Barcelona School of Economics

WORKING PAPERS

Rules Versus Discretion in Public Procurement

BSE Working Paper 1232

October 2019 (Revised February 2022)

Rodrigo Carril

Rules Versus Discretion in Public Procurement

Rodrigo Carril*

First draft: October 2019
This draft: February 2022

Abstract

I study the trade-off between rules and discretion in the context of US federal procurement. Below an arbitrary threshold amount, contracts can be awarded using procedures that are subject to significantly fewer rules and less oversight. Leveraging a change in the threshold value, I document three key empirical findings. First, there is substantial bunching of contracts at the threshold. Second, the added scrutiny introduced by rules distorts the award amount of some contracts, while discouraging other purchases altogether. Third, contracts subject to more scrutiny perform worse ex post. I propose and estimate a stylized model of public procurement that is consistent with these findings. I find that, at current levels, the benefits from waste prevention are modest relative to the size of the compliance costs introduced by regulation. I find that the optimal threshold is substantially higher than the current one, and that a proposed increase in the threshold will leave the government better off. The model highlights the key role of incentive misalignment in bureaucracies, and shows quantitatively how increased discretion can be optimal as misalignment is reduced.

*Universitat Pompeu Fabra and Barcelona School of Economics. E-mail: rodrigo.carril@upf.edu. I thank Liran Einav, Mark Duggan, and Matt Gentzkow for their invaluable guidance and support. For helpful comments and suggestions, I also thank Nano Barahona, Michael Best, Nick Bloom, Tim Bresnahan, Juan Camilo Castillo, Raj Chetty, Ignacio Cuesta, Andres Gonzalez-Lira, Caroline Hoxby, Neale Mahoney, Petra Persson, Luigi Pistaferri, Francisco Sahli, and Heidi Williams. I gratefully acknowledge financial support from the Bradley Graduate and Postgraduate Fellowship through a grant to the Stanford Institute for Economic Policy Research (SIEPR), the Institute for Research in the Social Sciences (IRiSS) at Stanford University, and the Spanish *Agencia Estatal de Investigación* (AEI) through the grant PID2020-115044GB-I00//AEI/ 10.13039/501100011033 and through the Severo Ochoa Programme for Centres of Excellence in R&D (CEX2019-000915-S).

1. Introduction

To carry out their mission, governments often rely heavily on private firms. From fighter jets and cutting-edge R&D to office supplies and janitorial services, developed countries spend almost a third of their government budgets purchasing goods and services from the private sector.[1] The US is no exception: in fiscal year 2018, the federal government alone awarded $835 billion in procurement contracts .[2] Yet despite these large stakes, the design and functioning of public procurement systems remains a relatively under-studied area of government activity.

In this paper, I analyze how the strength of procurement regulation affects the incentives of public buyers and private sellers and its ultimate effect on the benefit that the government derives from public procurement. By regulation I mean the set of rules that constrain the allowable behavior of market participants, sometimes specifying the exact procedures that transacting parties need to follow, and establishing a variety of compliance steps and oversight mechanisms. Most public purchases are decentralized, with specialized government officials at each organizational unit in charge of awarding and managing their contracts. By enacting procurement regulation, lawmakers delineate the scope of action of these officers and of the private firms with whom they interact.

How much to regulate behavior and how much to leave at the discretion of procurement officers is a central question in public procurement.[3] A useful lens to analyze this trade-off is the more general problem of the delegation of authority within organizations (Aghion and Tirole, 1997). Following this framework, the principal (e.g., Congress) may prefer rules over discretion for high degrees of preference misalignment with the agent (contracting officers). Indeed, calls for tighter rules and increased oversight of public contracts are typically justified by referencing fears of wasteful spending and abuse.[4] On the other hand, the principal may be less inclined to regulate heavily if the agent has valuable local information and if discretion promotes the agent's initiative. Kelman (1990, 2005) has been a key advocate of this view in the case of procurement. He has argued that a proliferation of rules can lead to an intricate system that removes all incentives for officers to obtain good deals for the government, and instead makes them focus on costly compliance (i.e., *red tape*).[5]

I empirically analyze this trade-off in the context of the US federal procurement system, where two sets of acquisition procedures coexist, imposing different levels of regulatory strength. I leverage the fact that the procedures that public agencies use to select sources and award contracts

[1] Government procurement as a share of total government expenditures is 30.6% on average across OECD countries. This is an average of 13.2% as a share of GDP. Source: *Government at a Glance 2017, OECD*. Data available at: `http://www.oecd.org/gov/government-at-a-glance-2017-database.htm`.

[2] `http://www.usaspending.gov`.

[3] Gutman (2014) describes it as "the most difficult and contentious debate in public procurement."

[4] For example, the OECD describes procurement as "the government activity most vulnerable to waste, fraud and corruption." (`https://www.oecd.org/governance/public-procurement.htm`).

[5] According to Kelman (1990, 2005), procurement regulation, as typically envisioned and implemented, "is more often the source of the problem than the solution to it." In fact, "the fear of allowing public officials to use good sense and good judgment in procurement works against both the selection of the best contractor and the quality of performance of those selected." This is because rules mostly focus on "control of abuse —particularly favoritism, corruption, and cheating— by government officials or contractors, rather than on (...) supporting agency missions. Frequently, practices justified only on the ground that they reduced abuse were applied at the cost of sacrificing good decisions under normal circumstances. (...) Thus the rules misdirected the behavior of the many to stop the abuses of the few."

change discontinuously with the size of the award. Below an arbitrary threshold, buyers use so-called simplified acquisition procedures, which streamline the purchase process and give broad discretion to procurement officers. Above the threshold, procurement procedures are significantly more complex and lengthy, with the full set of regulatory statutes put in place to prevent waste and fraud coming into effect. Therefore, this hybrid regulatory regime provides a unique opportunity to evaluate the relative merits of rules and discretion in public procurement.

To assess the implications of this regulation, I organize my empirical analysis around two key questions. I first ask whether and to what extent market participants respond to this rule, and I analyze the implications of these responses for aggregate transaction patterns. I then explore how the level of regulatory strength affects the quality of procurement spending. My analysis relies on administrative records of contract awards by all US federal agencies between 2006 and 2016, and it leverages a policy change in 2011 that raised the simplified acquisition threshold from $100,000 to $150,000.

Conceptually, the policy makes it discontinuously more costly for both parties to transact for an amount above the statutory threshold. Therefore, we can interpret this policy as a regulatory *notch* in the award amount space (Slemrod, 2013; Kleven and Waseem, 2013). Relative to a counterfactual world in which no such added cost is imposed, two possible responses are expected. First, contracts above the threshold could reduce their award amount to qualify for the simplified procedure (an *intensive margin* effect). Second, it is possible that some purchases above the threshold are discarded altogether in response to the added regulatory burden (an *extensive margin* effect).

I document that the simplified acquisition rule indeed generates strong transaction responses. There is a substantial spike in contracting activity right below the simplified acquisition threshold and a discontinuous drop just above it. Several pieces of evidence indicate that this bunching reflects real distortions in purchasing decisions and is not artificially created by splitting large contracts into multiple smaller transactions.

To separately quantify intensive and extensive margin responses, I use a bunching design combined with a difference-in-differences strategy. Leveraging the change in the threshold value, I propose a nonparametric estimator of the distribution of contracts that we would have observed post 2011, had the threshold been kept constant at $100,000 and had not been raised to $150,000. Comparing this counterfactual to the actual observed distribution, I conclude that along with intensive margin effects, the policy generates strong extensive margin responses. I estimate that raising the simplified acquisition threshold generated more than 4,000 additional contracts per year (equivalent to 7.5% of contracts between $50,000 and $300,000), which translated into $546 million in additional annual contract spending (9.1% of spending over the same range).

I then explore how rules and discretion affect the quality of procurement contracts. Recall that this relationship is ex-ante ambiguous. Rules may reduce the scope of wasteful spending, particularly by misaligned agents. Yet if misalignment is not a first-order concern, preventing officers from exercising discretion might lead to worse contracts ex-post (Kelman, 1990, 2005). To test these competing hypotheses, I construct a series of contract quality proxies and exploit variation across award size and time periods, leveraging the simplified acquisition regulation and its change in 2011. The quality proxies include measures of costly post-award renegotiation and adaptation (contract modifications, cost overruns, and delays), contract termination, and the likelihood of an award at the very end of the fiscal year.[6] These measures have the advantage of being observed for the universe of contracts in the data, but have the disadvantage of focusing on the quality of the contract execution, as opposed to the quality of the goods and services obtained. To alleviate potential concerns with this, I show that my execution-based quality proxies correlate positively with actual measures of service quality in a sample of IT contracts for which these are systematically recorded. Then, using the validated quality proxies for the full sample, I find that tightly regulated awards result in worse ex-post performance, which is in line with the Kelman view of procurement regulation. I argue that sorting and selection are unlikely to account for the bulk of these effects.

Motivated by these findings, I propose a stylized empirical model of public procurement. The model formalizes how regulation affects contracting decisions and the observed quality of spending. There are two types of bureaucrats: officers who are fully aligned with the agency's mission; and misaligned ones, who care only about their private benefit. Purchases are the result of bilateral bargaining between bureaucrats and private firms. While aligned bureaucrats seek to obtain the highest possible share of the surplus for the government, misaligned agents —because of either corruption or simply inefficiency— overpay for goods and services. The introduction of tight rules effectively eliminates misaligned agents' wasteful spending, but it imposes red tape costs on transacting parties. Additionally, by removing the discretion from the contracting officer, regulation also affects the distribution of post-award contract performance. This leads to an improvement in quality for misaligned agents but a reduction in quality when an aligned agent is in charge.

I characterize the equilibrium for a *notched* regulation: contract awards above a certain threshold face high regulatory scrutiny, whereas contracts below the threshold are awarded with full discretion. Above the threshold, compliance costs effectively reduce the expected surplus from the transaction. In some transactions, it is possible for both parties to increase their payoffs by bunching at the regulatory threshold. In some other cases, the surplus might not be enough to justify the red tape costs, so if bunching is not feasible, some transactions are forgone because they are subject to regulation. The first key primitive of the model is the distribution of government valuations and private costs, which in combination define the transaction surpluses. The second object of interest is the distribution of unobserved compliance costs that the regulation imposes above the threshold, and which determines the magnitude of the behavioral responses. The third key primitive is the share of misaligned agents across federal agencies. Finally, other economic parameters describe the relative bargaining power of the transacting parties and the presence of adjustment frictions. I parameterize and estimate the model using simulated method of moments to match the descriptive patterns described above.

Estimated with data that predates the 2011 increase in the statutory threshold, the model predicts reasonably well the aggregate effects of this policy change. Compliance costs are quantitatively relevant, with a mean of \$12.8 thousand per contract, rationalizing the strong bunching in the data. On the other hand, I find that over 82% of transactions are led by aligned bureaucrats. This estimate reflects the quality drop when discretion is removed from contracting

[6] This latter measure has been found to be associated with wasteful spending by Liebman and Mahoney (2017).

officers at the simplified threshold: since most bureaucrats act in alignment with the government's goals, the introduction of rules leads to an aggregate drop in contract performance.

I use the estimated model to assess the implications of different statutory threshold choices. I consider the optimality of these alternative policies from the point of view of the government. This is due to the fact that the analysis only allows me to recover the valuations of public agencies for contracted goods and services, as opposed to citizens' valuations of such products, which would be a key input for a full welfare analysis. One choice that is of particular interest is a value of $250,000, which has been approved by Congress but has yet to be implemented.[7] Given the size of compliance costs and the relatively high degree of agent alignment, I estimate that this change —which reduces the stringency of regulation— will leave the government better off. This is so despite estimated spending increases in moderately-sized contracts ($50,000 to $300,000) of 8%. In fact, I estimate that the optimal regulation should set the simplified acquisition threshold at $345,000.

