy of contract performance (Williamson [1975], Tadelis [2002]).[15] Since not all modifications are equally problematic, we split the set of

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

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

[15]The web appendix discusses these sample selection choices. In the literature, post-award modifications are a 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

amendments to two broad categories: *in-scope* and *out-of-scope* revisions.[16] In line with other studies that use FPDS data, we consider *in-scope* amendments only.[17] The quantitative relevance of modifications is evident from the summary statistics reported in the bottom panel of Table 2. Our sample ranges from 2010 to 2015 and consists of 131,686 projects, associated to 953 categories (i.e., the typology of work or service procured). Although the overall value of the contracts is $72.6 billion using the initial awarding price, it increases to $146 billion if cost overruns are included). The distribution of contract amounts is highly skewed: fifty percent of contracts are for amounts below $91,000, while 10 percent of contract spending is accounted for by contracts worth more than $824,000. The average award per contract is $551,000, while the total cost, inclusive of any cost overrun, is $1.1 million. Correspondingly, the average contractual duration is 250 days, while the final contract duration inclusive of any delay is 502 days. In both cases, the medians are lower than the average.

To operationalize the data on time and cost renegotiations into a proxy for contract performance we proceed as follows. We define: *Time Overrun* as the difference - in days - between the actual completion date and the estimated date, and *Cost Overrun* as the sum - in thousands of dollars - of all renegotiated amounts. Finally, 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. Each performance measure ranges between zero, worst performance, and one, perfect performance (i.e., no overrun). In the data,

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

[16] 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*.

[17] 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.4 in appendix).

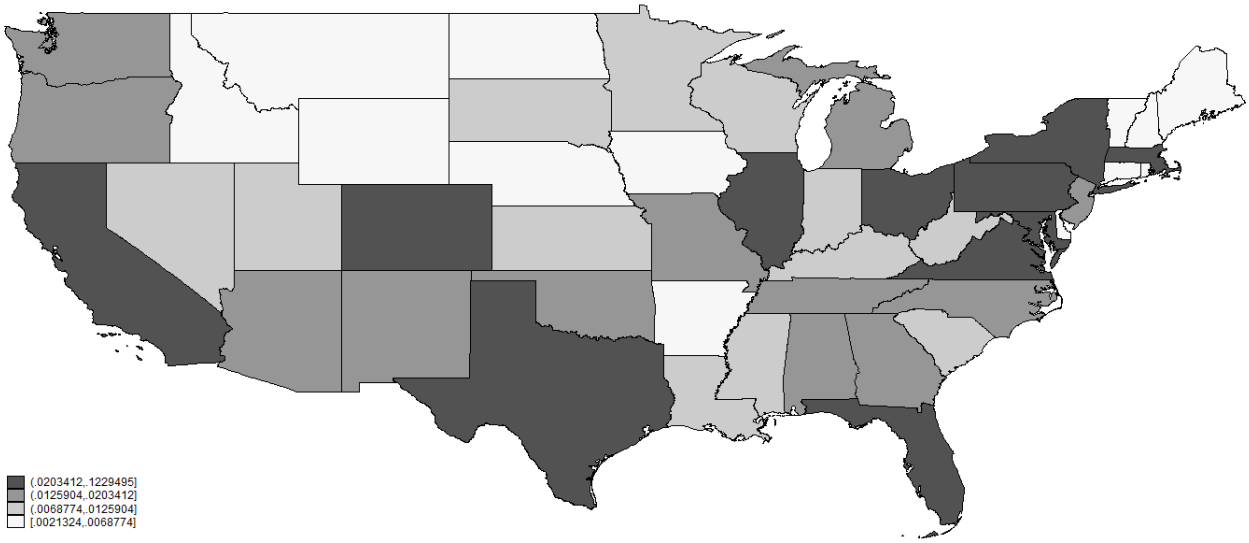
about half of the observations show no cost or time overruns. The coefficient of the linear correlation between the two equals 0.52 with a Spearman ρ of 0.57.

In addition to the two performance measures, the other relevant variables from the FPDS that will play a relevant role as controls in our analysis and that are reported in Table 2, are: *Negotiation*, a dummy variable indicating whether the contract uses negotiated procedures (i.e., the contract is awarded on the basis of a direct agreement with a contractor, after solicitation of a number of sources and without going through the competitive bidding process);[18] *Cost Plus*, a dummy variable equal to one if the contract is cost plus (i.e., the supplier is entitled to obtain compensation in proportion to its costs plus a mark-up) and zero if it is fixed price (i.e., the supplier is paid a fixed price, regardless of the cost incurred); and *Bureau Experience* and *Bureau Size*, which are the number of times a bureau has appeared in the past in the data for the same contract category and the cumulative value of contracts a bureau has awarded in the same year, respectively.

Table 2 shows that cost plus is only employed for a small fraction of contracts, 7 percent, leaving 93 percent of contract spending on fixed price format. Negotiated procedures are used in 26% of the contracts. Furthermore, 36% of total spending is by the Department of Defense, with the next largest purchasing agencies being the Department of Veteran Affairs and the Department of Homeland Securities. Finally, the data also exhibits useful geographical variation related to the place of contract execution that we document in Figure 1. Not surprisingly, more contracts take place in more densely populated states (12% of all contracts take place in California), but all states have at least some contracts. This is relevant because as discussed earlier, central and local procurement offices play different roles in terms of contract design and contract management, with the latter delegated to local

[18]Specifically, negotiation in procurement is broadly defined as a tendering method, used as an alternative to competitive processes such as auction, in which a request for proposals is sent only to qualified suppliers. The request for proposals details the scope, specifications, and terms and conditions of the proposed contract and the criteria for evaluating the bids. Then separate negotiations are carried out with each bidder whose bid falls within the preset competitive range. The process concludes with the award of contract to the bidder who offers most advantageous price, quality, and service combination. Negotiation as a method of contracting is intended to foster an impartial negotiation process between the government and potential contractors, leading to the selection of the proposal representing the best value to the government. The relevance of controlling for the difference between negotiated and more competitive procedures is stressed in several studies. See, for instance, Bajari, McMillan and Tadelis [2009].

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).

offices. This distinction will be used to devise the IV strategy presented below.

C. Public Workforce Characteristics: FedScope Data

The Office of Personnel Management (OPM) is an independent agency that functions as the central human resources department of the executive branch. In fulfilling its mission, OPM collects, maintains, and publishes data on a large portion of the federal civilian workforce. In FY 2010, OPM established a system called the Enterprise Human Resources Integration Statistical Data Mart (EHRI-SDM). This system provides access to personnel data for 96% of federal civilian executive branch employees.[19] These data are released through the Federal Human Resource (Fedscope) database, which represents the most comprehensive resource available on the size and scope of the federal workforce. Fedscope is the third data source that we use by merging it with FPDS at bureau level.[20] FedScope data are presented in five subject categories, called "cubes." For the purposes of the present paper, we only consider the "Employment" cube and the "Separations" cube for the years 2010-2015. The Employment cube contains several demographic characteristics along with information on

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

[20]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.

appointments and tasks: 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.

Table 4: 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 98 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 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 white-collars workers (i.e., managers and other white-collars workers) from the other employees, we will focus on this group of employees, whose separations from a bureau is most likely to have an impact on the bureau's competence. In Table 4, we report quantiles of age and salary of the managers and other white-collar employees: a total of 2.5 million employees, subdivided into 98 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 employeed at the local level. Finally, the geographical information in FedScope enables us to match the location (state) of each single federal employee with that of contract performance.[21] More details on the specific ways in which these data are used to construct our instrument are presented in section VI. Before that, however, in section V we present some relevant descriptive facts about the data that serve to establish the link between the FEVS and FPDS data.

V Survey Responses and Procurement Outcomes: Descriptive Evidence

Before trying to assess any causal effect of bureaus' competencies on procurement outcomes, it is useful to explore the data to establish three facts. First, contractual performance is a persistent characteristic of a bureau. Second, the relevant variation in performance occurs at the bureau and not only at the agency level. Third, the naive association between the competencies measures from the FEVS and the procurement 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 measure based on on the procurement data only. Thus, we aggregate time 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 in all contracts i that, at any given date t, the bureau had previously procured for the same contract category j:

$$Bu.Performance_{ijt}^{g} = \frac{\sum\limits_{ij\{\mathbf{1}|t'<t\}} w_{ijt'} * performance_{ijt'}^{g}}{\sum\limits_{ij\{\mathbf{1}|t'<t\}} w_{ijt'}}, \tag{1}$$

where w are Bartlett weights that are larger for more recent contracts.

We use these two bureau-level performance measures for time and cost to establish the

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

three features reported above. First, we seek to assess whether contractual performance is a transitory characteristic associated with each specific contract or whether it is persistent at bureau-level, as would be required to justify our focus in this study. Therefore, we begin by concentrating exclusively on the FPDS data and exploit the two bureau-level performance measures described above to estimate the following OLS model:

$$performance^{g}_{ijt} = \beta\, Bu.Performance^{g}_{jt} + \theta\, X_{ijt} + \iota_j + \kappa_t + \varepsilon_{ijt}, \tag{2}$$

where X contains additional covariates, and ι_j and κ_t indicate agency and year fixed effects, respectively. A positive β indicates persistency in bureaus' performance across contracts. Regarding the other covariates in X, we will gradually expand the set of controls to include: *Cost Plus*, *Negotiation*, *Bureau Experience* and *Bureau Size*. For the first two, since they tend to be used for more complex contracts, we expect a negative association with performance. Both more experienced and larger bureaus, instead, are expected to have higher performance.

Table 5 presents results for *cost performance* (columns 1-4) and *time performance* (columns 5-8).[22] Model (4) (or 8) is our preferred specification (baseline model) that includes all the controls for contract and bureau characteristics that are gradually added as we move from left to right along the table columns. In the baseline model, performance is regressed on bureau performance, along with fixed effects for agency, service category, year, and state as well as a dummy variable for each decile of the expected cost and duration distributions. Past performance has a positive and highly significant effect for both performance measures. The magnitude of the coefficients is 0.25 for Bu.Performance^C on Cost Performance and 0.21 for Bu.Performance^T on Time Performance, indicating a relatively high degree of persistency in performance. Despite the inclusion of different sets of controls (all entering with the expected signs), the magnitude and significance of the coefficient on past performance remains rather stable. Table 5 thus reveals persistency in performance: bureaus with a good performance in the past are more successful in having contracts completed within initial

[22]To facilitate the interpretation of the estimates, both the outcomes and endogenous regressors are replaced throughout all the regressions by their z-scores, i.e. the variables have been rescaled to have a mean of zero and a standard deviation of one. Hence, for example, a value of 1 indicates that the value for that case is one standard deviation above the mean.

contractual terms.

Table 5: Persistency of Cost and Time Performance

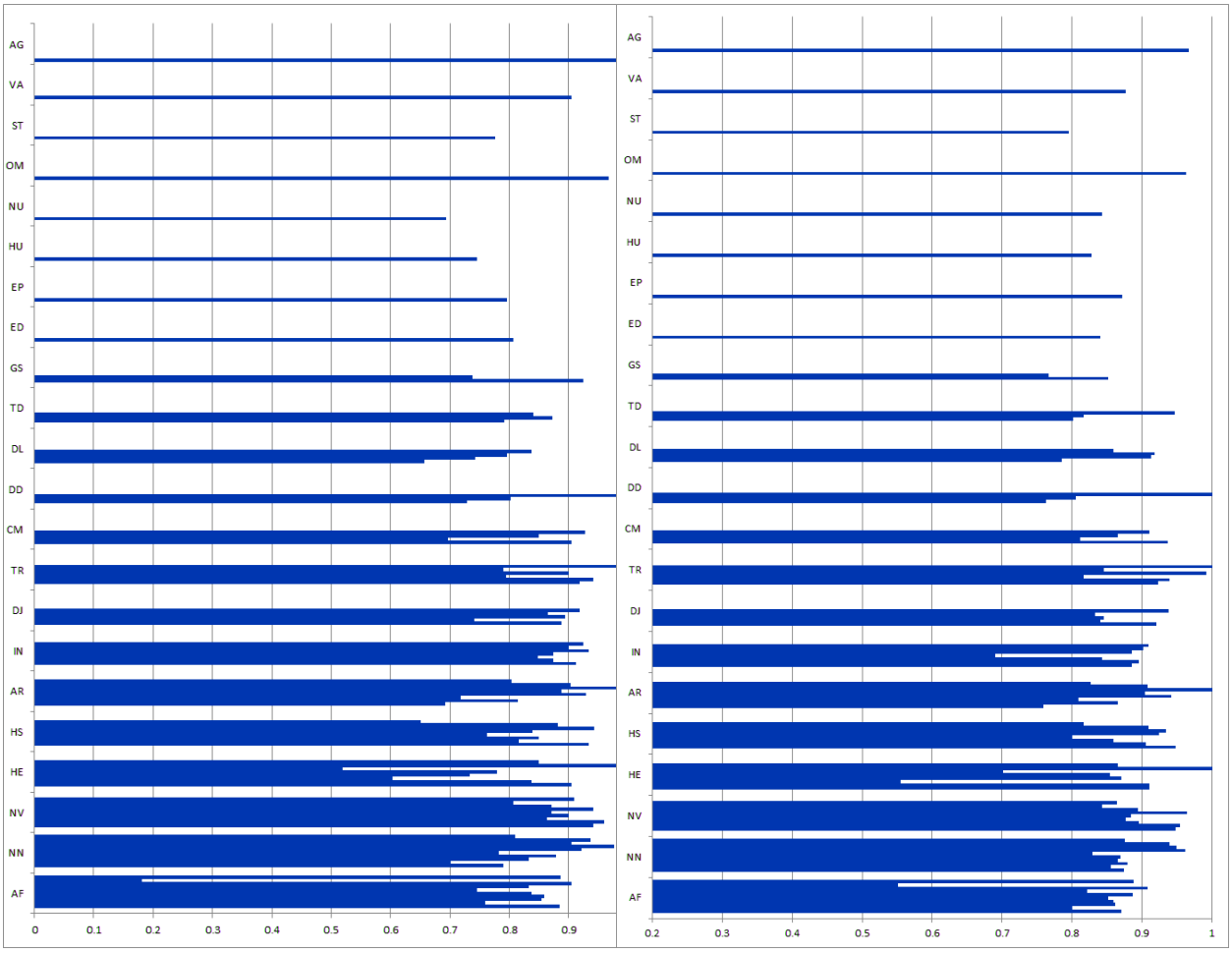
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Bu.Performance^C	0.26***	0.25***	0.25***	0.25***				
	(0.01)	(0.01)	(0.01)	(0.01)				
Bu.Performance^T					0.22***	0.22***	0.21***	0.21***
					(0.01)	(0.01)	(0.01)	(0.01)
Cost Plus		-0.49***	-0.49***	-0.48***		-0.18***	-0.18***	-0.18***
		(0.04)	(0.04)	(0.04)		(0.03)	(0.03)	(0.03)
Negotiation		-0.06***	-0.06***	-0.06***		-0.08***	-0.08***	-0.08***
		(0.02)	(0.02)	(0.02)		(0.02)	(0.02)	(0.02)
Bureau Experience			0.01	-0.02			-0.03	-0.03
			(0.02)	(0.02)			(0.03)	(0.03)
Bureau Size			0.02**	0.03***			0.02***	0.03***
			(0.01)	(0.01)			(0.01)	(0.01)
R-squared	.2	.21	.21	.21	.16	.16	.16	.16
Observations	121107	121107	121107	121107	121107	121107	121107	121107
Amount&Duration FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Agency FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year FEs	No	No	No	Yes	No	No	No	Yes
State FEs	No	No	No	Yes	No	No	No	Yes

Notes: OLS regressions. Both contract outcomes and bureau characteristics are replaced by their standard scores, as explained in footnote 24. Standard errors are clustered by agency and service category and are in parentheses. *Bureau Experience* and *Bureau Size* coefficients are scaled up by three and six orders of magnitude for readability purpose. *** Significant at the 1 percent level; ** Significant at the 5 percent level; * Significant at the 10 percent level.

The second data feature that we want to establish regards whether 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 miss a substantial share of the variation in performance. There we report the distribution of the bureau-level performance measures across all bureaus and agencies. For each agency, we report the performance of all bureaus with which the agency appears in the FPDS. The length of the horizontal lines measure the performance of $Bu.Performance^C$ (left) and $Bu.Performance^T$ (right). 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.

Figure 2: Procurement Performance across Bureaus and Agencies



(a) $Bu.Performance^C$ (b) $Bu.Performance^T$

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

Finally, to better understand the relationship between these two competence variables, as well as between them and contract performance, we present the case of 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 6 reports competence and performance measures of the top two agencies in terms of bureau competence - averaged across all the bureaus in the agency - which are the NRC (Nuclear Regulatory Commission) and NASA, both with an average *competence* equal to 0.86, and the worst

two, which are DVA (Department of Veteran Affairs) and DOJ (Department of Justice), with an average *competence* equal to 0.79 and 0.76, respectively. The corresponding values of *Ag.competence* across these four agencies in Table 6 also indicate a marked difference between the top and bottom two agencies. The last two columns of Table 6 report the values of the two performance measures for the four agencies considered.

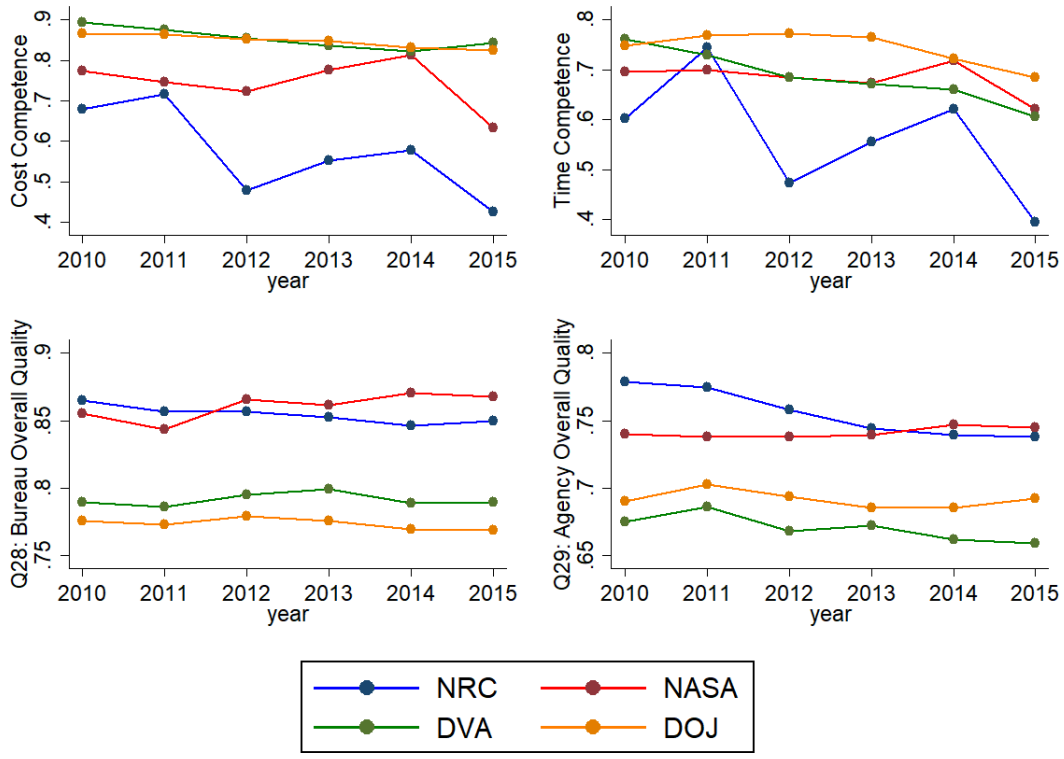
Table 6: Best and Worst Agencies (Competence)

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

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

The point that we want to stress through Table 6 is that, 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. Before turning to the explanation of how our empirical strategy deals with this issue, it is useful to conclude this section with some simple dynamics of the main variables. Figure 3 shows the evolution over time of the four variables for each of these four agencies. This figure shows two relevant features. First, the evidence based on the sample averages reported in Table 6 is persistent over time. Second, for each variable, while the time variation is not major, there are nevertheless changes over time which will be crucial to identify the effect of competence variations within the agencies.

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)).

VI Competence and Its Components: IV Analysis

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

$$performance^g_{ijt} = \beta\, competence_{jt} + \theta\, X_{ijt} + \iota_j + \kappa_t + \lambda_s + \varepsilon_{ijt}, \tag{3}$$

where $g = \{C, T\}$ indicates whether the outcome variable is cost or time performance, i, j, and t indicate contract, bureau, and time, respectively, X is a matrix of covariates, and ι_j and κ_t indicate agency and year fixed effects, respectively. Fixed effects for service categories and states will also always be included. The coefficient of interest is β, the effect of the bureau competence of contract performance, conditional on the other regressors. Analogous

regressions will be estimated substituting competence with its three components, as measured in the FEVS: cooperation, incentives and skills. For all these regressions, there are several difficulties in interpreting the OLS coefficient on the bureau's characteristic as a causal effect. Below, we illustrate this point referring to the estimate of performance on competence, but analogous concerns hold for the competence components.