While interest in the theoretical aspects of procurement and regulation is not new (e.g., Laffont and Tirole, 1993), until recently, empirical evidence on the subject remained scarce. This paper relates to a rapidly growing number of studies that examine the consequences for procurement systems of particular design choices and institutional features.[8] In particular, a number of recent and concurrent studies have analyzed the role of buyer discretion in public procurement.[9] My results are in line with evidence from Pertold and Palguta (2017) in the Czech Republic and Szucs (2021) in Hungary, showing that procurement rules that change at arbitrary thresholds can induce transaction responses that result in "bunching" of contracting activity. It is also consistent with concurrent work by Bandiera, Best, Khan, and Prat (2021) in Pakistan and Decarolis, Fisman, Pinotti, and Vannutelli (2021) in Italy showing that discretion can lead to improved contract performance, even when it increases the risk of corruption. My work advances this literature by providing an assessment of the rules versus discretion trade-off that emphasizes and rationalizes both transaction responses and effects on spending quality. In the spirit of the argument proposed by Kelman (1990), I provide novel quantitative estimates of substantial transaction costs associated with procurement regulation. Furthermore, my proposed model allows me to draw policy implications both within and beyond the scope of this particular setting. Consistent with Bosio, Djankov, Glaeser, and Shleifer (2022), my framework highlights that the optimal degree of discretion is determined by the underlying level of preference misalignment among government bureaucrats, a parameter that I can quantitatively estimate. My conceptual and empirical analyses suggest that it is optimal to

[7]The National Defense Authorization Act of 2018 increased the simplified acquisition threshold from $150,000 to $250,000 for all agencies. As of October 2019, however, the Federal Acquisition Regulation has not yet been updated to reflect this change.

[8]Examples include the effects of centralized purchase agreements (Bandiera, Prat, and Valletti, 2009), information and disclosure requirements (Coviello and Mariniello, 2014; Carril, Gonzalez, and Walker, 2020), the use of electronic platforms (Lewis-Faupel, Neggers, Olken, and Pande, 2016; Carril, Gonzalez, and Walker, 2020), the use of different auction formats (Decarolis, 2014, 2018), expiring annual budgets (Liebman and Mahoney, 2017), bureaucrat's characteristics (Coviello and Gagliarducci, 2017; Decarolis, Giuffrida, Iossa, Mollisi, and Spagnolo, 2018; Best, Hjort, and Szakonyi, 2018), workload (Warren, 2014), external audits (Gerardino, Litschig, and Pomeranz, 2017), and industry consolidation (Carril and Duggan, 2020).

[9]See Coviello, Guglielmo, and Spagnolo (2018); Szucs (2021); Pertold and Palguta (2017); Decarolis, Fisman, Pinotti, and Vannutelli (2021); Kang and Miller (2022); Bosio, Djankov, Glaeser, and Shleifer (2022); Colonnelli and Prem (2022); Bandiera, Best, Khan, and Prat (2021).

increase bureaucratic discretion in institutional settings that feature a low degree of misalignment, which rationalizes the cross-country patterns presented by these authors.[10] It also helps reconcile the results from different country-specific studies, where discretion is sometimes found to improve (e.g., Bandiera et al., 2021; Decarolis et al., 2021) or worsen (e.g., Szucs, 2021) outcomes.

This paper also belongs to a growing literature that uses bunching designs to estimate behavioral responses by economic agents. Methods originally developed in the context of income taxation (Saez, 2010; Chetty, Friedman, Olsen, and Pistaferri, 2011; Kleven and Waseem, 2013) have been extended and applied to a variety of different settings characterized by nonlinear incentives.[11] The nonparametric estimator of spending effects that I propose in this paper extends the standard approach used in the early papers to incorporate extensive margin responses. In this respect, it adds to other approaches that seek to disentangle intensive and extensive margin responses to nonlinear schedules (Kopczuk and Munroe, 2015; Best and Kleven, 2018; Persson, 2019; Gelber, Jones, Sacks, and Song, 2018; Marx, 2018).

More broadly, this paper contributes to the literature that studies how the incentives of public officers inform policy design and affect public service delivery.[12] The results of this paper are consistent with a series of studies that highlight that failing to consider the incentives of public employees can make policies and regulations yield unintended consequences (Olken, 2007; Yang, 2008; Bandiera, Prat, and Valletti, 2009; Gerardino, Litschig, and Pomeranz, 2017; Rasul and Rogger, 2018; Duflo, Greenstone, Pande, and Ryan, 2018).

Finally, the rules-versus-discretion debate in public procurement can be analyzed through the more general lens of the theory of allocation of authority within organizations. In that respect, this paper belongs to the large literature that has followed Aghion and Tirole (1997).[13]

The rest of the paper proceeds as follows. Section 2 provides a brief background on the US federal procurement system, its regulation, and the data. In Section 3, I present my main empirical findings, including descriptive evidence of behavioral responses to the simplified acquisition rule, the estimation and quantification of intensive and extensive margin responses, and the effect of regulation on spending quality. In Section 4 I describe the model setup and its estimation. Section 5 presents the model estimation results, counterfactuals, and normative analysis. Section 6 concludes.

2. Setting and Data

2.1. US Federal Procurement

Procurement is a large component of the US federal budget. In fiscal year 2018, federal contract awards totaled $835 billion, representing 20% of total federal outlays, and two-thirds

[10] It is also noteworthy to highlight two related recent papers that also study the simplified acquisition regulation in the US, albeit with a different focus. Calvo, Cui, and Serpa (2019) study the reduced-form effect of oversight on procurement outcomes, and Giuffrida and Rovigatti (2020) study the effects of performance bonding. When appropriately comparable, the results on contract quality presented here are qualitatively similar to the ones in these papers.

[11] See Kleven (2016) for a review.

[12] See Finan, Olken, and Pande (2017) for a survey of papers that use field experiments in developing countries.

[13] See Aghion, Bloom, and Van Reenen (2013) for a survey of theoretical and empirical work.

of discretionary spending.[14]

Federal contracts can be broadly categorized into two types: definitive contracts (DCs) and indefinite delivery vehicles (IDVs). DCs are stand-alone one-time agreements with a single vendor for the purchase of goods or services under specified terms and conditions. In contrast, IDVs are agreements with one or more vendors, and are characterized at the time of the award by uncertainty about the quantity of goods or services to be provided, the timing of delivery, or the scope of the agreement. IDVs notably include purchases through centralized and government-wide agreements, like the General Services Administration (GSA)'s schedules. DCs and IDVs each account for roughly half of the contract spending (GAO, 2018). To simplify the analysis, in this paper I will focus exclusively on DCs. This will allow me to assess how regulation shapes the acquisition process in the case of well-defined requirements from individual government units.

DCs are awarded at highly decentralized levels. Contracts are awarded by over 3,000 different contracting offices that are part of an executive or independent agency.[15] The workforce in charge of public contracting is made up of over 35,000 contracting officers (Warren, 2014) whose primary role is to plan, carry out, and follow-up on purchases made by their units.

The scope of action of contracting officers is defined and limited by the Federal Acquisition Regulation (FAR), a 2,010-page document[16] that lays out policy goals and guiding principles as well as a uniform set of detailed policies and procedures to guide the procurement process. My empirical analysis will leverage a specific section of the FAR —Part 13: Simplified Acquisition Procedures—as a convenient natural experiment to study rules versus discretion.

2.2. Simplified Acquisition Procedures

As part of an effort to simplify the federal procurement process, the Federal Acquisition Streamlining Act of 1994 stipulated that contracts worth up to a statutory threshold value shall be awarded using a set of streamlined procedures. The main purpose of these simplified acquisition procedures (SAPs) is to reduce the administrative burden associated with the awarding of federal contracts. The simplified acquisition threshold was originally set by the legislation at $100,000. Then, effective the first day of fiscal year 2011, it was raised to $150,000. More recently, the National Defense Authorization Act of 2017 increased the threshold to $250,000, although full implementation has not yet occurred.

Above the threshold, virtually all contracts are awarded using a procedure called *negotiated acquisition*. This requires the officer to set up an evaluation plan, stipulating precisely the factors that will determine how the source will be selected and how proposals will be technically scored. Firms need to follow all the requirements in these evaluation plans when preparing their proposals, which constitute legally binding documents that the firm cannot later disavow. The scoring of proposals —performed by a source selection *team*— typically leaves a subset of highly-rated offers, which are

[14] Discretionary spending excludes mandatory programs (such as Social Security and Medicare) and interest on debt.

[15] Executive agencies are headed by a Cabinet secretary, like the Department of Defense, the Department of State, or the Department of Health and Human Services. Independent agencies, which are not part of the Cabinet, include the Central Intelligence Agency, the Environmental Protection Agency, and the Federal Trade Commission.

[16] https://www.acquisition.gov/sites/default/files/current/far/pdf/FAR.pdf

said to be *within the competitive range*. The officer then completes a next stage of negotiation with these offerers, before deciding which source to select and the precise terms of the award. Each of these steps is associated with significant paperwork requirements, vigorous oversight, and the possibility that prospective vendors will formally protest decisions. The process, therefore, leaves little scope for abuse by individual officers and firms.

SAPs waive most of these requirements, and instead give broad discretion to contracting officers, allowing them to operate on the basis of informal quotations from prospective vendors. Therefore, SAPs relieve contracting officers from a significant regulatory burden. This is also true for contractors, who now have to navigate a less cumbersome process. The key goal is to minimize red tape, allowing transacting parties to flexibly work on mutually beneficial agreements rather than focus on regulatory compliance.

While the simplified acquisition threshold applies generally, the FAR establishes an exception for the acquisition of *commercial items*. These are defined as "customarily used by the general public or by non-governmental entities for purposes other than governmental purposes" (FAR Part 2.101). Commercial items have a much higher threshold (currently $7 million) to qualify for simplified procedures. This means that in practice the simplified acquisition threshold will only be relevant for a subset of the observed contracts. Unfortunately, I will not be able to distinguish in the data commercial acquisitions at the individual contract level. However, it will be important to keep this exception in mind when interpreting the findings, and I will come back to this explicitly when I discuss the results.

2.3. Data

The main source of data is the Federal Procurement Data System - Next Generation (FPDS-NG), which tracks the universe of federal awards that exceed $5,000 (otherwise called micro purchases). The data are publicly available and can be downloaded from www.usaspending.gov. An observation in this dataset is a *contract action*, representing either an initial award or a follow-on action, such as a modification, termination, renewal, or exercise of options. For each observation, I observe detailed information, such as the dollar value of the funds obligated by the transaction; a four-digit code that describes the product or service; codes for the agency, sub-agency, and contracting office making the purchase; the identity of the private vendor; the type of contract pricing (typically, fixed-price or cost-plus); the extent of competition for the award; characteristics of the solicitation procedure; the number of offers received; and the applicability of a variety of laws and statutes. To make the contract the unit of observation, I collapse all actions by contract ID.

The analysis sample consists of all definitive contracts in excess of $5 thousand[17] awarded between fiscal years 2006 and 2016 for products and services other than Research and Development (R&D).[18] Table 1 presents summary statistics for the data. Since parts of the analysis will restrict

[17]This means non-IDV awards or micro purchases. Therefore, it includes definitive contracts that go through the standard procurement process (mostly negotiated acquisition) as well as those that are awarded using simplified acquisition procedures, referred to in the FAR as *purchase orders*.

[18]R&D awards, considered a special category of contracts, are subject to exceptional rules (see FAR Part 35). Additionally, maximum grant value has coincided at times with the simplified acquisition threshold. To avoid

attention to a window around the simplified acquisition threshold —e.g., between \$50,000 and \$300,000—the table presents statistics for all non-R&D contracts and for these restricted sample. We see that among federal agencies, the Department of Defense receives the most awards (54% overall and 61% in the restricted sample), followed by the Department of Veteran Affairs (12% and 9%, respectively), the State Department (6% and 4%, respectively), the Department of Justice (5% and 3%, respectively), and the Department of Agriculture (4% and 4%, respectively). Roughly two-thirds of the awards are competitively awarded, and virtually all contracts are fixed-price (as opposed to cost-plus). 80% of the contracts around the threshold use SAPs, and the average contract receives 3 offers (with the median contract receiving a single offer).

The bulk of contracting activity consists of relatively small purchases. Table 2 shows that two-thirds of the contracts are awards below \$25,000, 80% are below \$50,000, and less than 1% exceed \$5 million. Of course, the pattern is very different if we consider the expenditures by each size group, with contracts above \$5 million accounting for more than 80% of the total spending. The sample of "mid-size contracts," between \$50,000 and \$300,000, accounts for 15% of the awards and 4% of contract spending.

The sample consists of awards for a vast array of goods and services (excluding R&D). Each award is classified into one of 2,400 possible standardized 4-digit alphanumeric codes. These can be aggregated into 101 broader product *categories*: 77 goods, and 24 services. Table 3 shows the most common products. The upper panel presents the top 10 product categories, while the lower panel shows the top 10 products at the more granular 4-digit classification level. The most common product categories are professional services, electronic equipment, and medical supplies.

3. Empirical Evidence

3.1. Descriptive Evidence of Behavioral Responses to Regulation

"Bunching" Evidence

To shed light on the existence of transaction responses to the regulation, I start by examining contracting activity around the simplified acquisition threshold. Figure 1 presents visual evidence inconsistent with a null hypothesis in which regulation does not affect contracting behavior. Before 2011, when the statutory threshold equaled \$100,000, there was a pronounced spike in the distribution of contracts right up to this level and a discontinuous drop to the right of it (Figure 1(a)). Something similar is observed after 2011, when the spike and discontinuous drop in the distribution occurred at the award amount of \$150,000, the threshold value in force during those years (Figure 1(b)).[19] To get a better sense of the dynamics of adjustment to the new regulation,

confounding this with responses to the simplified acquisition regulation, I drop all contracts in the R&D category.

[19] A noteworthy feature of these distributions is the presence of round number effects expressed as spikes in the frequency of contracts in all bins that contain multiples of \$5,000 and \$10,000. This means that even in the absence of any regulation, we might have expected some spike at \$100,000, like the one we observe in Panel (b). In some of the estimation exercises below, I correct for these round number effects using the approach proposed by Kleven and Waseem

Appendix Figure A1 shows these distributions for every individual year between 2006 and 2016. The bunching follows the location of the threshold, albeit with some inertia. In particular, fiscal year 2011 appears as a transition year, with the distribution exhibiting large spikes both at the $100,000 and $150,000 levels. For this reason, in all of the analysis that follows I drop 2011, distinguishing between "pre" (2006-2010) and "post" (2012-2016) periods, as in Figure 1.[20]

The observed patterns are consistent with a preference by officers and firms for simplified contracting vis-à-vis the regular procurement process. In Section 4 I formalize this intuition by assuming that, relative to simplified acquisition, contracting for an amount that exceeds the threshold imposes a fixed cost on transacting parties. This creates a "notch" in the award amount space, whereby agent's payoffs discontinuously drop when they exceed the regulatory threshold. The responses documented here are qualitatively similar to those found in other contexts that feature notched tax schemes and regulations (see Kleven, 2016).

__Heterogeneity by Products and Agencies__

The patterns observed in the aggregate sample of contracts mask significant heterogeneity across both product categories and contracting agencies. An important source of this heterogeneity is the regulatory exception for commercially available products. Recall that the simplified acquisition threshold for commercially available products is equal to $7 million, and therefore we should not expect to see bunching around the $100,000 threshold for these types of award. Unfortunately, I do not observe at the individual contract level whether a given award was eligible for the simplified procedures as a commercial product. Yet, ex-ante, we expect that in certain product categories this exception will be claimed in a higher share of awards.

Appendix Figure A2 shows an example of this heterogeneity by product category, plotting the aggregate distribution, the distribution of electronic equipment awards, and the distribution of contracts in the "subsistence" (food) product category. Electronic equipment contracts, which are likely to be subject to a lot of customization for government's purposes, exhibit significantly starker bunching than the aggregate, while awards for food have essentially a smooth distribution around the threshold.

Because different agencies purchase different product mixes, this heterogeneity will be reflected across agencies as well. On top of this, it is possible that office-specific factors (e.g., experience, training, volume of purchases) also generate additional heterogeneity. Appendix Figure A3 illustrates this point by showing that the distribution of awards made by the Defense Logistics Agency has a much larger spike and subsequent drop than the one by the State Department.

(2013), who documented a similar phenomenon in the context of income taxation.

[20]The $150,000 threshold was first published in the Federal Acquisition Regulation (FAR) on the first day of fiscal year 2011. One possible explanation for this slow adjustment is that, because the observations reflect the date on which the contract was *awarded*, some of the contracts in early FY2011 may had had their terms and conditions defined before the new threshold was in place. Of course, this slow adjustment could also be rationalized simply by inattention or lack of perfect information by transacting parties.

Does Bunching Reflect Contract Splitting?

While the evidence from Figure 1 is inconsistent with the hypothesis that simplified regulation plays no role in procurement decisions, it is not clear whether these responses reflect any real effects. The fact that a contract can —in principle— be divided into pieces that add up to the original purchase suggests that there is a relatively straightforward way to bypass the regulation. For example, a \$120,000 contract would not have qualified for simplified procedures pre-2011, but two awards for the same product —say, of \$100,000 and \$20,000 value, respectively— would have. If the totality of observed responses reflects this type of behavior, then changing the policy would have no effect on either the total amount of goods and services purchased or on total procurement spending, even though we would see changes in the observed distribution of awards.

Several considerations suggest that this is unlikely to be a pervasive practice. First, this type of behavior is explicitly forbidden by the Federal Acquisition Regulation.[21] This prohibition is featured explicitly in most agencies' manuals and guidance documents,[22] and it is not uncommon for audit processes to be triggered by suspicions of split purchases.[23]

Second, these bunching patterns remain essentially unchanged when I restrict the analysis to offices that have had a single transaction with a given firm during a fiscal year. If there were coordinated efforts between contracting offices and vendors to split a large contract into two or more smaller awards, then the bunching should be driven by office-firm pairs with multiple transactions within the budget cycle. Appendix Figure A4 shows that this possibility is not supported by the data: the award distribution by single-transaction office-firm pairs is almost identical to that of the full sample. Similarly, if instead the splitting is driven by offices dividing purchases of the same product across different suppliers, we should see that the bunching is explained by offices with multiple purchases for the same detailed product code within the fiscal year. Yet as Appendix Figure A4 shows, the distribution has a very similar shape if we restrict the sample to single-transaction office-product pairs.

Third, if the bunching of awards is entirely explained by split purchases, then, all else being equal, the number of transactions should have *decreased* following the 2011 change in the simplified acquisition threshold. The logic is that, with a higher threshold, some of the splitting would no longer be necessary, leading to fewer overall transactions. I test this implication with a triple-difference design (DDD), where I exploit the heterogeneity across contracting offices that I documented in the previous subsection, before and after the policy change, and across different award sizes. The starting point is that under the null hypothesis of pure splitting, offices that

[21] FAR 13.003(c)(2) states that officers shall not "break down requirements aggregating more than the simplified acquisition threshold (...) into several purchases that are less than the applicable threshold merely to permit use of simplified acquisition procedures."

[22] See, e.g., the Department of Defense acquisition site's frequently asked questions (https://www.acq.osd.mil/dpap/pdi/pc/faq.html#q14, or the National Oceanic and Atmospheric Administration's acquisition training documents on the subject (www.ago.noaa.gov/acquisition/docs/ago_standdown/splitting_of_requirements.ppt).

[23] See, e.g., audits by the Department of Housing and Urban Development (https://www.hudoig.gov/sites/default/files/documents/IED-11-003R.pdf), or by the Department of Veteran Affairs (https://www.va.gov/oig/pubs/VAOIG-15-05519-377.pdf).

have more bunching in the pre-period would have reduced their level of transactions relative to low-bunching offices. Yet because differential budget changes can affect offices' transaction volumes independently of the regulation change, I look at these DD estimates across the award distribution and relative to the change that we see for awards that are not plausibly affected by the regulation —namely, awards above \$500,000. Appendix C describes this exercise formally and in detail, and the key results are presented in Appendix Figure A6. The plotted coefficients should be interpreted as the change in the number of contracts for offices with high pre-period bunching relative to those with low pre-period bunching for awards of different sizes. Panel (a) shows that the number of transactions dropped for contracts just below the initial threshold and increased to the right of it. This is, of course, consistent with the evidence from Figure 1, where we see the mass moving from the old to the new simplified threshold. Importantly, in Panel (b) we see that the net effect is an *increase* in the total number of transactions around the threshold. Furthermore, the coefficients outside the (\$85,000,\$200,000] window are very close to zero, suggesting that there is no compensation for this increase elsewhere in the distribution, as we would expect if split purchases were driving the bunching results.

All these results imply that the simplified acquisition regulation generates real transaction responses and that these were likely confined to a window around the threshold. In the next section, I seek to quantify the magnitude of these responses as well as their aggregate implications for contract spending.

3.2. Quantifying Behavioral Responses to Regulation

Conceptually, if non-simplified acquisition is seen as costly by government agencies and firms, we could expect responses along both the intensive and extensive margin.[24] Relative to a world in which no such cost exists, an intensive margin response would entail reducing the award amount of a contract from the right of the threshold, just enough to qualify for simplified contracting. This would create an "excess mass" of contracts to the left of the threshold and a corresponding "missing mass" to the right. An extensive margin response would entail contracts to the right of the threshold becoming inviable under the regulation, so that the original transaction is forgone altogether. This creates an *additional* missing mass to the right of the threshold.

Figure 2(a) depicts how these two effects affect the distribution of contracts. The excess mass A equals the number of contracts that see their award amount modified due to the regulation, while the difference between the missing mass B and the excess mass A equals the number of contracts that do not occur as a consequence of the regulation.

Now, starting from a situation in which the regulation is in place, we can ask what would be the effect of increasing the value of the simplified acquisition threshold. Note that this implies *relaxing* the regulation because fewer contracts will be affected by the regulatory cost. The same type of intensive and extensive margin effects would follow, this time in opposite directions. Some contracts would *increase* their award amount because the policy would become less stringent, while some additional contracts would now become profitable, increasing the overall level of purchases.

[24]This argument is formalized in Section 4.

The bunching of contracts would now occur at the new threshold level. Figure 2(b) shows these effects. Area C represents the volume of contracts responding through the intensive margin, while area $(D - E - C)$ accounts for the extensive margin response.

In a world with intensive and extensive margin responses (i.e., $D - E - C > 0$), raising the threshold would increase both the volume of contracts and total contract spending due to both higher average amounts and higher numbers of awards. In the absence of extensive margin responses (i.e., $D - E - C = 0$), we would see an increase in total spending due to higher average transacted amounts and no change in the number of contracts.

Leveraging the 2011 Threshold Change as Variation

Motivated by this framework, I analyze the 2011 change in federal regulation that raised the simplified acquisition threshold from \$100,000 to \$150,000. My goal is to create an empirical analogue of Figure 2(b), which will allow me to assess the quantitative relevance of intensive and extensive margin responses to regulation. The main challenge to this exercise is that a direct comparison of the pre-2011 and post-2011 award distributions is misleading because contract levels may vary over time for reasons unrelated to this regulation.[25] For example, if federal budgets change over the course of years, this will shift the distribution of contracts, even in the absence of changes to regulation.[26] To accurately quantify transaction responses, the relevant comparison is between the post-2011 distribution and the *counterfactual* distribution that we would have observed in this same period, had the threshold been kept fixed at \$100,000.

To estimate this counterfactual, I propose a non-parametric estimator that extends standard bunching methods to account for time variation that affects the absolute levels of the distribution. Like standard bunching designs (Saez, 2010; Chetty et al., 2011; Kleven and Waseem, 2013), I use parts of the distribution that are far from the threshold to obtain information about the counterfactual close to it. But because of the confounding variation over time, I apply this interpolation logic not to frequency levels directly but to frequency *changes* relative to the pre period. If transaction responses are circumscribed to a window around the initial and final thresholds, then changes to the award distribution far from these thresholds should reflect only the influence of variables not related to regulation (e.g., budget changes). I can then use this information to estimate how the pre-2011 distribution would have evolved in the absence of a threshold change. The logic of this exercise is to extend the standard bunching design using a difference-in-differences strategy: bins from outside the excluded window can serve as a counterfactual for bins within this region, and we compare their trajectories before and after the policy change to account for time-varying factors that affect both parts of the distribution in similar ways.

More concretely, I first classify all awards between \$5,000 and \$1,000,000 into bins of \$1,000 length, where bin $b = 6$ includes all awards in $(\$5,000, \$6,000]$, and so forth. Let n_b^t the number

[25] Note that simply normalizing the frequencies to compare the distributions would not be appropriate in the presence of extensive margin responses because the regulation is affecting the total number of contracts, not just their location along the distribution.