First, our survey measure of competence is likely to be a noisy proxy for the set of characteristics that would ideally measure a bureau's competence. Individuals could misreport their bureau quality for a variety of reasons ranging from simple biased perceptions to sophisticated strategies to exploit how the OPM ensuing recommendation might benefit them. Furthermore, 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.

Our approach to addressing these potential concerns is twofold. First, we exploit the richness of our data to include in the model specifications all observable characteristics likely contributing to explaining performance. In particular, we always include agency and service category fixed effects to capture the differences in the types of procurement across both agencies and contracts. Then, we gradually include controls for *cost plus* and *negotiation*, as these types of contracts and awarding procedures tend to be associated with more complex projects. There are, however, multiple features of the project design and management that 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 employees' deaths. Notably, for each bureau and year, we compute the share of deaths involving those employees who are more likely to impact our *competence* indexes. We exploit the richness of the data to evaluate the public workforce under different aspects, resulting in the construction of two instruments that capture the distinct roles in that central and local

procurement officers can have on the procurement process.

First, inspired by the vast literature on CEO deaths, we focus on deaths of those employees more likely to have positive roles for the productivity of their office. We thus look 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 4, this implies looking at employees with a salary of $50,000 or more and an age of 50 years or less. Such thresholds value are able to capture 95% of the manager population and the upper half of the other white-collar employees. We thus build our first instrument as a dummy indicating whether a death of at least one employee in this age/salary groups occurred within a bureau-year:

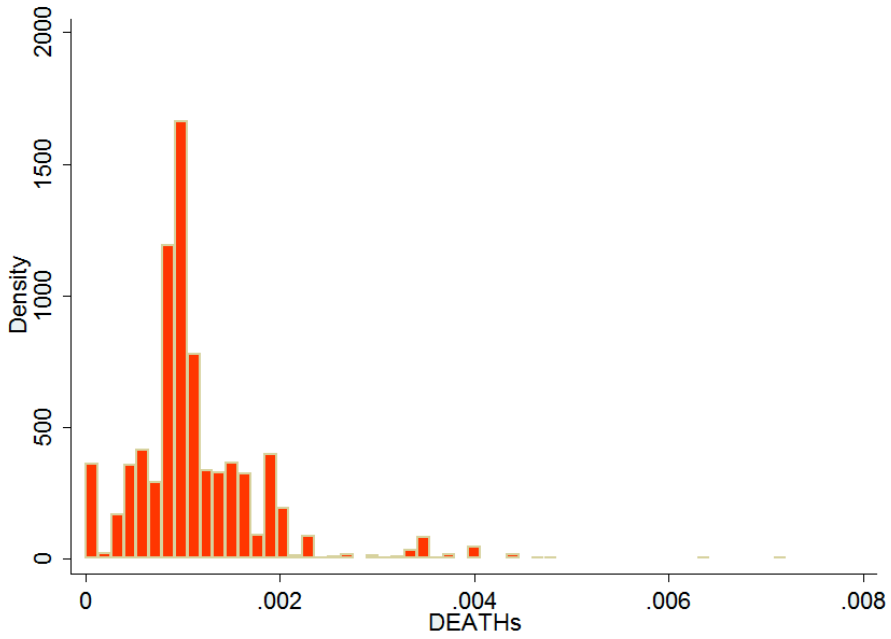
$$Relevant\ deaths_{jt} = \mathbb{1}Death[age <= 49, salary >= 50k]_{jt}, \tag{4}$$

where j is the bureau and t the year. Table 7 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 relevant age/salary population. It reveals a well-behaved distribution with 9% of the bureau-year observations being zero deaths and only a few extreme observations (to the exclusion of which our estimates are robust). The reason for the effectiveness of this variable as an instrument for competence can be deducted from Figure 4b. In this figure, we report the median value of *deaths* for each combination of age and salary levels. The median value of *deaths* increases monotonically in age, with salary having little effect (especially below the $100,000 salary, where most observations lie). This implies that for the group of individuals that we consider to be important for the well functioning of a bureau (i.e., young with a relative high salary), deaths are particularly unlikely. Thus their occurrence will be particularly disruptive. We return to this aspect after having introduced the other instrument.

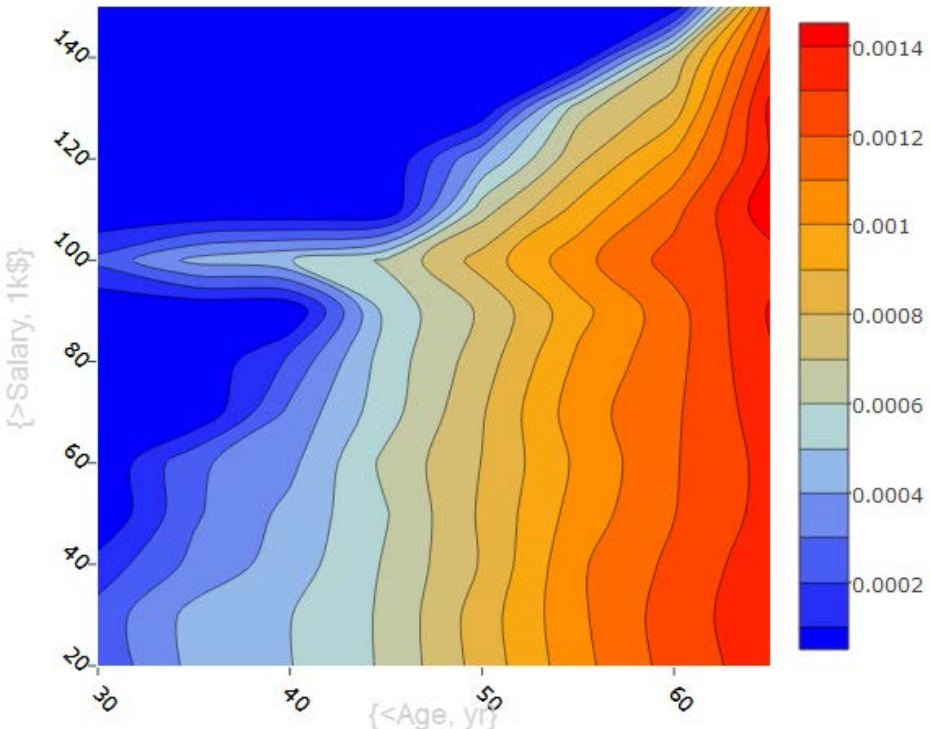
For a second instrument, we follow Bruce, Figueiredo and Silverman [2017] who suggest that the spatial proximity of a death event in the procurement agency can be relevant to contractual performance. By exploiting this variation, we construct our second instrument, *proximal deaths*: a binary variable indicating whether at least one death event among white-

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

(a) Histograms at contract level



(b) Median frequency by Age and Salary



Notes: In panel (a), we report the histogram of the ratio between the count of death events and the workforce dimension for each bureau. In panel (b), we report the median value of the ratio for each combination of age and salary.

Table 7: Instruments Summary Statistics

	(1) Mean	Median	S.D.	N
Relevant Deaths	0.91	1	0.30	131,686
Proximal Deaths	0.64	1	0.48	131,686

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

collars employees of the bureau awarding the contract has occurred in the same state of the contract's place of performance and in the same year of the contract awarding. To avoid ambiguities in interpreting a value of zero for this instrument, we exclude 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).[23]

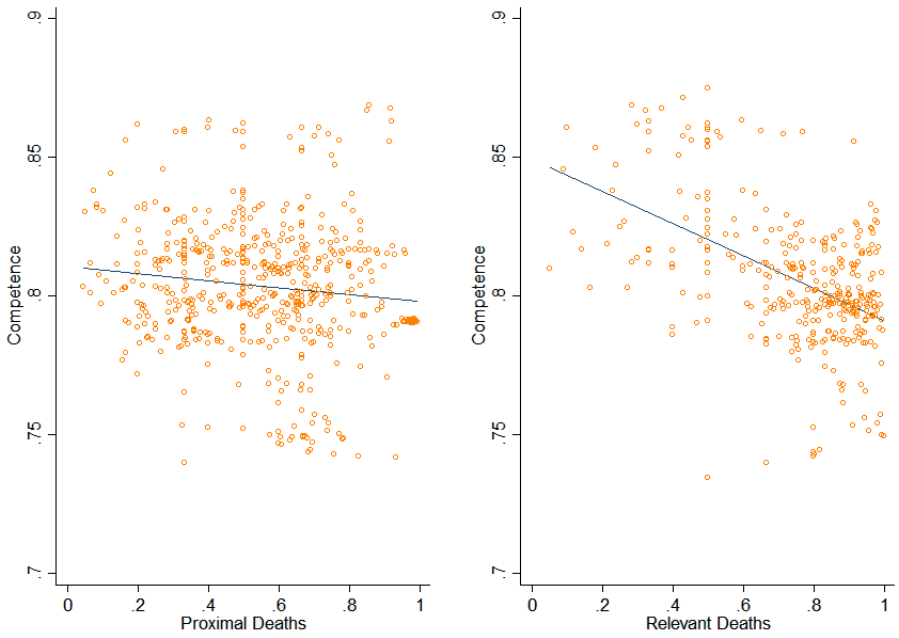
The relationship between deaths affecting the bureau and its competence is apparent from the "visual first stage" reported in Figure 5. This figure shows the relationship between our two instruments, *relevant deaths* and *proximal deaths*, and *competence*. A clear negative association is present in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

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.[24] This literature suggests that individuals in charge of high-responsibility tasks, such as managers, supervisors and team leaders, account for most of the competence in an institution. The validity of the instrument is supported by the fact that as-good-as-random separations of office managers negatively affect the competence level of the whole office through two obvious channels. First, a sudden separation determines a vacancy

[23]See also footnote 23 and the appendix table referenced therein.

[24]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 his coworkers. Other related papers using deaths as an exogenous shock 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.

Figure 5: Visual Representation of the First-Stage



Notes: Graphical representation of the relationships between *Competence* with *Proximal Deaths* - left panel - and *Relevant Deaths* - right panel. Observations are collapsed at cluster level (agency and service category).

of skills in terms of knowledge and prompt decisions of management. This is particularly true when high-level employees are involved or a complex task has to be accomplished. Second, the managerial literature evaluates the so-called onboarding effect, and estimates as the time a newly hired officer needs to reach full productivity to be eight months. The latter feature is clearly present in the federal workforce, as new hirings are notoriously slow due to the need to resort to public evidence procedures while transfers of workers are hindered by the limited ability to negotiate financial incentives. Regarding the first aspect, however, we do not have direct evidence on the causes of employees' deaths. An obvious concern is the potentially endogenous relationship existing between the quality of the workplace and deaths.

However, there are two pieces of evidence suggesting this feature is not prominent in the data. First, although FedScope does not allow to distinguish death's causes, we used 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.

Table 8: Death Occurrence Predictors

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

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

Nevertheless, evidence from the Bureau of Labor Statistics rules out this possibility as both the Survey of Occupational Injuries and Illnesses and the Census of Fatal Occupational Injuries show no suicides among federal managers in our sample years. Second, we perform a regression analysis (see Table 8) to identify the determinants of our two instruments. We find that while these deaths are associated in an intuitive way with education, salary and health security of the workplace, they are not associated with any of the procurement measures appearing at the top of Table 8: the total available budget, the number of pledged contracts, or the average amount of pledged contracts. The coefficients' magnitude is nearly zero and none reaches statistical significance.

VII Results

We begin the presentation of our results from Table 9 where we show the OLS estimates corresponding to equation (3). Panel A displays the results for *cost performance*, while panel B reports those for *time performance*. Columns 1-4 display the estimates for competence as we gradually expand the set of covariates. In all specifications, the effect of competence is highly significant and essentially identical for both time and cost performance. In line with the earlier discussion, their magnitude grows as we increase the set of controls.

Using the same set of controls of column 4, we also repeat the analysis substituting competence with cooperation (column 5) or with all the three components (cooperation, skills and incentives, column 6). We then proceed analogously in the two final columns of the table for the two PCA variables. Incentives are either insignificant (for cost performance) or significant but with a smaller magnitude than cooperation (for time performance). Skills are negative (and so is the PCA skill/incentive), denoting either a particularly severe downward bias or a more subtle issue regarding the interaction between cooperation and skills that we discuss below.

Among the other covariates, the coefficients on *Cost plus* and *Negotiation* are consistently negative and significant across all specifications. Their inclusion improves the R-squared as well as the magnitude of the competence coefficient, although only very slightly. These

Table 9: OLS regressions

Panel A: Cost Performance

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Q28	0.03**	0.03***	0.03**	0.04***				
	(0.01)	(0.01)	(0.01)	(0.01)				
Q20					0.07***	0.10***		
					(0.02)	(0.02)		
Q21						-0.03**		
						(0.01)		
Q23						-0.01		
						(0.02)		
PCA - Cooperation							0.08***	0.08***
							(0.02)	(0.02)
PCA - Skill/Incentive								-0.00
								(0.01)
Cost Plus		-0.59***	-0.59***	-0.58***	-0.59***	-0.59***	-0.59***	-0.59***
		(0.05)	(0.05)	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)
Negotiation		-0.07***	-0.07***	-0.07***	-0.07***	-0.08***	-0.08***	-0.08***
		(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
Bureau Experience			0.02	-0.01	-0.02	-0.01	-0.01	-0.01
			(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)
Bureau Size			0.04***	0.05***	0.05***	0.05***	0.05***	0.05***
			(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
R-squared	.16	.18	.18	.18	.19	.19	.19	.19

Panel B: Time Performance

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Q28	0.03***	0.03***	0.03***	0.04***				
	(0.01)	(0.01)	(0.01)	(0.01)				
Q20					0.06***	0.08***		
					(0.01)	(0.02)		
Q21						-0.08***		
						(0.01)		
Q23						0.05***		
						(0.01)		
PCA - Cooperation							0.05***	0.05***
							(0.01)	(0.01)
PCA - Skill/Incentive								0.01
								(0.01)
Cost Plus		-0.22***	-0.22***	-0.22***	-0.22***	-0.23***	-0.22***	-0.22***
		(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)
Negotiation		-0.08***	-0.09***	-0.09***	-0.09***	-0.09***	-0.09***	-0.09***
		(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
Bureau Experience			-0.05	-0.05	-0.05	-0.04	-0.05	-0.05
			(0.04)	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)
Bureau Size			0.04***	0.04***	0.04***	0.04***	0.04***	0.04***
			(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
R-squared	.12	.13	.13	.13	.13	.13	.13	.13
Observations	131686	131686	131686	131686	131686	131686	131686	131686
Amount&Duration FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Agency FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year FEs	No	No	No	Yes	Yes	Yes	Yes	Yes
State FEs	No	No	No	Yes	Yes	Yes	Yes	Yes

Notes: Both contract outcomes and bureau characteristics are replaced by their standard scores. Standard errors are clustered by agency and service category and are in parentheses. * Significant at the 10 percent level; ** Significant at the 5 percent level; *** Significant at the 1 percent level.

results are in line with the earlier discussion on how cost plus contracts and contracts awarded through negotiations are systematically associated with tasks exposed to more renegotiations. Interestingly, *Bureau experience*, which the earlier literature has used as a proxy for competence, is not significant when entered with our competence measure from the FEVS. But a positive and significant effect is found for *Bureau size*. These latter findings are also consistent with the evidence in Table 5 on performance persistency. Despite the inclusion of these controls, a concern with the potential downward bias in the OLS competence estimates remains.

Table 10

Panel A: Reduced-Form Regressions

	Cost Performance			Time Performance		
	(1)	(2)	(3)	(4)	(5)	(6)
Proximal Deaths	-0.06***		-0.06***	-0.06***		-0.05***
	(0.01)		(0.01)	(0.01)		(0.01)
Relevant Deaths		-0.06**	-0.04*		-0.06***	-0.05**
		(0.02)	(0.02)		(0.02)	(0.02)
R-squared	.18	.18	.18	.13	.13	.13

Panel B: First-Stage Regressions

	Competence			Cooperation				PCA	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Proximal Deaths	-0.14***		-0.12***	-0.12***	-0.13***	-0.11***	-0.11***	-0.13***	-0.13***
	(0.02)		(0.02)	(0.01)	(0.01)	(0.01)	(0.01)	(0.02)	(0.02)
Relevant Deaths		-0.18***	-0.16***	-0.02	0.05**	0.05**	0.06**	-0.04	-0.04
		(0.03)	(0.03)	(0.03)	(0.02)	(0.03)	(0.03)	(0.03)	(0.03)
R-squared	.6	.59	.6	.72	.81	.85	.85	.74	.74
Observations	131686	131686	131686	131686	131686	131686	131686	131686	131686

Notes: Panel A reports reduced-form regressions of *cost performance* and *time performance*, respectively, on the instruments. In Panel B we present the first stage for each IV regression from table 11. Both contract outcomes and bureau characteristics are replaced by their standard scores. Standard errors are clustered by agency and service category and are in parentheses. All models include controls for contract features (cost plus format and solicitation procedure), buyer characteristics (experience and yearly procurement budget), fixed effects for service category, agency, deciles for contract value and duration, year, and U.S. state of performance. * Significant at the 10 percent level; ** Significant at the 5 percent level; *** significant at the 1 percent level.

To address this concern, we implement an IV strategy based on the two instruments

presented above. Table 10 reports the reduced-form (panel A) and first-stage (panel B) estimates. For the former, the coefficients on both instruments enter with a negative and significant effect, both when used individually and jointly. For the first-stage regressions, we notice again the expected negative and significant effect of both instruments when competence is the outcome (columns 1-3). But one of the two instruments - relevant deaths - is either insignificant or positive when the outcome is cooperation (columns 4-7) or PCA cooperation (columns 8-9). This underscores the greater difficulty in devising valid instruments that are likely to satisfy the exclusion restriction for each one of the three individual components of competence. Indeed, while consider the IV estimates presented below as a valid causal assessment of the role of competence, to establish the relevance of cooperation we rely on a somewhat more articulated argument.

Table 11 reports our main IV estimates. The first three columns report the results for competence using as IV either relevant or proximal death (columns 1 and 2) or both (column 3). The set of controls is that of column 4 of the corresponding OLS regressions in Table 9. Compared to those OLS estimates, the magnitude of all IV estimates is substantially larger, by always exceeding the upper bound of the OLS 95% confidence interval.[25] The estimates remain quite similar between cost and time performance. Interestingly, despite the two instruments having a relatively low mutual correlation (15%), the estimates in column 1 and 2 are statistically identical. This is suggestive of these estimates plausibly representing an average treatment effect and not a LATE.[26]

[25]Although there is also the possibility of reverse causality that prevents us from interpreting the OLS coefficient on the bureau's characteristic as a causal effect (i.e., better contract performance inducing a more positive answer to the FEVS question on bureaus' competence), the magnitude of OLS estimates are smaller than the IV estimates. This suggest that the source of upward bias is less relevant than that of downward bias (i.e., more competent bureaus are associated with more complex procurement projects). The possibility of reverse causality, nevertheless, indicates that an IV strategy like the one we undertake is preferable to a different approach based on first regressing performance on bureau fixed effects and, subsequently, regressing these fixed effects on bureau competencies.

[26]IV estimates differing when using different instruments, is an indication of heterogeneous treatment effects due to different compliers associated with the instruments. This is the LATE interpretation of an IV estimator given by 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. Further evidence supporting the ATE interpretation of the estimates is visible in Figure A.2, where the IV estimates are pretty stable throughout the entire span of possible instruments obtainable by the combination of salary and age.

Table 11: IV regressions

Panel A: Cost Performance

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Competence	0.46***	0.31**	0.40***						
	(0.10)	(0.13)	(0.08)						
Cooperation				0.55***	0.42***	0.46***	0.43***		
				(0.12)	(0.10)	(0.12)	(0.11)		
PCA - Cooperation								0.51***	0.52***
								(0.10)	(0.10)
Weak Id. F-Test	57.67	36.45	40.11	31.62	50.14	41.41	43.47	34.37	34.6
Underid. F-Test	54.96	33.35	70.34	56.85	66.85	61.7	62.61	56.5	57.42
Overid. F-Test	0	0	.95	1.47	6.59	6.09	7.02	.79	.75
Centered R-squared	.11	.16	.13	.12	.17	.16	.17	.14	.14

Panel B: Time Performance

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Competence	0.40***	0.32**	0.37***						
	(0.10)	(0.13)	(0.08)						
Cooperation				0.49***	0.35***	0.38***	0.34***		
				(0.11)	(0.10)	(0.12)	(0.11)		
PCA - Cooperation								0.46***	0.46***
								(0.10)	(0.10)
Skills	No	No	No	No	Yes	No	Yes	No	No
Incentives	No	No	No	No	No	Yes	Yes	No	No
PCA Sk./in.	No	No	No	No	No	No	No	No	Yes
IV: Proximal Deaths	Yes	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes
IV: Relevant Deaths	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Weak Id. F-Test	57.67	36.45	40.11	31.62	50.14	41.41	43.47	34.37	34.6
Underid. F-Test	54.96	33.35	70.34	56.85	66.85	61.7	62.61	56.5	57.42
Overid. F-Test	0	0	.28	2.38	8.38	7.23	8.85	1.43	1.33
Centered R-squared	.08	.1	.09	.08	.12	.12	.12	.09	.09
Observations	131686	131686	131686	131686	131686	131686	131686	131686	131686

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

On the basis of these findings, we consider the estimates in column 3 as our baseline estimate of the effect of competence on performance. To offer a more transparent economic interpretation of the estimates, we can then consider what would happen if we were to use them to infer the effect of lifting the level of *competence* from all bureaus to that of the

bureaus at the 90th percentile of this distribution. This implies a saving of $102,619 on average per contract, or around $13.5 billions in total across all contracts in the dataset. Moreover, this would imply a saving of 54.4 days in effective execution time, corresponding to 7.2 million days across all the contracts in the dataset. The amounts are economically sizable and compare well to what the literature has indicated could be achieved by optimizing either the incentives given to suppliers (for instance through the choice between cost plus and fixed price contracts) or the type of awarding procedures (for instance through the selection of negotiations versus competitive auctions).