[26] In fact, the regulation change coincides with the enactment of the Budget Control Act, which cut discretionary spending during the subsequent years.

of contracts in bin $b \in \{6, ..., 1000\}$ and period $t \in \{PRE, POST\}$. I then compute for each bin the relative frequency change between the pre and post periods, $RFC_b = (n_b^{POST}/n_b^{PRE}) - 1$. Defining R to be the excluded region (i.e., the range of bins that are affected by the regulation), I fit the relative frequency changes on a polynomial of degree p of the bin values and a set of dummies for each of the bins in R:

$$RFC_b = \sum_{x=0}^{p} \beta_x \cdot b^x + \sum_{j \in R} \gamma_j \cdot \mathbf{1}[b = j] + \nu_b \quad . \tag{1}$$

Estimated frequency changes are then obtained by ignoring the contribution of the excluded region dummies, i.e., $\widehat{RFC_b} = \sum_{x=0}^{p} \hat{\beta}_x \cdot b^x$. This effectively means interpolating the polynomial fit into the excluded region, using only the observations outside of this window. Counterfactual frequencies are then obtained by applying the estimated changes to the pre-period frequencies: $\hat{n}_b^{CF} = n_b^{PRE} \cdot (1 + \widehat{RFC_b})$. By comparing these estimated counterfactual counts with the actual post-2011 distribution, I obtain estimates of the intensive and extensive margin responses, as in Figure 2(b).

Specifically, I compute the amount of missing mass ($\widehat{m}$) to the left of the initial threshold as the cumulative difference in counts between the counterfactual and actual distribution, normalized by the average counterfactual frequency in the excluded region. The excess mass ($\widehat{x}$) to the right of the initial threshold is defined analogously. In particular,

$$\widehat{m} = \frac{\sum_{b=\underline{R}}^{100} (\hat{n}_b^{CF} - n_b^{POST})}{\sum_{b=\underline{R}}^{\overline{R}} \hat{n}_b^{CF} / (\overline{R} - \underline{R})} \qquad ; \qquad \widehat{x} = \frac{\sum_{b=100}^{\overline{R}} (n_b^{POST} - \hat{n}_b^{CF})}{\sum_{b=\underline{R}}^{\overline{R}} \hat{n}_b^{CF} / (\overline{R} - \underline{R})} \quad . \tag{2}$$

Of course, the estimate of missing mass ($\widehat{m}$) is directly linked to intensive margin responses, whereas extensive margin responses are related to *net* excess mass ($\widehat{x} - \widehat{m}$). Standard errors for these estimates are obtained via bootstrap.

Results

Figure 3 shows the implementation of the procedure graphically. In Figure 3(a), I plot frequency changes for each bin and fit a kernel function over the three key regions: below the original threshold, between the initial and final threshold, and above the final threshold. Bins closely below \$100,000 reduce their levels considerably, since in the pre period there was strong bunching at these levels, which largely disappears in the post period. In the middle area, all bins increase their frequency, and this effect is particularly strong as we approach \$150,000 from the left, where the new bunching emerged. Finally, all bins above \$150,000 decrease their frequency.

Figure 3(b) shows the fit of (1) once we have "dummied out" the excluded region, which in the baseline estimation is defined as between \$50,000 and \$300,000. These limits correspond to half the size of the initial threshold and twice the size of the final threshold.[27]

[27] This choice is obviously arbitrary, yet it is arguably a conservative one. Recall that the evidence from Appendix Figure A6 suggested that offices that had high- and low- pre period bunching did not respond differentially to the policy change below \$85,000 and above \$200,000. In Appendix D I show the sensitivity of my results to particular choices of this region.

Having estimated $\widehat{RFC}_b$ I can then compute the counterfactual distribution for the post-2011 period, in a world were the regulatory threshold was kept fixed at $100,000. Figure 4 presents this counterfactual, along with the actual distribution observed post-2011. We see that increasing the simplified acquisition threshold made the distribution shift to the right, as we expected. However, it is visually apparent that the amount of mass that was originally bunched at $100,000 is smaller than the additional mass that is generated in excess of the counterfactual, to the right of the initial threshold.

Indeed, the first column in Table 4 shows that the missing mass to the left of $100,000 is almost 10 times the average frequency within the excluded region, suggesting the presence of significant intensive margin responses. However, the excess mass estimate is roughly three times as large, suggesting that extensive margin responses account for two-thirds of the increased contracting that we see to the right of the initial threshold. These estimates imply that an additional 4,095 contracts per year are awarded as a result of the higher simplified acquisition threshold. These additional transactions and the higher award amounts of contracts that respond along the intensive margin generate an extra $526 million in annual contract spending (8.7% of counterfactual spending in contracts between $50,000 and $300,000).

Placebos

To assess the plausibility of the identifying assumptions, I conduct two placebo exercises that focus on periods when there was no change in the simplified acquisition threshold. Recall that the validity of my estimates relies on the assumption that, absent any changes to the policy threshold, the frequency distribution of contracts evolves smoothly. If this is the case, then interpolating frequency changes from outside the excluded region should accurately capture the evolution of the distribution around the threshold. By examining consecutive periods when the threshold actually remained fixed, I can test whether this assumption is plausible. Note that this is analogous to testing whether there are parallel pre-trends in the context of a difference-in-difference design.

In particular, I predict the award distribution in 2008-2009 using as a baseline the contract distribution during the previous two years (2006-2007). Because the policy should have a one-time effect, I can repeat this logic and apply it to consecutive periods *after* the threshold change had already materialized. I do this predicting the last two years of my sample (2015-2016), using the previous two years as baseline.

Figure 5 shows the results of these placebo exercises. Panel (a) presents the results for 2008-2009, while Panel (b) shows the estimation for 2015-2016. In both cases, the estimated counterfactual distribution tracks the actual distribution closely, bolstering the plausibility of the main counterfactual results of Figure 4. Consistent with this visual evidence, the second and third columns of Table 4 report that estimates of $\widehat{m}$ and $\widehat{x}$ are very close to zero.

3.3. Effects of Regulation on the Quality of Awarded Contracts

My estimated effects of regulation on the distribution of contract awards suggest that the simplified acquisition regulation significantly changes purchasing decisions by public agencies. In particular, the threshold regulation induces a reduction in the overall size of certain awards, and it discourages some purchases altogether. These results, however, say nothing about the quality of the contracts that get awarded.

Regulation may be an effective way of curtailing waste. We can think of non-simplified acquisition as a means of ensuring that government agencies exercise due diligence before they award federal contracts. The increased oversight and paperwork requirements may lead to a more effective screening of vendors. All of this would eventually translate into better post-award performance.

Yet regulation could adversely affect contract quality. As suggested by Kelman (1990, 2005), reducing discretion can hinder the ability of contracting officers to reward and punish vendors for aspects of quality that are not easily contractible. Increasing the complexity of the procurement process can also lead to a poorer selection of firms: contractors that become experts in navigating the system of rules, rather than those that provide the best value for the government, tend to be selected. Consequently, highly regulated procurement leads to lower quality contracts.

Whether the quality of procurement spending is better or worse under simplified acquisition is, therefore, an empirical question. In this section I explore this issue and find that the data are consistent with the Kelman view. Relative to simplified awards, contracts subject to tighter regulation fare worse in terms of various performance measures.

Measuring contract quality

An important challenge faced by the empirical procurement literature is the availability of spending quality measures. I deal with this difficulty by constructing a series of quality proxies based on the observed execution of the contract. In particular, I develop five measures of (poor) contract quality. While no measure is perfect, I find reassuring the robustness of results across different measures. Furthermore, below I present a validation of these measures using data from a sample of IT contracts where explicit quality ratings are observed.

The first three measures are: (i) the number of post-award modifications, (ii) cost overruns (in dollars), and (iii) time delays (in days). These are based on the important idea that a crucial component of procurement contracts' costs is ex-post renegotiation and adaptation.[28] Because the data specify for each contract the total sum of payments and completion date *expected* at the time of the award, I can construct measures of overruns and delays by comparing these expectations to the realized payments and duration. Following recent studies that use these types of measures as performance proxies, I consider only renegotiations that occur *within the scope* of the original

[28]See Crocker and Reynolds (1993); Bajari and Tadelis (2001); Spiller (2008); Gagnepain et al. (2013); Bajari et al. (2014). Ganuza (2007) provides an alternative rationale for observed renegotiation and adaptation, which arises from strategic underinvestment in initial project design by the procurer, in order to intensify competition.

agreement.[29] The fourth performance measure is the probability of contract termination, which typically occurs when the contractor fails to comply with some term in the original agreement. The fifth and final proxy is the probability that a contract is awarded during the last week of the fiscal year, building on prior evidence that finds these type of awards to be of lower quality (Liebman and Mahoney, 2017).

While I present results separately for each of these quality measures, I also combine them in a single *quality index*. I follow Kling et al. (2007) in constructing this summary index. Let Y_i^k be a quality measure k for contract i, with $k \in \{1, ..., 5\}$, and let $\bar{Y}^k$ and σ^k be their mean and standard deviation during the 2006-2010 period. The normalized quality measure k for contract i is $\hat{Y}_i^k = (Y_i^k - \bar{Y}^k)/\sigma^k$, and the quality index is simply the average across the normalized quality measures: $q_i = \frac{1}{5}\sum_{k=1}^{5} \hat{Y}_i^k$.

Validating quality measures with IT procurement data

Before presenting the results, it is important to address a potential concern with the proposed quality proxies. Four out of five of these measures are constructed only based on the observed execution of the contract. Therefore, they may not reflect —or, worse, may even correlate negatively with— the underlying quality of the purchased product or service. Suppose that contractors that put more effort into the quality of service delivery tend to have more delays, then a lower value of the index q_i might actually imply *higher* quality received by the government.

To explore this issue empirically, I follow Liebman and Mahoney (2017) and use data from the IT Dashboard. This is a federal platform that tracks the performance of the most important federal IT projects. Beyond measures of schedule and cost, which capture the same execution information contained in some of my quality proxies, the projects are rated in a discrete 1-to-5 scale by chief information officers (CIOs) at the responsible agency. CIOs evaluation is ultimately subjective, but it must be informed by a series of objective quality criteria. The specific metrics vary depending on the project, but they directly capture the quality of the delivered product or service.[30] Furthermore, CIOs have incentives to rate projects accurately (Liebman and Mahoney, 2017).

The IT Dashboard provides us with an opportunity to validate the quality proxies used in this paper. Appendix E presents this validation exercise in detail, while we briefly mention here the main results. Appendix Table B2 shows summary statistics for the sample of 390 projects in the IT dashboard that can be merged to 3,576 contracts in our analysis sample. We use this sample to study whether the quality proxies based on contract implementation used in the main analysis correlate with CIO evaluations for the associated IT project.[31]

Our results suggest that contract implementation quality as measured in FPDS is strongly

[29] E.g. Lewis and Bajari, 2011; Decarolis et al., 2018; Giuffrida and Rovigatti, 2020; Kang and Miller, 2022.

[30] Examples include "percent of the time that the system is available", "percent of servers reduced as a result of virtualization", "number of applicants using ePermits", "number of repeat customers using system", "reduction in number of field office data storage devices", "percent of customer issues that are addressed and fully resolved within 24-hours", etc.

[31] An alternative would be to perform all the analyses of this paper in the sample of IT projects. However, very low sample sizes make the nonparametric exercises presented infeasible.

positively correlated with CIOs evaluations. Appendix Figure A5 (a) shows that the unconditional mean of the quality index ($\bar{q}$) is -0.214 for contracts associated with projects rated 1 or 2, $\bar{q} = -0.015$ for projects with rating equal to 3, $\bar{q} = 0.015$ for projects with a rating of 4, and $\bar{q} = 0.038$ for projects with the highest rating. Panels (b) and (c) show that this positive relationship is robust to controlling for agency fixed-effects and weighting by the size of the project.

The main take-away is that, to the extent that we can measure variables reflecting the quality of the purchased good or service, these are positively correlated with the quality proxies constructed based on contract execution. We then proceed using the latter, since these have the advantage of being available for the universe of contracts analyzed.

Results

I start by examining raw measures of contract quality over the award distribution. Figure 6 Panel (a) and Panel (b) respectively show the average quality index for awards of different amounts during both the "pre" (2006-2010) and "post" (2012-2016) periods. The plots allow us to concentrate on two sources of variation: in the size of the award (since, for a given period, contracts to the left and right of the threshold differ in that the latter are generally subject to a higher regulatory burden); and over time (since, across periods, the location of the threshold changed from \$100,000 to \$150,000). These figures show that contracts subject to increased regulation present discontinuously worse performance, with a quality drop that occurs precisely at the simplified acquisition threshold.

To compare these patterns directly, I estimate the following specifications:

$$\begin{aligned} Y_i^{pre} &= \alpha_{t(i)}^{pre} + \alpha_{b(i)}^{pre} + \sum_{s \in S} \beta_s^{pre} \cdot \mathbf{1}[b(i) \in s] + \delta^{pre} \cdot X_i + \epsilon_i^{pre} \quad \text{for } t(i) \leq 2010 \\ Y_i^{post} &= \alpha_{t(i)}^{post} + \alpha_{b(i)}^{post} + \sum_{s \in S} \beta_s^{post} \cdot \mathbf{1}[b(i) \in s] + \delta^{post} \cdot X_i + \epsilon_i^{post} \quad \text{for } t(i) > 2011 \quad , \end{aligned} \tag{3}$$

where the unit of observation is a contract i, $t(i)$ is the fiscal year when the contract is signed, $b(i)$ is an award amount bin of \$1,000 width, X_i are controls, and S represents some partition of the award amount space.[32]

Figure 6 Panel (c) plots the set of estimated coefficients β_s^{pre} and β_s^{post}, along with 95% confidence intervals, for versions of (3) with the quality index as an outcome and no controls. Naturally, the coefficients replicate the discontinuous drop in quality at the simplified acquisition threshold for each period. Particularly informative is the gap in quality that emerges in the medium region of \$100,000 to \$150,000 awards; contracts in this range were subject to regulation in the pre period but are eligible for simplified procedures in the post period.

To explicitly look at these differences across periods, I pool observations across both periods and estimate:

$$Y_i = \alpha_{t(i)} + \alpha_{b(i)} + \sum_{s \in S} \beta_s \cdot \mathbf{1}[b(i) \in s] \times Post_{t(i)} + \delta \cdot X_i + \epsilon_i \quad , \tag{4}$$

where $Post_{t(i)}$ is a dummy equal to 1 for contracts awarded after 2011. Figure 6 Panel (d) shows the

[32] In all regression analyses in this section, standard errors are clustered by awarding agency of the contract.

estimated β_s coefficients and confidence intervals for the quality index measure and no controls. Again, the main result is that the quality of contracts increases for awards in the range where simplified acquisition is now allowed.

These patterns are also present, albeit with varying degrees of precision, when we consider separately each of the five performance proxies that constitute the quality index. Appendix Figure A7 replicates the plot in Figure 6 Panel (c) for each underlying measure.[33] Again, both the within-period and across-period variation are consistent with regulation associated with worse performance.

These results need to be interpreted carefully. Because of the evidence presented in the preceding sections, we should not interpret these effects as reflecting purely a "treatment effect" of regulation on contract quality. Rather, they are the combination of any such effect with the difference in quality that can arise from an endogenous selection of contracts. On the one hand, the intensive margin responses documented above imply that some awards will sort into the low-regulation regime regardless of the location of the threshold. On the other hand, the extensive margin responses discussed previously imply that the region between \$100,000 and \$150,000 in the post period will include awards that were forgone when regulation was tighter.

Nevertheless, three arguments suggest that selection is unlikely to be the main driver of these results. The first is the direction of the bias that selection would introduce. Presumably if there is any benefit from increased oversight and regulatory scrutiny it is to screen out ex-ante wasteful contracts. But if this is the case, then the awards that select into the \$100,000 to \$150,000 region in the post period should be negatively selected in terms of quality, relative to the pre period contracts. This would make the "true" gap attributable to the causal effect of regulation on quality even larger.

A second argument against selection playing a first-order role is given by the mean quality patterns to the left of \$100,000. The "pre" and "post" coefficients in this left region of Figure 6 Panel (c) almost fully overlap, even when the underlying distribution is changing substantially. If "bunched" contracts were selected in terms of quality, then we would see large changes in the post period quality index just below the \$100,000 threshold. However, the mean quality remains very similar to that of the pre-period, just like awards that we know are unlikely to be affected by selection— say, those below \$75,000, which is a part of the distribution for which there is no evidence of intensive or extensive margin responses (see Figure 4).

A third and final argument is that these results are essentially unchanged when we include controls for observable characteristics of the contracts. As discussed in Section 3, we see significant variation in the extent of responses, depending on the awarding agency and the product that is being purchased. To assess whether the pre-vs-post gaps in quality are still present once we account for these observables, I re-estimate (3) and (4), this time including a full set of detailed product codes (4-digit PSC codes) and awarding office fixed-effects. The results are presented in Appendix Figures A9 and A10, and are essentially equivalent to those in Appendix Figures A7 and A8.

In terms of magnitudes, Table 5 presents estimates from (4) with the full set of controls and a coarser partition of the award amount space: $(50, 75]$ (the base region), $(75, 100]$ (simplified

[33]Note that in the case of the individual measures, we see a positive "jump", because the variables correspond to measures of poor performance.

acquisition pre and post), $(100, 150]$ (regulation pre, simplified acquisition post), and $(150, 300]$ (tight regulation both pre and post).[34] Consistent with the raw data and less parametric specifications discuss above, there is at most very modest evidence of selection effects right below $100,000, yet there is a strong and significant *positive* effect on quality for contracts between $100,000 and $150,000, which in the post period became eligible for simplified acquisition. These effects are consistent across all quality measures, as well as reflected on the aggregated quality index.[35] In terms of magnitudes, the middle row in Table 5 indicates that contracts that become eligible for simplified procedures see on average: a reduction of 0.07 in the number of modifications (-7.5% relative to the mean), (ii) $2.8 thousand less cost overruns (-32%), (iii) a reduction of 5.7 days on average delays (-7%), (iv) a drop of 0.1 percentage points in their probability of termination (-7.5%), and (v) a 0.88 percentage points decline in the share of contracts awarded during the last week of the fiscal year (-10%). All this is reflected in an increase of roughly 0.03 units of standard deviation in the normalized quality index that aggregates these different performance proxies.

In sum, across a variety of measures, contracts subject to higher levels of regulation perform worse than lightly-regulated simplified contracts. This appears to be driven by the effect of regulation itself rather than by the selection of transactions into different acquisition procedures. These results are also consistent with work that finds that delays and overruns increase above the simplified acquisition threshold (Calvo et al., 2019; Giuffrida and Rovigatti, 2020).

4. A Stylized Model of Public Procurement

Motivated by the evidence presented in the previous section, I propose a simple model of public procurement that formalizes how these patterns can arise. I then estimate the model to examine the extent to which it can quantitatively replicate the empirical findings. This allows me to estimate the consequences of counterfactual policies. Furthermore, the added structure makes it possible to evaluate the regulation from a normative standpoint, clarifying the efficiency trade-off associated with the strength of procurement regulation.

4.1. Model Description

The basic model structure consists of bilateral bargaining between a government agency (buyer) and a private contractor (seller).[36] The agency needs to purchase a good or service, for which it has an *expected* valuation of v. There is a private firm that can produce such a product at cost c. If the bargaining process leads to an award p, then the agency obtains an expected payoff of $\tilde{U}(p) = v - p$ at the time of the award, and the firm obtains $\Pi(p) = p - c$. After the award,

[34] All numbers in thousand dollars.

[35] The negative coefficients imply a positive effect on quality for columns (2)-(6) because these specifications have a dependent variable that measures poor contract performance.

[36] In this respect, the model is closely related to Kopczuk and Munroe (2015), who study housing market responses to notched transaction taxes. As we will see below, regulatory costs will effectively act as a notched tax on procurement transactions, and, thus, many of the results here will be identical to their model. However, the equilibrium responses between the models will differ. They consider a pre-bargaining matching stage that I will abstract from, while the agency considerations that are key to the procurement problem are not present in their model of the housing market.

the firm implements the contract, and a shock ϵ is realized, which affects the actual value that the government obtains ex-post, $w(\epsilon)$.[37] This implementation shock captures contract incompleteness, leading to uncertainty about the post-award transaction costs, as emphasized by Bajari and Tadelis (2001). Once uncertainty is resolved, the government obtains an ex-post payoff of $U(p) = w(\epsilon) - p$.

The government agency is represented by a bureaucrat who has local information. In particular, the bureaucrat knows c, while the principal (e.g., Congress) does not. There are two types of bureaucrats: aligned (A) and misaligned (M) agents. Aligned bureaucrats act in complete correspondence with the goals of the agency —that is, their personal payoff (B_A) is identical to the government's payoff (U). On the other hand, misaligned agents' personal payoff differs from the government's ($B_M \neq U$). The principal can allow bureaucrats to exercise full discretion ($R = 0$), or it can force the application of a set of rules ($R = 1$). Finally, the distribution of implementation shocks depends on the identity of the buyer and on whether or not the purchase is subject to a strict regulatory regime: $\epsilon | X, R \sim f(\cdot|X, R)$ with $X \in \{A, M\}$ and $R \in \{0, 1\}$. This implies that expected valuations also depend on both the agent's type and regulation: $v = E_\epsilon[w(\epsilon)|X, R]$.

Bargaining with full discretion

In the full discretion case, the principal is unable to scrutinize the transaction. Bureaucrats are instructed to bargain with the firm, leveraging the government's bargaining power of ϕ. However, this cannot be verified by the principal because c is only observed by the agent.

Aligned agents follow the prescribed behavior. The expected valuation in this case is $v_A \equiv E_\epsilon[w(\epsilon)|A, R = 0]$, and so the bureaucrat and the firm bargain over the expected surplus $(v_A - c)$. The equilibrium award p_A is determined by maximizing the Nash bargaining objective:

$$p_A = \arg\max_p (v_A - p)^\phi (p - c)^{1-\phi} \tag{5}$$

which will leave the government (and the bureaucrat) with a share ϕ of the expected surplus. The remaining share will be profits for the firm.

Misaligned agents do not reach the Nash bargaining solution. Instead, they simply award p_M equal to the highest feasible amount and receive a payoff of $B_M = b_M(v_A)$ if they complete the transaction, or $B_M = 0$ otherwise. I assume that the principal cannot detect misaligned agents unless p_M is inconsistent with *any* award that an aligned agent would choose. This means that the highest possible award that can go undetected is $p_M = v_A$.[38]

I do not take a stance on the exact nature of this misalignment. One interpretation is that the agent is corrupt and, thus, that this inflated award maximizes the profits of the firm, which then retributes the bureaucrat by paying a kickback equal to b_M. Another interpretation is that the bureaucrat is simply lazy or inefficient. For example, suppose that bargaining requires effort, but

[37] I.e., v is the expectation of $w(\epsilon)$ at the time of the award. Importantly, because ϵ will depend on the agent's type X and the regulatory regime R (see below), v, too, will be a function of these two variables.

[38] If an aligned agent is in charge, the highest possible award would occur when $c = v_A$, which leads to $p_A = \phi v_A + (1 - \phi) v_A = v_A$. A higher award would be inconsistent with the aligned agent's behavior because it would yield a negative payoff to the government.

that the agent only receives some fixed benefit b_M for completing the transaction, without being the residual claimant of any payment reductions that result from bargaining. In this case, the agent exerts no effort and simply accepts any offer from the firm (as long as it does not raise suspicions from the principal). Regardless of the interpretation, b_M is allowed to depend on the expected valuation v_A to capture the fact that misaligned agents may have more to gain from larger value transactions —either because the size of the bribe increases or because bureaucrats are incentivized to complete larger value transactions.