The last columns of Table 11 report the IV estimates for cooperation (columns 4-7) and PCA cooperation (columns 8-9). We are particularly interested in assessing the role of cooperation due to the key role that cooperation is believed to have in successful procurement, as discussed earlier. While simply applying the same IV strategy used for competence is problematic due to potential violations of the exclusion restriction,[27] we find support for a positive and significant effect of cooperation on time and cost performance based on what we report in columns 4-9. In column 4, cooperation replaces competence using the same model specification and instruments of column 3: the estimates are positive and highly significant, with a magnitude that is slightly larger than that of competence in column 3 (statistically both estimates lie within each other's 95% confidence interval). In the following columns, we expand the model specification to include skills (column 5), incentives (column 6), or both (column 7). For both cost and time performance, the estimate of cooperation becomes slightly lower, moving even closer to that of competence. Qualitatively, the same results hold in columns 8 and 9 where the endogenous regressor is PCA cooperation (and where, in column 9, we control for PCA skill/incentives). The consistency among all these results on cooperation and their close resemblance to those on competence are reassuring that our analysis is indeed able to reveal a positive and significant role for cooperation on procurement performance.

[27]This is also indicated by the various tests on the instruments reported in the bottom portion of Table 11. There we report the usual tests of IV models: F-statistics for the test of whether our instruments have zero coefficients, F-statistic of the LM underidentification test and Sargan's over-identification test statistic (see details in the table note). While the instruments pass all tests for competence, the latter test fails for cooperation, as expected given that the instruments correlate with other components of competence.

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

Second, we consider the robustness to alternative subsamples. Table 12 reports the IV obtained when estimating on different subsamples models analogous to those in columns 3 of Table 11. For convenience, we report in the first column of each panel the baseline estimate from Table 11. In the second and third column, we restrict the sample to exclude the most extreme observations with cost and time performance lower than 0.1 and 0.25, respectively.[28] While the magnitudes of the estimates declines with a more stringent definition of the sample, the signs and significance are broadly in line with those in the baseline estimates. In columns 4-6, we iteratively discard one-per-time all the observations belonging to the three largest agencies: DOD, DVA, and GSA, respectively. All estimates appear robust. Finally, in column 7, we restrict the sample to fixed price contracts not awarded through negotiated procedures. These are the contracts for which renegotiations should be particularly problematic, as their simple nature advised (at the tendering stage) against the use of a cost plus contract and a negotiated procedure. For these contracts we still find that greater competence has a strong, positive effect, in line with our baseline estimates.

The third set of robustness checks involves alternative measures for both procurement performance and bureau competencies. Regarding the former, contracts amendment records

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

Table 12: Robustness Checks: Sample Selection

Panel A: Competence - Cost Performance

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Competence	0.40***	0.34***	0.18***	0.31***	0.33***	0.37***	0.40***
	(0.08)	(0.07)	(0.05)	(0.09)	(0.09)	(0.08)	(0.08)
Centered R-squared	.13	.13	.13	.15	.17	.14	.09
Weak Id. F-Test	40.11	40	38.25	28.89	43.55	39.25	39.11
Underid. F-Test	70.34	70.17	68.11	50.5	76.3	69.29	77.54
Overid. F-Test	.95	1.75	3.4	.01	.71	.28	2.4

Panel B: Competence - Time Performance

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Competence	0.37***	0.35***	0.17***	0.20**	0.24***	0.34***	0.40***
	(0.08)	(0.07)	(0.05)	(0.08)	(0.08)	(0.08)	(0.09)
Centered R-squared	.09	.08	.11	.14	.14	.09	.07
Observations	131686	125575	108395	105468	98444	121975	91507
Weak Id. F-Test	40.11	40	38.25	28.89	43.55	39.25	39.11
Underid. F-Test	70.34	70.17	68.11	50.5	76.3	69.29	77.54
Overid. F-Test	.28	.57	.87	.33	.21	.01	.15

Notes: Column 1 reports the baseline estimate for *Cooperation* or *PCA Cooperation* from table 11. In Column 2 and 3 we do not consider contracts associated to cost and time performance lower than 0.1 and 0.25, respectively. In columns 4,5, and 6, we discard DOD, DVA, and GSA, respectively. Finally, column 7 refers to contracts priced as fixed price and not awarded through negotiated procedures. Both contract outcomes and bureau characteristics are replaced by their standard scores. Standard errors are clustered by agency and service category and are in parentheses. All models include controls for contract features (cost plus format and solicitation procedure), buyer characteristics (experience and yearly procurement budget), fixed effects for service category, agency, deciles for contract value and duration, year, and U.S. state of performance. * Significant at the 10 percent level; ** Significant at the 5 percent level; *** significant at the 1 percent level.

in the FPDS data are classified according to the reason for contract modification, which can be *in-scope* or *out-of-scope* revisions.[29] In line with other studies, we have considered *in-scope* amendments only. 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 (results reported in Table A.4), we find very similar results to those in our baseline estimates. Second, regarding the robustness to alternative measures of competence, we consider different implementations of the PCA relative to that presented earlier. That is, we assess the robustness of the estimates in model

[29] 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*.

(8) of our IV baseline to PCA measures constructed through different methodologies. As shown in the detailed discussion in the appendix, the estimates (reported in Table A.3) obtained with both parametric and non-parametric versions of the PCA, with both iterated principal factors and maximum likelihood, all conform closely with our baseline estimates.

The fourth and final robustness check involves inference. Recent research by Young, Alwyn [2017], indicates that IV studies sometimes have inference problems driven by the finite sample estimator performance. Our large sample size limits this problems but, nevertheless, in the appendix we report standard estimates obtained via bootstrap. We replicate our IV analysis by drawing 200 bootstrap samples in a fashion consistent with the error dependence within our cluster of observations (agency and service category) and independence across observations. The findings confirm the baseline estimates presented above.

VIII Conclusions

This paper is the first comprehensive study of the determinants of the impact of public employees' competencies on public procurement outcomes. 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. Our identification strategy exploits the exogeneity of death events involving public officials to allow for a causal interpretation of bureaus' competencies on procurement performance.

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

Our second main result, to be taken somewhat more cautiously in terms of causal inter-

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

We see several avenues for further research. First, given the crucial role we have identified for competence, it would be important to develop an understanding of what factors can promote this trait within public offices, especially with regard to the ability to maintain cooperation among employees. Although a detailed exploration of this issue is beyond the scope of this paper, our data are indicative of the key role played by young managers. To further explore this aspect, we report in Figure 6 plots of how our baseline estimates would differ with instruments constructed by altering the definition of the *relevant deaths* instrument. In the baseline estimates, the median values of age and salary are the cut offs used to select relatively low age and high salary workers. In Figure 6 we report the IV estimates interactively replacing the *relevant deaths* instrument 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. The results in the figure indicate that for all age cutoffs up until the age of 50, deaths of workers with higher than median salary produce estimated effects of competence on performance that are statistically larger than the corresponding ones for below median salary workers. Above age 50, the estimates become statistically identical. This evidence is further corroborated by a full heterogeneity analysis presented in the appendix where we explore all combinations of age and salary. This analysis is indicative of interesting heterogenous effects across employees that might even offer a simple policy prescription to help low-performing bureaus to improve: infuse relatively young, competent and well paid managers.[30] A similar policy prescription is offered by Bertrand et al. [2016]

[30]This type of analysis thus also allow us to relate our estimates to an important strand of the literature

who, in the rather different context of India's higher bureaucrats, find that officers that enter older and in larger cohorts are less effective and more likely to be suspended. Career concerns of public officials thus seem to play key roles in how the economy will be affected by the bureaucracy. Related to this, it will also be interesting in future work to try to understand whether our results on the importance of cooperation may be related to the role that trust is thought to play in organizations (see La Porta et al. [1997]), as cooperation and trust are typically closely associated.

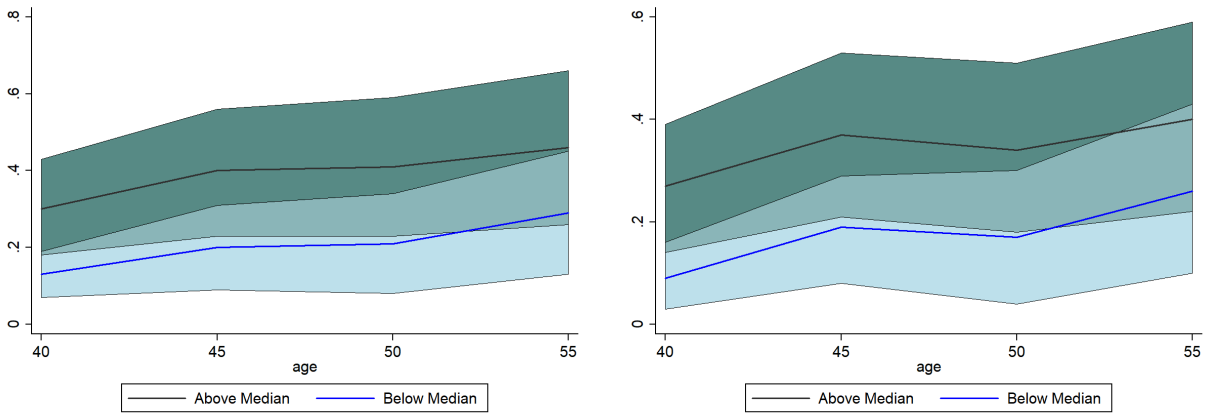
Second, it would be relevant to investigate whether buyers who perform better do so because they are better at selecting good contractors or because they are more capable of monitoring contractors while performing the contract. In other words, what is the relative importance of selecting good quality contractors versus managing the post-selection contractual relationship with them (selection vs moral hazard)? In the context of Russian procurement, Best, Hjort and Szakonyi [2017] find that more effective procurers are able to attract different bidders relative to less effective ones. For US federal procurement, Giuffrida and Rovigatti [2017] exploit a regulatory change determining which contracts are subject to greater public monitoring to study the issue of selection vs moral hazard. When they integrate in their analysis the measures of bureaus' competence that we developed in this study, they find that public oversight negatively affects outcomes, in particular for the least competent buyers. This suggests that a fruitful avenue for future research would involve a more in-depth evaluation of the channels through which more competent buyers improve procurement outcomes. In this respect, a more holistic approach to contracting that accounts for all the phases of the procurement process, from project design stage to the contract award and ex post management, is likely to be essential.