Regulated procurement

Because of the existence of misaligned agents, the principal considers making procurement transactions subject to regulation ($R = 1$). This has three practical effects. First, it reveals the cost parameter c, so that all agents are forced to reach the Nash bargaining solution. Second, it imposes red tape costs on transacting parties. In particular, the transaction is subject to a total cost of κ, which is borne by the bureaucrat and the firm according to shares γ and $(1-\gamma)$, respectively. Third, the regulation affects the distribution of post-award contract performance: implementation shocks are now drawn from $f(\cdot|X, R = 1)$, instead of from $f(\cdot|X, R = 0)$, for $X \in \{A, M\}$. Furthermore, when regulation is introduced, the identity of the agent becomes irrelevant for the distribution of post-award shocks, i.e., $f(\cdot|A, R = 1) = f(\cdot|M, R = 1) = f(\cdot|R = 1)$.

Let the expected product valuation under regulation (for both aligned and misaligned agents) be $v_R \equiv E_\epsilon[w(\epsilon)|R = 1]$. Compliance costs κ effectively reduce the size of the expected surplus, with parties now bargaining over $(v_R - c - \kappa)$. The equilibrium award p_R maximizes the Nash bargaining objective is:

$$p_R = \arg\max_p (v_R - p - \gamma\kappa)^\phi (p - c - (1-\gamma)\kappa)^{1-\phi} \tag{6}$$

Aligned bureaucrats obtain an expected payoff identical to that of the government (i.e., a share ϕ of this reduced surplus). Misaligned bureaucrats get some reservation value from completing the transaction, $B_M = b_R$, or zero otherwise.

4.2. Equilibrium with a threshold regulation

Motivated by the simplified acquisition threshold, I focus on equilibrium behavior when the principal requires transactions to be subject to $R = 1$ if they exceed a certain award amount $\bar{P}$. To be precise, the timing of the game is as follows. First, nature draws a type for the bureaucrat $X \in \{A, M\}$ and a vector of expected valuations and cost (v_A, v_M, v_R, c), which are observed by the agent and the firm. Second, the bureaucrat decides whether to use simplified ($R = 0$) or regulated ($R = 1$) procedures. Third, bargaining takes place and a contract may be awarded (if $R = 1$ was chosen, this award may not exceed $\bar{P}$). Fourth, if a contract is awarded, the implementation shock $\epsilon \sim f(\epsilon|X, R)$ is realized. I now summarize the possible outcomes that can arise in equilibrium for each of the two types of agents. Derivation details and the exact conditions that determine each case are presented in Appendix F.

Let p_A^* be the equilibrium award made by an aligned bureaucrat facing a simplified acquisition threshold $\bar{P}$. Then p_A^* and the government's expected payoff $\tilde{U}(p_A^*)$ must correspond to one of the following cases:

Case A:	$p_A^* = \phi c + (1-\phi)v_A;$	$\tilde{U}(p_A^*) = \phi(v_A - c)$
Case B:	$p_A^* = \bar{P};$	$\tilde{U}(p_A^*) = (v_A - \bar{P})$
Case C:	$p_A^* = \phi c + (1-\phi)v_R + (\phi - \gamma)\kappa;$	$\tilde{U}(p_A^*) = \phi(v_R - c - \kappa)$
Case D:	$p_A^* = \emptyset;$	$\tilde{U}(p_A^*) = 0$

On the other hand, let p_M^* be the equilibrium award made by a misaligned bureaucrat who faces a simplified acquisition threshold $\bar{P}$. Then p_M^*, the bureaucrat's personal payoff (B_M), and the government's expected payoff $\tilde{U}(p_A^*)$ must correspond to one of the following cases:

Case A':	$p_M^* = v_A;$	$\tilde{U}(p_M^*) = (v_M - v_A);$	$B_M = b_M(v_A)$
Case B':	$p_M^* = \bar{P};$	$\tilde{U}(p_M^*) = (v_M - \bar{P});$	$B_M = b_M(\bar{P})$
Case C':	$p_M^* = \phi c + (1-\phi)v_R + (\phi - \gamma)\kappa;$	$\tilde{U}(p_A^*) = \phi(v_R - c - \kappa);$	$B_M = b_R$
Case D':	$p_M^* = \emptyset;$	$\tilde{U}(p_M^*) = 0;$	$B_M = 0$

Despite having very different incentives, both aligned and misaligned agents behave in qualitatively similar ways. Relative to the full discretion case, the introduction of this threshold regulation can either have no effects (Cases A and A'), generate intensive margin responses (Cases B, C, B' and C') —some of which will involve bunching (Cases C and C')—or generate extensive margin responses that imply forgoing the transaction (Cases D and D').

From the point of view of the principal, the overall effects of introducing the policy are ambiguous. There are clear direct effects in both directions: regulation prevents the overpayment by misaligned agents but it introduces compliance costs for everyone. On the other hand, the regulation also generates indirect effects through the behavioral responses of the agents. Bunching effectively reflects an increase in the bargaining power of the government, as the agency is able to procure the same contract for a smaller total payment than the one that would occur without regulation, and without having to incur in compliance costs. On the other hand, extensive margin responses reflect pure loss in the case of aligned agents, with compliance costs effectively destroying transactions that would have generated a positive surplus for the government. For misaligned agents, extensive margin responses are positive as long as they prevent transactions that were generating a negative net payoff for the government.

The relative magnitude of these effects in the aggregate will critically depend on how regulation affects the distribution of post-award performance and on the pervasiveness of the presence of misaligned agents. To quantitatively assess this trade-off, I parametrize the model and estimate it using the data described at length in Sections 2 and 3.

4.3. Model Estimation

I estimate the model using the sample of awards in 2006-2010 (prior to the threshold change). I treat each observed award p_i for $i = 1, ..., N$ as the equilibrium outcome of the model, given a set of model primitives. To take the model to the data, I make a series of parameterizations and simplifying assumptions.

First, the ex-post government valuation is $w_i = \bar{v}_i(1+\epsilon_i)$. This means that prior to the purchase, the agency has a reference valuation $\bar{v}_i$ determined by the characteristics of the project, and it expects this to be shifted by the post-award implementation shock ϵ_i. $\bar{v}_i$ is log-normally distributed: $\bar{v}_i \sim \log N(\mu_v, \sigma_v^2)$. Implementation shocks, on the other hand, are normally distributed. I normalize the distribution of post-award shocks under regulation to be standard normal, whereas the shock distributions under discretion represent shifts in either a positive or a negative direction depending on the bureaucrat's type. In particular, $(\epsilon_i|R=1) \sim N(0,1)$, $(\epsilon_i|A, R=0) \sim N(\Delta_A, 1)$, and $(\epsilon_i|M, R=0) \sim N(\Delta_M, 1)$, where $\Delta_A \geq 0 \geq \Delta_M$. Finally, costs are uniformly distributed on a range below ex-ante valuations in order to focus on cases with positive expected surplus, $c_i|\bar{v}_i \sim U\left(\frac{\bar{v}_i}{2}, \bar{v}_i\right)$.

The econometrician does not observe valuations, costs or post-award shocks. However, in addition to award amounts p_i, there is an observable index of implementation quality q_i that is related to the latent implementation shocks. In particular, $q_i = \delta_0 + \delta_1 \cdot \bar{v}_i + \epsilon_i$. The implementation quality index is allowed to depend linearly on $\bar{v}_i$ to capture the fact that, for reasons unrelated to the agent's actions or the regulatory framework, the implementation quality might vary for awards of different sizes. For example, we might expect that larger awards are —ceteris paribus— more likely to incur delays or overruns because of higher average project complexity. In practice I do not estimate the structural parameters δ_0 and δ_1. Instead, I simply calculate the linear reduced form relationship between $E[q_i]$ and p_i for awards above the regulatory threshold, and I use this to extrapolate to contracts below $\bar{P}$. This restriction fits well the evidence from Figure 6 while having the benefit of reducing the total number of parameters that need to be estimated.

The agent in charge of the transaction is aligned with probability λ and is misaligned otherwise. For the misaligned agent, I assume a fixed reservation value (which may be negative) of completing a regulated transaction b_R, and a stochastic reservation value of completing a discretionary transaction $b_M|v_A \sim U\left(0, \frac{v_A}{2}\right)$.

Red tape costs are normally distributed and censored at zero: $\kappa_i = \max\{0, \tilde{\kappa}_i\}$, with $\tilde{\kappa}_i \sim N(\mu_\kappa, \sigma_\kappa^2)$. The heterogeneity in compliance costs captures the fact that different agencies and contractors might have diverse levels of skill, resources, or experience dealing with highly regulated procurement; in other words, the effective burden that a given agent faces is not fixed.

Inspired by the commercial item exception, I assume that a fraction ψ of the products is exempt from the regulation and that misaligned agents cannot deviate from the Nash bargaining award in that case. In other words, for commercially available goods and services the cost is easy to verify and so regulation is not necessary. Since I do not observe commercial exceptions at the award level, I calibrate $\psi = 0.25$ to match the aggregate share of commercial item acquisitions based on an estimate by the US Government Accountability Office.[39]

Finally, I introduce adjustment frictions: with probability η, agents face frictions and are unable to âĂIJbunchâĂİ at the threshold. In this case they must choose only between going ahead with the transaction at the Nash bargaining award level with regulation and not transacting at all. This

[39]See GAO, "Contracting Data Analysis: Assessment of Government-wide Trends", http://www.gao.gov/products/GAO-17-244SP.

additional parameter is necessary to rationalize the data because in a frictionless model we would see much starker bunching than the one observed and a dominated region to the right of the threshold that would create a hole in the distribution.

I estimate the model via simulated method of moments (see McFadden, 1989, Pakes and Pollard, 1989). That is, I choose a vector of parameters θ to generate simulation-based moments that closely resemble key moments from the data. Formally, consider a vector of J moments from the data, given a sample size of n. Denote these *target moments* by m_n. On the other hand, I generate analogous moments by simulating $n \cdot s$ observations with the model. Denote these *simulated moments* as $m_s(\theta)$, noting that these depend on the parameters $\theta \in \Theta \subset \mathbb{R}^P$.

The estimated parameters are chosen to minimize a standard distance metric:

$$\hat{\theta} = \arg\min_{\theta} \; (m_n - m_s(\theta))' W_n (m_n - m_s(\theta)) \tag{7}$$

were W_n is a weighting matrix. Letting $M_s(\theta)$ be the $(P \times J)$ Jacobian matrix of the vector of simulated moments, we have under standard regularity assumptions:

$$\sqrt{n}\left(\hat{\theta} - \theta_0\right) \xrightarrow{d} N\left(0, \left(1 + \frac{1}{s}\right)\left(M'WM\right)^{-1} M'W\Omega WM \left(M'WM\right)^{-1}\right) , \tag{8}$$

where W is the probability limit of W_n, M is the probability limit of $M_s(\theta_0)$, and Ω is the asymptotic variance of m_n. The vector of parameters corresponds to:

$$\theta = \left(\mu_v, \sigma_v^2, \Delta_A, \Delta_M, \phi, \lambda, \gamma, \mu_\kappa, \sigma_\kappa^2, b_R, \eta\right) \quad .$$

I use two sets of target moments. The first set of moments is a smoothed vector of normalized frequencies on a window around the simplified acquisition threshold. In particular, I start by dividing the range of award amounts between \$50,000 and \$200,000 into 150 bins of \$1,000 width—i.e., (\$50,000,\$51,000], ..., (\$199,000,\$200,000]—and I compute the normalized frequency of awards at each bin, as in Figure 1. I then smooth this distribution to eliminate the spikes at round numbers, adjusting the distribution proportionally to satisfy an integration constraint. Appendix G describes this procedure formally.

The second set of moments is given by a smoothed vector of mean contract quality. Over the same range of awards (between \$50,000 and \$200,000), I regress the quality index on a polynomial of award value at each side of the simplified acquisition threshold. I then extract a vector of 15 quality moments that correspond to the fitted value at \$55,000, \$65,000, ..., \$195,000 (see Appendix G for details).

Stacking together these two vectors, I obtain m_n, the vector of 165 moments that I will seek to match with the model. For the weighting matrix, I start from a consistent estimate of the inverse of the asymptotic variance of m_n (i.e., the optimal weighting matrix). This estimate is obtained via bootstrap, sampling contracts with replacement from the estimation sample. I then follow Einav, Finkelstein, and Mahoney (2018) and modify this matrix to assign greater weights to moments that are close to the simplified threshold, since most of the identification comes from this part of the distribution. In particular, the weights decrease by a constant amount (1/100) per thousand dollars away from the threshold. I use the stochastic optimization algorithm Differential Evolution (DE) (Storn and Price, 1997) to perform the objective minimization.