Third, it would be interesting to study procurement in other sectors where other relevant outcome measures could be analyzed, such as patients' outcomes in the health procurement, or patents registration and citations in innovation procurement (see Decarolis et al. [2017] for some preliminary evidence on this).

that tries to quantify how specific groups of employees affect outcomes of the teams/units/firms they work for (see Jäger [2017] for a recent study in this area).

To conclude, whilst economists have emphasized the importance of government quality, the lack of data has traditionally limited the possibilities to quantify the effects of both public employees and their organizations. Most of recent empirical studies on bureaucracies have focused developing countries. Although often based on accurate field experiments, these studies are inherently specific to environments with a low government effectiveness. Yet, our paper suggests that the degree of heterogeneity can be large even in developed countries and so too can be the gain from policy initiatives directed to reduce government inefficiencies. This calls for more empirical economic research aiming to understand how government units work and perform also under strong institutions.

Figure 6: Heterogeneity of IV Estimates for Competence Components



(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 11. 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.

References

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

Angrist, Joshua D., Guido W. Imbens, and Donald B. Rubin. 1996. "Identification of Causal Effects Using Instrumental Variables." *Journal of the American Statistical Association*,

91(434): 444–455.

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

Bloom, Nicholas, and John Van Reenen. 2007. "Measuring and Explaining Management Practices Across Firms and Countries." *The Quarterly Journal of Economics*, 122(4): 1351–1408.

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

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

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

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

Chong, Alberto, Rafael La Porta, Florencio Lopez-de Silanes, and Andrei Shleifer. 2014. "Letter Grading Government Efficiency." *Journal of the European Economic Association*, 12(2): 277–299.

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

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

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

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

Decarolis, F., G. de Rassenfosse, L. Giuffrida, E. Iossa, V. Mollisi, E. Raiteri, and G. Spagnolo. 2017. "Buyer Competencies and Innovation Procurement." *Working Paper.*

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

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

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

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

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

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

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

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

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

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

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

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

La Porta, Rafael, Florencio Lopez-de Silane, Andrei Shleifer, and Robert W. Vishny. 1997. "Trust in Large Organizations." *American Economic Review*, 87(2): 333–38.

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.

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

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

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

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

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

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

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

Williamson, Oliver E. 1975. *Markets and Hierarchies, Analysis and Antitrust Implications: A Study in the Economics of Internal Organization.* New York:Free Press.

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

World Bank. 2017. "Benchmarking Public Procurement."

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

For Publication on the Authors' Web Pages

Web Appendix

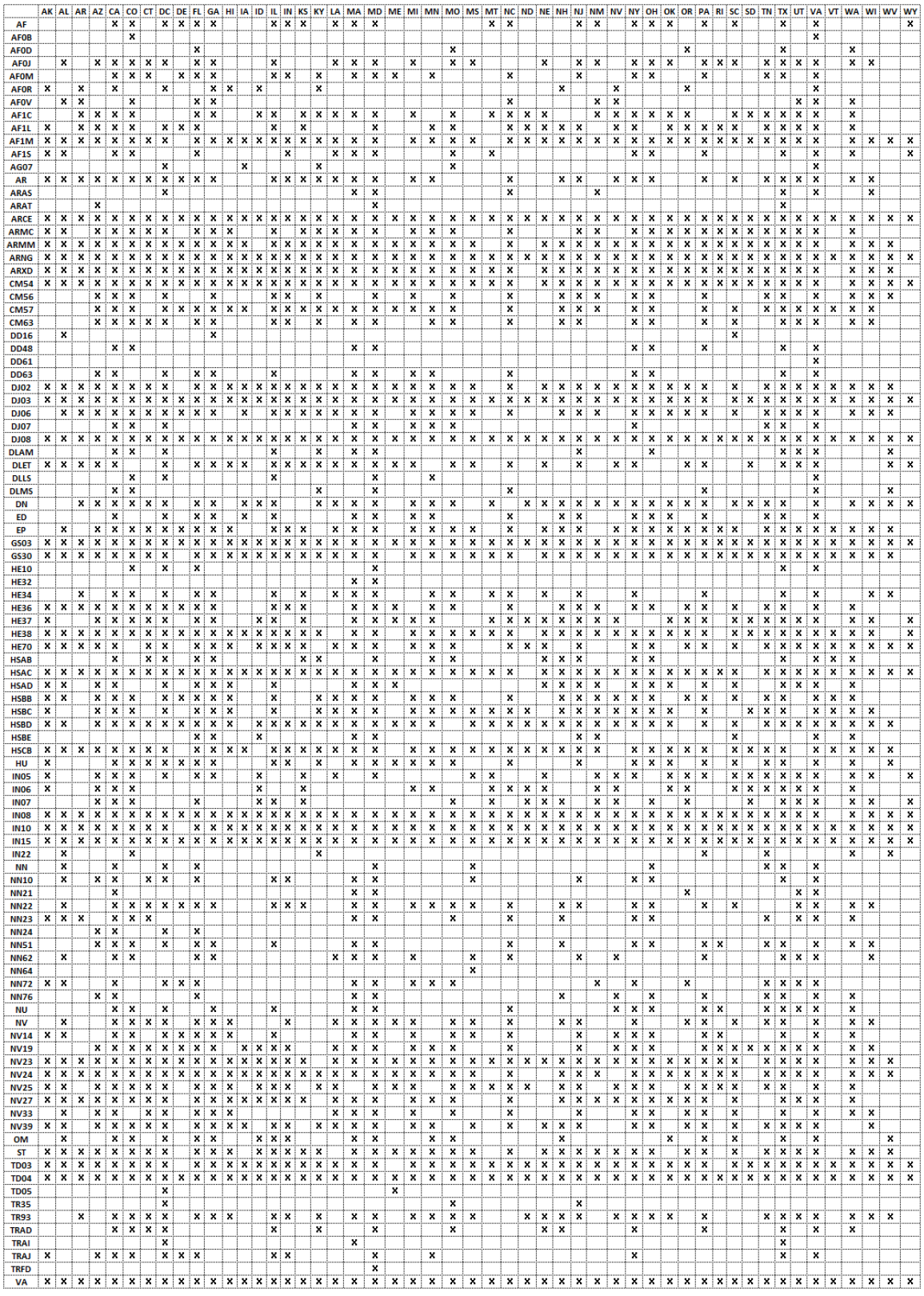
I Sample Selection

For the purpose of our analysis, we will focus on the years where the FEVS has an yearly frequency and where the two datasets, FEVS and FPDS, overlap. Thus, we focus on the years between 2010 and 2015. Although the data contain contracts for both supplies, R&D projects, services, and constructions, the first two are ruled out of the analysis. Supplies typically do not exhibit any ex post variation in price or delivery time, while the outcome of R&D contracts cannot be reasonably assessed in terms of costs and duration. Thus, for our analysis we focus exclusively on the procurement of services and constructions and refer jointly to these as *services*.[31] We restrict our sample to those contracts awarded through competitive solicitations because the effect of the treatments would otherwise not be observable. We consider as competitive a lot for which the extent of competition is labelled "Full and open"; those whose participation is not set aside to any specific group of firms; those at or below the micro-purchase acquisition threshold - \$3,000 - as allocated without soliciting competitive quotations. FPDS contains every base contract that exceeds an individual transaction value of \$2,500. We focus on contracts worth more than \$25,000.[32] In non-competitive awardings, the participation criteria restrict the competition *ex-ante* to dimensions other than quality.

[31] 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. Constructions include: construction, maintenance, repair, alteration of structures/facilities.

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

Figure A.1: Federal Employees by State



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

For similar reasons, we focus on contracts whose tasks are such that the vendor can influence the outcome metrics through effort. Supply contracts do not allow for renegotiations. Hence, for these contracts our measure of performance does not proxy for outcome quality and we exclude them from the analysis.[33] The same applies to the subcategory "Lease or Rental of Equipment, Structures, or Facilities". We consider only contracts awarded within the U.S. border. Finally, the sample includes only contracts awarded in states where the awarding bureau has at least one employee. This restriction leads us to drop 4% of the sample, but serves to insure that we can match the locations of the bureaus, local offices and of the contracts that they are likely to supervise. Figure A.1 reports the location of each bureau by indicating with an "X" the sate in which they employ at least one white-collar worker.

II Robustness Checks

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

- Table A.1: LIML estimates. As is standard for checking for weak instruments, LIML estimates are provided as a robustness for the 2SLS estimates presented in the main text. All the point estimates are very close to those in Table 11, thus limiting concerns about a weak instruments problem.

- Table A.2: control function estimates. Since $performance^{C}_{ijt}$ and $performance^{T}_{ijt}$ are fractional variables on (0,1][34] 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 $\mathbb{E}(y|\mathbf{x}) = \Phi(\mathbf{x}\gamma)$.[35] This fractional probit model handles continuous endogenous explanatory through a *two-step control function approach*. The control function approach relies on similar identification conditions of the

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

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

[35] 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 $\mathbf{x}$.

linear IV described in the main text.[36] Table A.2 presents the estimates obtained via control function, using the same four instruments used for the main analysis. All the qualitative implications described for our baseline estimates are confirmed. The significance of the first stage residuals leads further support to our endogeneity concerns.[37]

- Table A.3: robustness on the construction of the PCA. In Table A.3, we report the IV estimates obtained with three common alternatives of factor analysis. Suppose that we have a partitioning of the variance of vector $\mathbf{y}$ into a component due to the common factors, $\Lambda\Lambda^{'}$, which is a vector of the so-called communalities, h_i^2 , and another, Ψ, called the residual variance. In particular, Ψ is a diagonal matrix of residual variances, $(\psi_1, \psi_2, \cdots, \psi_p)$.[38] In Table A.3, the first column reports for comparison the factor estimates from Table 11. This is iterated principal factors.[39] Columns

[36]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. $\mathbb{E}(performance_{ijt}|competence_{jt}, \hat{cf}_{jt}, \mathbf{X}) = \Phi(\beta\, competence_{jt} + \zeta\, \hat{cf}_{jt} + \theta\, X_{ijt} + \iota_j + \kappa_t)$.

[37]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.

[38]To see where this comes from, consider that the factor analysis model is a linear combination of the underlying latent variables, $f_1, f_2, \cdots, f_m$, these are hypothetical and not observed. For the variables $y_1, y_2, \cdots, y_p$ in any of the observation vectors sample, the model is defined as: $y_i = \lambda_{1i} f_i + \lambda_{2i} f_2 + \cdots + \lambda_{im} f_m + \epsilon_i$ for all $i = 1, ..., p$. This can be expressed more compactly in matrix notation: $\mathbf{y} = \Lambda f + \epsilon$, where $\boldsymbol{y}$ is a standardized outcome vector and $\epsilon$is a random vector of error terms due to idiosyncratic factors. Assume latent variables are independent of each other and of the error terms. Factors $j = 1, 2, \cdots, m$, have the expected value of the jth factor, $E(f_j) = 0$. The variance of the factor model is, $var(f_j) = 1$ and the covariance of two factor models f_j and f_k is $cov(f_j, f_k) = 0$ with $j \neq k$. The error terms ϵ_i are independent of each other, $cov(\epsilon_i, \epsilon_j) = 0$ with $E(\epsilon) = 0$ and $var(\epsilon_i) = \psi_i$. The covariance of the error terms and the factor is $cov(\epsilon_i, f_j) = 0$, where the assumption of $cov(\epsilon_i, f_j) = 0$ implies that the factors represent all the correlations among the outcome vector $\mathbf{y}$. Thus, factor analysis accounts for the explained components of the variance. With the stated assumptions, the variance of $\mathbf{y}$ can be defined as: $S = \Lambda\Lambda^{'} + \Psi$.

[39]Here, the squared multiple correlation coefficients are used as starting points of the communalities, but better estimates are obtained through iteration of $\hat{h}_i^2 = [j = 1]m\sum\hat{\lambda}_{ij}^2$. The values of $\hat{h}_i^2$ are then substituted into $S - \hat{\Psi}$ and a new $\hat{\Lambda}$ is calculated. The stopping rule is such that $\hat{h}_i^{(i+1)} - \hat{h}_i^{(i)} < k\ \forall i$, where k is a very small value.

(2) and (3) show a parametric and non parametric version of the PCA,[40] column (4) uses maximum likelihood.[41] The principal factor method and iterated principal factor method usually produce results close to the principal component method if either the correlations or the number of variables is large.[42] The two methodologies in column (2)-(3) have better statistics of the underidentification rank tests, but poorer results for the overidentification. The loadings from each method are rather similar and don't differ significantly. However, the factors resulting from the principal component method and the PCA methodology explain 87% of the cumulative variance compared to 99% from the principal factor method and iterated principal factor method. This drastic difference is due to the presence of negative eigenvalues, so we are more confident in our baseline PCA, and the other principal component methods.

[40]In Column 2, the approach of the principal factor method or the principal axis method starts with an estimate of $\hat{\Psi}$ and factors $S-\hat{\Psi}$. Rearranging the sample covariance yields: $S-\hat{\Psi}=\hat{\Lambda}\hat{\Lambda}'$. The diagonal of S is subtracted of the residual variances, $\hat{\psi}_i$, to obtain:

$$\hat{\Lambda}\hat{\Lambda}'=\begin{bmatrix}\hat{h}_1^2 & s_{12} & \cdots & s_{1p}\\ s_{21} & \hat{h}_2^2 & \cdots & s_{2p}\\ \vdots & \vdots & \ddots & \vdots\\ s_{p1} & s_{p2} & \cdots & \hat{h}_p^2\end{bmatrix}$$

An initial estimate of the communalities, $\hat{h}_i^2$, is made using the squared multiple correlation between the observation y_i and the other $p-1$ variables: $\hat{h}_i^2=1-\frac{1}{s^{ii}}$, where s^{ii} is the ith diagonal element of S^{-1}. Identification of $\hat{\Lambda}$ is performed again through spectral decomposition of $S-\hat{\Psi}$ by retaining all the eigenvalues which pass the stated rule.

In Column 3, the approach of the principal component method is to calculate the sample covariance matrix and then find an estimator, denoted $\hat{\Lambda}\hat{\Lambda}^{'}$, that can be used to factor S. The vector Ψ is not estimated and set to zero, so that the communalities are treated as all 1, meaning that there are no unique factors. Decomposition of S is performed through spectral decomposition of S. $\hat{\Lambda}$ is identified retaining during the spectral decomposition all the eigenvalues and the relative eigenvectors which pass some specific rule.

[41]The maximum likelihood method in column 5 assumes that the data are multivariate normal distributed. This is going to be a drawback to this method, since data in social sciences are usually discrete and bounded. The maximum likelihood estimator for the communalities, $\Lambda\Lambda^{'}$, and the specific variances $\hat{\Psi}$ are obtained by finding, $\hat{\Lambda}$, and $\hat{\Psi}$ that maximizes the following log likelihood:

$$L\left(\Lambda,\Psi\right)=-\frac{np}{2}\log\left(2\pi\right)-\frac{n}{2}\log\left|\Lambda\Lambda'+\Psi\right|-\frac{1}{2}\overset{[}{n}=1]N\sum\mathbf{Y}_n'\left(\Lambda\Lambda'+\Psi\right)\mathbf{Y}_n$$

where N is the observation length of the observed vector $\mathbf{y}_i$. To obtain a unique solution a possible additional constraint is that $\Lambda'\Psi^{-1}\Lambda$ is a diagonal matrix.

The rotation criterion is varimax and applies to all the procedures. Varimax has the nice feature to confine loadings (λ_{im}) to the extremes of the space dimensions, such that factors are easy to identify.

[42]Rencher, Alvin C. 2003. Principal Component Analysis. Methods of Multivariate Analysis, 380407. John Wiley & Sons, Inc.

- Table A.4: alternative measurement of procurement performance. The estimates in Table A.4 are the analogous to our baseline estimates, but are obtained with outcome variables calculated with the definition of contract renegotiations adopted in Kang and Miller [2017]. Compared to our definition, a broader set of contract modifications are included to calculate the final time and cost of the contract. Nevertheless, all the qualitative results from our baseline are robust if compared with the estimates reported in Table A.4.

- 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.5 we replicate our IV analysis by drawing 200 bootstrap samples in a fashion consistent with the error dependence within our cluster of observations (agency and service 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.

- Figure A.2: heterogeneous effects of different employees. These figures further expand the analysis reported in Figure 6 by reporting the IV estimates obtained by varying the IV construction for all possible combinations of age and salary. Figure A.2 reports 6 panels, corresponding to the IV estimates for model (3) of the baseline IV estimates. The estimates for cost performance are on the left-hand side panels, those for time performance on the right. The top panel regards Competence, the following Cooperation, and PCA-Cooperation appears last. The colors distinguish, within each plot, the areas where the point estimates are lower (blue) from those where they are higher (red). The most relevant feature is the confirmation of what has already been described in relation to Figure 6: for cooperation the estimated effects are higher whenever we restrict the set of relevant deaths to those of the highest paid individuals.

Table A.1: LIML Estimates

Panel A: Cost Performance

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Competence	0.46***	0.31**	0.40***					
	(0.10)	(0.13)	(0.08)					
Cooperation				0.56***	0.43***	0.47***		
				(0.12)	(0.10)	(0.12)		
PCA - Cooperation							0.52***	0.52***
							(0.10)	(0.10)
Centered R-squared	.11	.16	.13	.12	.16	.16	.14	.13
Weak Id. F-Test	57.67	36.45	40.11	31.62	50.14	41.41	34.37	34.6
Underid. F-Test	54.96	33.35	70.34	56.85	66.85	61.7	56.5	57.42
Overid. F-Test	0	0	.95	1.47	6.54	6.01	.78	.75

Panel B: Time Performance

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Competence	0.40***	0.32**	0.37***					
	(0.10)	(0.13)	(0.08)					
Cooperation				0.50***	0.36***	0.39***		
				(0.12)	(0.10)	(0.12)		
PCA - Cooperation							0.46***	0.47***
							(0.11)	(0.11)
Centered R-squared	.08	.1	.09	.08	.12	.11	.09	.09
Skill	No	No	No	No	Yes	No	No	No
Incentive	No	No	No	No	No	Yes	No	No
PCA Sk./In.	No	No	No	No	No	No	No	Yes
Observations	131686	131686	131686	131686	131686	131686	131686	131686
Weak Id. F-Test	57.67	36.45	40.11	31.62	50.14	41.41	34.37	34.6
Underid. F-Test	54.96	33.35	70.34	56.85	66.85	61.7	56.5	57.42
Overid. F-Test	0	0	.28	2.36	8.34	7.15	1.42	1.32

Notes: The 2SLS IV analysis presented in Table 11 is replicated by using Limited Information Maximum Likelihood. For more details, see notes from Table 11.

Table A.2: Control Function Estimates

Panel A: Cost Performance

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Competence	0.10***	0.07	0.09***					
	(0.02)	(0.04)	(0.02)					
Cooperation				0.12***	0.10***	0.11***		
				(0.02)	(0.02)	(0.02)		
Skill					-0.04***			
					(0.01)			
Incentive						-0.05***		
						(0.01)		
PCA - Cooperation							0.11***	0.11***
							(0.02)	(0.02)
PCA - Skill/Incentive								-0.00
								(0.00)
FS Residual	-0.09***	-0.06	-0.08***	-0.11***	-0.08***	-0.09***	-0.10***	-0.10***
	(0.02)	(0.04)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)

Panel B: Time Performance

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Competence	0.13***	0.09*	0.12***					
	(0.03)	(0.06)	(0.03)					
Cooperation				0.15***	0.13***	0.14***		
				(0.03)	(0.03)	(0.04)		
Skill					-0.06***			
					(0.01)			
Incentive						-0.06***		
						(0.02)		
PCA - Cooperation							0.14***	0.14***
							(0.03)	(0.03)
PCA - Skill/Incentive								0.01
								(0.00)
FS Residual	-0.11***	-0.08	-0.10***	-0.13***	-0.09***	-0.11***	-0.12***	-0.12***
	(0.03)	(0.06)	(0.03)	(0.03)	(0.03)	(0.04)	(0.03)	(0.03)
Observations	131686	131686	131686	131686	131686	131686	131686	131686

Notes: Table 11 is replicated by using the two-step fractional probit approach proposed in Wooldridge [2002]. For more details, see notes from Table 11.