4.4. Identification

Although the previous section makes clear that all the parameters are jointly estimated, certain features of the data are related more closely to each of the individual parameter estimates. In particular, certain estimates are most sensitive to different points of the award distribution moments, while other parameters are driven largely by the quality moments.

The parameters related to the government's valuation (μ_v, σ_v) are determined by the shape of the award distribution well below the simplified threshold, where the award density is simply a fixed transformation of the primitive valuation distribution. The shape of the excess mass below the threshold pins down the estimates of bargaining weights (ϕ) and red tape cost shares (γ), while the *magnitude* of the spike below the threshold identifies the average size of the compliance costs (μ_κ). The area just to the right of the threshold —and, in particular, the absence of a hole in the distribution— will determine the dispersion in red tape costs (σ_κ) as well as the adjustment frictions (η).[40] Finally, the levels of the distribution well to the right of the threshold will identify the outside option parameter b_R, as they govern part of the extensive margin responses.

The rest of the parameters $(\lambda, \Delta_A, \Delta_M)$ are identified by the quality moments below the simplified acquisition threshold. Because the parameters discussed in the previous paragraph fully characterize the award distributions for both types of agents, they pin down how aligned and misaligned agents differentially sort into different parts of the distribution (how they bunch at different rates, how they locate slightly below the threshold at some other rates, etc.). Therefore, at any bin b at or below the threshold, the share of aligned agents will be $\tau_b(\lambda)$, where $\tau_b(\cdot)$ is an increasing function that is identified given the parameters discussed above. Therefore, mean post-award performance shocks at that bin will be $\bar{\epsilon}_b = \tau_b(\lambda) \cdot \Delta_A + [1 - \tau_b(\lambda)] \cdot \Delta_M$, which means that $(\lambda, \Delta_A, \Delta_M)$ are identified given at least three linearly independent observations of $\bar{\epsilon}_b$. The quality moments below the threshold provide such observations.

Note that the quality moments above the threshold do not provide additional identification. Above the threshold, post-award performance shocks are simply normalized to have mean zero, and the quality moments are used to determine the parameters δ_0 and δ_1, which are not part of the main estimation. This helps to see intuitively where the identification of λ comes from. Given that $\Delta_A \geq 0$ and $\Delta_M \leq 0$, a high value of λ will be necessary to have positive average performance shocks below the threshold, whereas a low value will be needed to justify negative average shocks below $\bar{P}$. This implies that a quality drop (jump) at the threshold will reflect a high (low) degree of agent alignment.

[40] This is analogous to the model in Kleven and Waseem (2013), where the dispersion in elasticities and the adjustment frictions fill part of the missing mass that we would expect just to the right of the notch.

5. Model Results

5.1. Parameter Estimates and Model Fit

Table 6 presents the parameter estimates. Several facts are worth highlighting. First, compliance costs are substantial and highly heterogeneous. The estimated mean red tape costs are \$12.8 thousand, which can represent up to 12.8% of the award amount (the smallest contract subject to regulation is \$100,000). The standard deviation corresponds to \$10.2 thousand.

Second, a large majority (82%) of the contracting officers act as if they are fully aligned with the government's goal. The average post-award performance shock is 0.08 units of standard deviation larger than when non-simplified acquisition is used. The converse is true for misaligned agents, who under simplified acquisition have contracts with 0.03 units of standard deviation worse post-award shocks. As a result, the average performance under simplified acquisition is significantly better than above the threshold, consistent with the results from Figure 6.

Third, the government's bargaining weight is estimated to be 0.4, while the agency absorbs a quarter of the compliance costs. However, note that, at least for the aligned agents, the true incidence of compliance costs is affected not by the γ parameter shares but by the bargaining power parameter. Compliance costs shrink the size of the surplus, and this reduces the parties' payoffs in proportion to their bargaining weights, regardless of how red tape costs are shared. This is perhaps more evident if we think about regulation as a transaction tax, as in Kopczuk and Munroe (2015): γ represents the statutory incidence, but the economic incidence will be determined by the bargaining power of each agent.

Overall, the model is able to closely replicate the key empirical patterns in the estimation sample. Figure 7 presents the data and corresponding simulated moments: Panel (a) presents the model fit for the award distribution moments, while Panel (b) depicts the fit for the contract quality moments. Because I estimate the model using data from 2006 through 2010 (when the threshold was fixed at \$100,000), I ask whether the model can also fit the data *after* the threshold was increased to \$150,000. Figure 8 presents the results of this exercise, with simulated moments using a value of $\bar{P} = 150$. Reassuringly, these out-of-sample predictions fit the data reasonably well.

5.2. Counterfactuals

Now we can go beyond the observed levels of regulation during the sample period and construct counterfactuals for other levels of the simplified acquisition threshold. Given the real spending responses documented in Section 3, I ask how much total contract spending in awards between \$50,000 and \$300,000 would be under different threshold values. Figure 9, which presents the results of this exercise, shows that contract spending increases at a decreasing rate as the threshold value becomes larger. This decreasing slope is explained by the fact that both intensive and extensive margin responses become less consequential as the threshold grows. On the one hand, the mass of contracts around the threshold that is susceptible to intensive margin responses shrinks as the threshold increases. On the other hand, and given the parametrization, the probability that the gross surplus is smaller than compliance costs goes to zero as the threshold grows, making extensive margin responses less likely for larger value contracts.

One of these counterfactuals is of particular interest: raising the value of the simplified acquisition threshold to \$250,000, as stipulated in the National Defense Authorization Act of 2018 (approved, but not yet fully implemented). The results from Figure 9 indicate that spending in awards between \$50,000 and \$300,000 would increase by 18% relative to the baseline pre-2011 period. With respect to the current regulation (the \$150,000 threshold), this constitutes a spending increase of roughly 8%. Extensive margin responses account for three-quarters of these effects.

Finally, we can also construct a counterfactual with an infinite value for the regulatory threshold—i.e., a world in which all purchases are allowed to use simplified acquisition. Appendix Figure A11 presents this counterfactual along with the simulated values for $\bar{P} = 100$, which is the estimated analog of Figure 2(a).

5.3. Optimal Regulation

Now we move beyond purely positive analysis and derive implications for the optimal level of regulation for the government. The model formalizes the trade-off associated with regulation as follows. First, rules help the government by controlling misaligned agents. By eliminating asymmetric information, regulation prevents these agents from overpaying for goods and services. Second, regulation introduces compliance costs on all regulated transactions, which hurts the government. Third, compliance costs generate transaction responses. Intensive margin responses benefit the government because they allow agencies to obtain goods and services for a lower total award: this regulation effectively increases the bargaining position of public buyers. The effect of extensive margin responses, on the other hand, depends on the type of agent in charge of the transaction. With aligned agents, extensive margin responses represent a pure loss because they reflect the destruction of surplus. With misaligned agents, extensive margin responses can be beneficial when they prevent wasteful transactions. Fourth, regulation modifies the expected post-award performance, which has ambiguous effects for the principal, again depending on the type of agent that represents the government. Rules negatively affect the distribution of post-award performance of aligned agents relative to a discretionary benchmark, yet the opposite is true for misaligned agents.

As this discussion highlights, the overall desirability of this type of regulation is ex-ante ambiguous for the government. Moreover, it emphasizes that the answer critically depends on the degree of misalignment that exists between government agencies and contracting officers. This intuition is at the heart of the rules vs. discretion debate: regulation can be an effective antidote to waste and abuse whenever these are pervasive, but it can backfire if most agencies would actually be effective when exercising discretion. Because the estimated degree of alignment for this case is remarkably high ($\lambda = 0.82$), the conclusion is that the government would benefit from reducing the stringency of regulation.

Figure 10 shows this quantitatively. The plot displays the government's payoff as a function of the simplified acquisition threshold, and shows that the \$100,000 level in force during the estimation period was lower than optimal. Indeed, the later increases to \$150,000 and \$250,000 are estimated to have improved —all else equal— the overall benefit for the federal government. The optimal simplified threshold for the estimation period is \$295,000, which in 2019 dollars corresponds to roughly \$345,000.[41] The model generates an interior optimum because compliance costs are fixed, whereas the key benefit of regulation —controlling misaligned agents— grows with the size of the award. As the size of the contract increases, red tape represents a small cost as a share of the contract value, whereas the potential loss from wasteful spending becomes substantial.

Reiterating the importance of the role that the degree of misalignment plays in this model, Figure 11 shows the relationship between λ and the optimal threshold. For low levels of λ—i.e., a high degree of misalignment—the government should set a stringent policy that has a low simplified acquisition threshold. As misalignment is reduced, the government can increase the threshold to allow more contracts to qualify for simplified acquisition.

This counterfactual highlights the appeal of the model for applications beyond this particular setting. Using the exact same structure, estimating this model on data from a different institutional environment—say, one in which preference misalignment in the form of either corruption is much more pervasive—might very well lead to the result that discretion should be restricted. Beyond telling us *if* more discretion is optimal, the model quantitatively highlights *when* more discretion is desirable: namely, when bureaucracies feature a high enough level of alignment, so that the risk of generating waste is low relative to the efficiency gains in the form of reduced red tape and quality changes. This feature of the model is consistent with the recently proposed framework by Bosio, Djankov, Glaeser, and Shleifer (2022), and can rationalize the cross-country evidence presented by these authors, enabling a quantitative assessment of when discretion enhances procurement efficiency.

5.4. Model Discussion

In order to simplify the analysis and focus on a few key mechanisms, the model abstracts away from several important issues. Here I discuss some considerations that are omitted from the current analysis but could motivate further extensions in future work.

An important modeling assumption is that compliance costs introduced by regulation are fixed. In reality, part of these costs may scale up with the size of the contract: e.g., drafting evaluation plans might become a more involved task as the project complexity increases. Note, however, that since the loss generated by misaligned agents grows one-to-one with award size, as long as compliance costs are less than proportional to the size of the award, the qualitative prediction of an interior optimum will prevail.

The model also assumes that each contract represents the purchase of a fixed and indivisible good or service, without the possibility of adjustments along the margins of the quantity purchased and/or product quality. The model then interprets intensive margin responses as reflecting pure price adjustments: a contract bunching at the threshold of \$100,000 that would have been awarded for \$110,000 in the absence of regulation represents a saving of \$10,000 for the government. While

[41]Of course, it is important to keep in mind that \$295,000 is a value that is significantly out of sample. Recall that the estimation used moments between \$50,000 and \$200,000, and, therefore, extrapolating above this range should warrant some additional caution.

incorporating quantity or quality adjustments would naturally affect the *quantitative* predictions, they would reinforce the conclusion that the government would benefit from more discretion, leading to an even higher estimate of the optimal threshold. This is so because when the threshold is raised, the government would obtain additional quantity/quality per award that currently is unaccounted for.

Another strong assumption is that the share of misaligned agents is exogenous. Although it is outside of the scope of the model, the degree of waste and abuse could endogenously respond to the stringency of regulation.[42] This could operate through at least two margins. First, higher thresholds might increase the rewards to corrupt behavior, which would increase the misalignment share. This force will shift the conclusions towards a more conservative (i.e., lower) threshold. Second, discretion might change the pool of agents who select into public procurement jobs. This effect is ex-ante ambiguous because discretion can attract agents who have more initiative or agents who are more likely to engage in wrongdoing.

The results might also be altered by considering different government's objectives. For example, risk aversion might call for more stringent regulation: Congress might be willing to pay a risk premium to eliminate expected losses generated by misaligned agents. Similarly, the government may have reasons to put a higher weight on waste generated by misalignment, above and beyond the monetary costs (say, either due to electoral incentives or because corruption scandals erode public trust in the government). Again, this would push towards a lower regulatory threshold.

Finally, the optimal regulation has been analyzed from the perspective of the party that designs the transaction mechanism —in this case, the buyer/government. This is analogous to the approach taken by several classic papers in the auction and procurement literature (e.g., Myerson, 1981; Che, 1993; Bajari and Tadelis, 2001). A different but related question is what regulation maximizes social welfare. Inevitably, such welfare analysis would require us to take a stance on how much citizens value the goods and services procured by the government. And while this might be a conceivable exercise in more specific setups,[43] translating such a diverse set of contracts (such as those analyzed here) into measures of consumer surplus is beyond the scope of this paper. Note, however, that the qualitative economic forces emphasized here would still be at play in the social welfare analysis: red tape costs would be socially costly, intensive margin responses would prevent costly transfers to firms (due to the marginal cost of public funds), and extensive margin responses would either destroy social surplus or prevent wasteful transactions.