Table A.3: Alternative PCA Measures: IV Estimates

Panel A: Cost Performance

	(1)	(2)	(3)	(4)
PCA Cooperation	0.52***	0.25***	0.22***	0.42***
	(0.10)	(0.08)	(0.08)	(0.09)
Centered R-squared	.14	.18	.18	.16
Overid. F-Test	.75	13.99	15.39	4.46

Panel B: Time Performance

	(1)	(2)	(3)	(4)
PCA Cooperation	0.46***	0.20**	0.17**	0.36***
	(0.10)	(0.08)	(0.08)	(0.09)
Centered R-squared	.09	.13	.13	.11
Overid. F-Test	1.33	14.67	15.9	5.61
Weak Id. F-Test	34.6	54.63	51.66	43.79
Underid. F-Test	57.42	79.40	86	67.39
Observations	131686	131686	131686	131686

Notes: Specification (8) of Table 11 is replicated using different methodologies to estimate the factor analysis. Columns (1) presents our preferred PCA specification, i.e. iterated principal factors employed in column (8) of Table 11. Column (2) is principal factor; Columns (3) is principal-component factor; Columns (4) maximum-likelihood factor. Both contract outcomes and bureau characteristics are replaced by their standard scores. Standard errors are clustered by agency and service category and are in parentheses. All models include controls for contract features (cost plus format and solicitation procedure), buyer characteristics (experience and yearly procurement budget), fixed effects for service category, agency, deciles for contract value and duration, year, and U.S. state of performance. * Significant at the 10 percent level; ** Significant at the 5 percent level; *** significant at the 1 percent level.

Table A.4: Alternative Performance Measures: IV Estimates

Panel A: Cost Performance

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Competence	0.45***	0.29**	0.38***					
	(0.10)	(0.13)	(0.08)					
Cooperation				0.54***	0.42***	0.46***		
				(0.12)	(0.10)	(0.12)		
PCA - Cooperation							0.50***	0.51***
							(0.10)	(0.10)
Centered R-squared	.11	.16	.14	.12	.16	.16	.14	.14
Weak Id. F-Test	57.52	36.39	40.02	31.51	50.17	41.31	34.38	34.62
Underid. F-Test	54.89	33.31	70.26	56.76	66.91	61.63	56.6	57.52
Overid. F-Test	0	0	1.17	1.18	5.92	5.48	.59	.57

Panel B: Time Performance

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Competence	0.39***	0.31**	0.35***					
	(0.10)	(0.13)	(0.08)					
Cooperation				0.47***	0.34***	0.37***		
				(0.11)	(0.10)	(0.12)		
PCA - Cooperation							0.44***	0.44***
							(0.10)	(0.10)
Centered R-squared	.08	.1	.09	.08	.12	.12	.09	.09
Weak Id. F-Test	57.52	36.39	40.02	31.51	50.17	41.31	34.38	34.62
Underid. F-Test	54.89	33.31	70.26	56.76	66.91	61.63	56.6	57.52
Overid. F-Test	0	0	.29	2.18	7.86	6.77	1.32	1.24
Observations	131470	131470	131470	131470	131470	131470	131470	131470
Q21	No	No	No	No	Yes	No	No	No
Q23	No	No	No	No	No	Yes	No	No
PCA Sk./In.	No	No	No	No	No	No	No	Yes

Notes: Results from Table 11 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 agency and service category and are in parentheses. All models include controls for contract features (cost plus format and solicitation procedure), buyer characteristics (experience and yearly procurement budget), fixed effects for service category, agency, deciles for contract value and duration, year, and U.S. state of performance. * Significant at the 10 percent level; ** Significant at the 5 percent level; *** significant at the 1 percent level.

Table A.5: IV regressions - Cluster Bootstrap

Panel A: Cost Performance

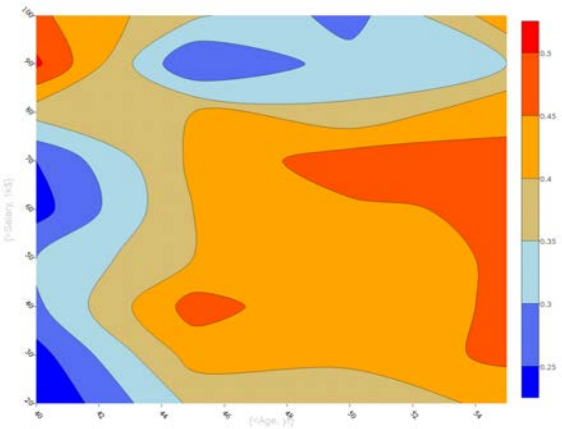
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Competence	0.46***	0.31***	0.40***					
	(0.10)	(0.09)	(0.08)					
Cooperation				0.55***	0.42***	0.46***		
				(0.12)	(0.07)	(0.07)		
PCA - Cooperation							0.51***	0.52***
							(0.10)	(0.10)
Centered R-squared	.11	.16	.13	.12	.17	.16	.14	.14
Skill	No	No	No	No	Yes	No	No	No
Incentive	No	No	No	No	No	Yes	No	No
PCA Sk./In.	No	No	No	No	No	No	No	Yes
Observations	131686	131686	131686	131686	131686	131686	131686	131686
Weak Id. F-Test	57.67	36.45	40.11	31.62	50.14	41.41	34.37	34.6
Underid. F-Test	54.96	33.35	70.34	56.85	66.85	61.7	56.5	57.42
Overid. F-Test	0	0	.95	1.47	6.59	6.09	.79	.75

Panel B: Time Performance

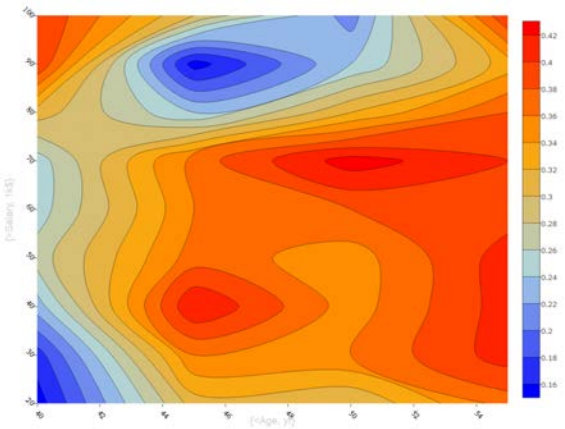
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Competence	0.40***	0.32**	0.37***					
	(0.11)	(0.08)	(0.08)					
Cooperation				0.49***	0.35***	0.38***		
				(0.11)	(0.09)	(0.11)		
PCA - Cooperation							0.46***	0.46***
							(0.10)	(0.10)
Centered R-squared	.08	.1	.09	.08	.12	.12	.09	.09
Q21	No	No	No	No	Yes	No	No	No
Q23	No	No	No	No	No	Yes	No	No
PCA Sk./In.	No	No	No	No	No	No	No	Yes
Observations	131686	131686	131686	131686	131686	131686	131686	131686
Weak Id. F-Test	57.67	36.45	40.11	31.62	50.14	41.41	34.37	34.6
Underid. F-Test	54.96	33.35	70.34	56.85	66.85	61.7	56.5	57.42
Overid. F-Test	0	0	.28	2.38	8.38	7.23	1.43	1.33

Notes: Results from Table 11 are replicated with standard errors - in parentheses - clustered by agency and service category and bootstrapped with 100 replications. Instruments are: *Relevant Deaths* and *Proximal Deaths*. Both contract outcomes and bureau characteristics are replaced by their standard scores. All models include controls for contract features (cost plus format and solicitation procedure), buyer characteristics (experience and yearly procurement budget), fixed effects for service category, agency, deciles for contract value and duration, year, and U.S. state of performance. * Significant at the 10 percent level; ** Significant at the 5 percent level; *** significant at the 1 percent level.

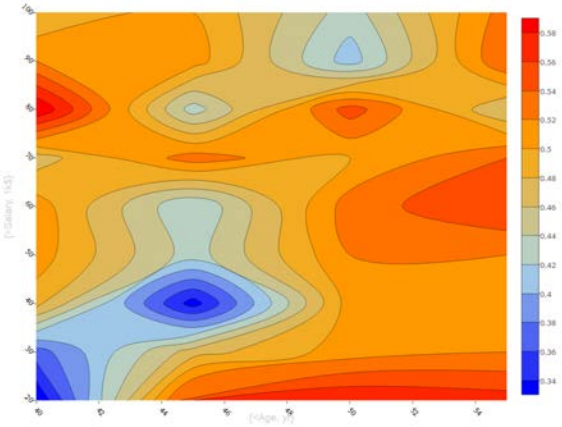
Figure A.2: Full Heterogeneity of IV Estimates for Competence Components



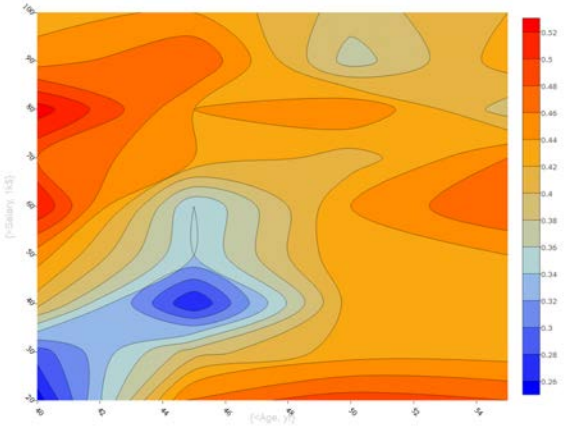
(a) Cost Performance: Competence



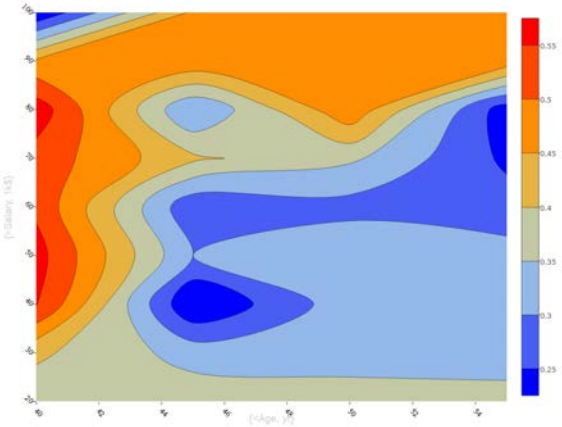
(b) Time Performance: Competence



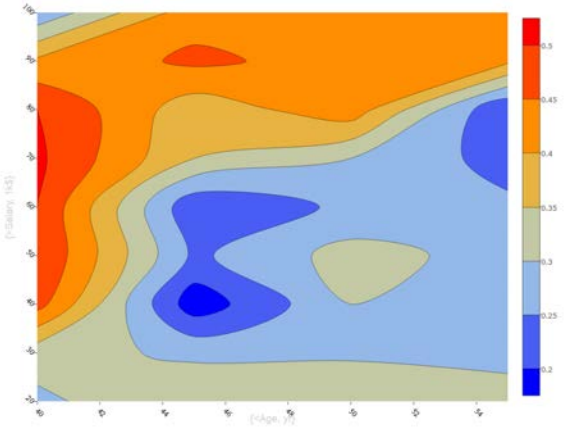
(c) Cost Performance: Cooperation



(d) Time Performance: Cooperation



(e) Cost Performance: PCA cooperation



(f) Time Performance: PCA cooperation

Notes: Heterogeneous effects of different employees. IV estimates obtained by varying the IV construction for all possible combinations of age and salary.