6. Conclusion

Designing an efficient public procurement system forces regulators to strike a balance between rules and discretion. Ultimately, they need to decide whether the benefits of preventing potential waste and abuse outweigh the costs of vigorous regulation. In this paper, I argue that the US federal

[42] Celentani and Ganuza (2002) propose a model of public procurement auctions where bureaucrats endogenously decide to become corrupt. The goal of their analysis is to examine how equilibrium corruption responds to changes in competition.

[43] For example, Lewis and Bajari (2011) perform welfare analysis in the context of procurement contracts for highway construction, where the time of project completion is linked to consumer surplus through commuting times.

government could obtain more value from procurement contracts by moving the balance in the direction of more discretion. I arrive at this conclusion after analyzing a large set of procurement contracts awarded over the course of a decade by all US federal agencies.

I first show that transactions are sensitive to the level of regulatory strength that is imposed upon them. The evidence suggests that market participants —agencies and firms— have a preference for transacting under a regime that has more discretion and less oversight. Because this simplified regime is only available for awards below a certain threshold, I argue that the government awards smaller and fewer contracts than in a world with full discretion.

How to interpret these transaction responses depends critically on how the regulatory environments affect the quality of contract spending. If rules are able to control misaligned bureaucrats and prevent wasteful awards, then contract performance should be superior under stronger regulation. Yet if abuse and inefficiency are the exception rather than the norm, contract performance will be enhanced by allowing agencies to exercise discretion. The evidence provided in this paper supports the latter hypothesis, and it is therefore consistent with Kelman (2005)'s assessment of the traditional procurement system, where "*rules misdirected the behavior of the many to stop the abuses of a few*."

I formalize this logic by proposing a stylized model of public procurement in which there is a principal-agent structure on the buying party's side. Regulation in the model eliminates the problem of preference misalignment, but it imposes compliance costs. The magnitude of the transaction responses reveal the size of these compliance costs, while the effect of regulation on contract performance identifies the degree of preference misalignment. Beyond rationalizing the observed findings, the model allows me to make quantitative assessments of policy-relevant regulatory choices. I find that a new proposed simplified acquisition threshold of $250,000 will increase the net benefit obtained by the government from its procurement contracts.

While results presented here are of course specific to this particular setting, the analysis can shed light on some of the general economic forces that will be at play in other procurement contexts. Indeed, the main trade-off highlighted by the model helps rationalize seemingly contradictory evidence from country-specific studies, and is consistent with cross-country evidence produced by Bosio et al. (2022). My work makes clear that both compliance costs and preference misalignment have to be simultaneously considered and provides a methodology to quantify these in a given setting. Considering the size of the markets for government contracts around the world, policy improvements that incorporate this analysis may result in very large savings of resources for taxpayers.

References

Aghion, P., N. Bloom, and J. Van Reenen (2013). Incomplete contracts and the internal organization of firms. *The Journal of Law, Economics, and Organization 30*(Supplement 1), i37–i63.

Aghion, P. and J. Tirole (1997). Formal and real authority in organizations. *Journal of Political Economy 105*(1), 1–29.

Bajari, P., S. Houghton, and S. Tadelis (2014). Bidding for incomplete contracts: An empirical analysis of adaptation costs. *The American Economic Review 104*(4), 1288–1319.

Bajari, P. and S. Tadelis (2001). Incentives versus transaction costs: A theory of procurement contracts. *The RAND Journal of Economics 32*(3), 387–407.

Bandiera, O., M. C. Best, A. Q. Khan, and A. Prat (2021, 08). The Allocation of Authority in Organizations: A Field Experiment with Bureaucrats*. *The Quarterly Journal of Economics 136*(4), 2195–2242.

Bandiera, O., A. Prat, and T. Valletti (2009). Active and Passive Waste in Government Spending: Evidence from a Policy Experiment. *American Economic Review 99*(4), 1278–1308.

Best, M. C., J. Hjort, and D. Szakonyi (2018). Sources of Variation in State Effectiveness and Consequences for Policy Design. *National Bureau of Economic Research Working Paper Series No. 23350*.

Best, M. C. and H. J. Kleven (2018). Housing Market Responses to Transaction Taxes: Evidence from Notches and Stimulus in the U.K. *Review of Economic Studies 85*(1), 157–193.

Bosio, E., S. Djankov, E. Glaeser, and A. Shleifer (2022). Public Procurement in Law and Practice. *American Economic Review* (Forthcoming).

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

Carril, R. and M. Duggan (2020). The impact of industry consolidation on government procurement: Evidence from department of defense contracting. *Journal of Public Economics 184*, 104141.

Carril, R., A. Gonzalez, and M. S. Walker (2020). Competition under Incomplete Contracts and the Design of Procurement Policies. *Mimeo.*

Celentani, M. and J.-J. Ganuza (2002). Corruption and competition in procurement. *European Economic Review 46*(7), 1273 – 1303.

Che, Y.-K. (1993). Design competition through multidimensional auctions. *The RAND Journal of Economics 24*(4), 668–680.

Chetty, R., J. N. Friedman, T. Olsen, and L. Pistaferri (2011). Adjustment Costs, Firm Responses, and Micro vs. Macro Labor Supply Elasticities: Evidence from Danish Tax Records. *Quarterly Journal of Economics 126*(2), 749–804.

Colonnelli, E. and M. Prem (2022). Corruption and Firms. *Review of Economic Studies* (Forthcoming).

Coviello, D. and S. Gagliarducci (2017). Tenure in Office and Public Procurement. *American Economic Journal: Economic Policy 9*(3), 59–105.

Coviello, D., A. Guglielmo, and G. Spagnolo (2018). The effect of discretion on procurement performance. *Management Science 64*(2), 715–738.

Coviello, D. and M. Mariniello (2014). Publicity Requirements in Public Procurement: Evidence from a Regression Discontinuity Design. *Journal of Public Economics 109*, 76–100.

Crocker, K. J. and K. J. Reynolds (1993). The efficiency of incomplete contracts: An empirical analysis of air force engine procurement. *The RAND Journal of Economics 24*(1), 126–146.

Decarolis, F. (2014). Awarding Price, Contract Performance, and Bids Screening: Evidence from Procurement Auctions. *American Economic Journal: Applied Economics 6*(1), 108–132.

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

Decarolis, F., R. Fisman, P. Pinotti, and S. Vannutelli (2021). Rules, Discretion, and Corruption in Procurement: Evidence from Italian Government Contracting. *Mimeo*.

Decarolis, F., L. M. Giuffrida, E. Iossa, V. Mollisi, and G. Spagnolo (2018). Bureaucratic Competence and Procurement Outcomes. *National Bureau of Economic Research Working Paper Series No. 24201*.

Duflo, E., M. Greenstone, R. Pande, and N. Ryan (2018). The Value of Regulatory Discretion: Estimates From Environmental Inspections in India. *Econometrica 86*(6), 2123–2160.

Einav, L., A. Finkelstein, and N. Mahoney (2018). Provider Incentives and Healthcare Costs: Evidence from Long-Term Care Hospitals. *Econometrica 86*(6), 2161–2219.

Finan, F., B. Olken, and R. Pande (2017). The Personnel Economics of the Developing State. In A. Banerjee and E. Duflo (Eds.), *Handbook of Economic Field Experiments*, Volume 2, pp. 467–514. North-Holland.

Gagnepain, P., M. Ivaldi, and D. Martimort (2013). The cost of contract renegotiation: Evidence from the local public sector. *The American Economic Review 103*(6), 2352–2383.

Ganuza, J.-J. (2007). Competition and cost overruns in procurement. *The Journal of Industrial Economics 55*(4), 633–660.

GAO (2018). Contracting Data Analysis: Assessment of Government-wide Trends. *Government Accountability Office: Report to Congressional Adressees GAO-17-244SP*.

Gelber, A., D. Jones, D. W. Sacks, and J. Song (2018). Using Non-Linear Budget Sets to Estimate Extensive Margin Responses: Method and Evidence from the Earnings Test. *National Bureau of Economic Research Working Paper Series No. 23362*.

Gerardino, M. P., S. Litschig, and D. Pomeranz (2017). Can Audits Backfire? Evidence from Public Procurement in Chile. *National Bureau of Economic Research Working Paper Series No. 23978*.

Giuffrida, L. M. and G. Rovigatti (2020). Supplier Selection and Contract Enforcement: Evidence from Performance Bonding. *Mimeo*.

Gutman, J. (2014). Is There Room for Discretion? Reforming Public Procurement in a Compliance-Oriented World. *The Brookings Institution — Global Economy & Development Working Paper No. 74*.

Kang, K. and R. A. Miller (2022). Winning by Default: Why is there so Little Competition in Government Procurement. *Review of Economic Studies* (Forthcoming).

Kelman, S. (1990). *Procurement and Public Management: The Fear of Discretion and the Quality of Government Performance*. Washington, D.C.: American Enterprise Institute.

Kelman, S. (2005). *Unleashing change: A study of organizational renewal in government*. Washington, D.C.: Brookings Institution.

Kleven, H. J. (2016). Bunching. *Annual Review of Economics 8*(1), 435–464.

Kleven, H. J. and M. Waseem (2013). Using Notches to Uncover Optimization Frictions and Structural Elasticities: Theory and Evidence from Pakistan. *Quarterly Journal of Economics 128*(2), 669–723.

Kling, J. R., J. B. Liebman, and L. F. Katz (2007). Experimental analysis of neighborhood effects. *Econometrica 75*(1), 83–119.

Kopczuk, W. and D. Munroe (2015). Mansion Tax: The Effect of Transfer Taxes on the Residential Real Estate Market. *American Economic Journal: Economic Policy 7*(2), 214–257.

Laffont, J.-J. and J. Tirole (1993). *A Theory of Incentives in Procurement and Regulation*. Cambridge, MA: MIT Press.

Lewis, G. and P. Bajari (2011). Procurement Contracting With Time Incentives: Theory and Evidence. *The Quarterly Journal of Economics 126*(3), 1173–1211.

Lewis-Faupel, S., Y. Neggers, B. A. Olken, and R. Pande (2016). Can Electronic Procurement Improve Infrastructure Provision? Evidence from Public Works in India and Indonesia. *American Economic Journal: Economic Policy 8*(3), 258–283.

Liebman, J. B. and N. Mahoney (2017). Do Expiring Budgets Lead to Wasteful Year-End Spending? Evidence from Federal Procurement. *American Economic Review 107*(11), 3510–3549.

Marx, B. M. (2018). Dynamic Bunching Estimation with Panel Data. *MPRA Paper No. 88647*.

McFadden, D. (1989). A method of simulated moments for estimation of discrete response models without numerical integration. *Econometrica 57*(5), 995–1026.

Myerson, R. B. (1981). Optimal auction design. *Mathematics of Operations Research 6*(1), 58–73.

Olken, B. A. (2007). Monitoring Corruption : Evidence from a Field Experiment in Indonesia. *The Journal of Political Economy 115*(2), 200–249.

Pakes, A. and D. Pollard (1989). Simulation and the asymptotics of optimization estimators. *Econometrica 57*(5), 1027–1057.

Persson, P. (2019). Social Insurance and the Marriage Market. *The Journal of Political Economy* (Forthcoming).

Pertold, F. and J. Palguta (2017). Manipulation of Procurement Contracts: Evidence from the Introduction of Discretionary Thresholds. *American Economic Journal: Economic Policy 9*(2), 293–315.

Rasul, I. and D. Rogger (2018). Management of bureaucrats and public service delivery: Evidence from the nigerian civil service. *The Economic Journal 128*(608), 413–446.

Saez, E. (2010). Do Taxpayers Bunch at Kink Points? *American Economic Journal: Economic Policy 2*(3), 180–212.

Slemrod, J. (2013). Buenas notches: lines and notches in tax system design. *eJournal of Tax Research 11*(3), 259–283.

Spiller, P. T. (2008). An institutional theory of public contracts: Regulatory implications. *National Bureau of Economic Research Working Paper Series No. 14152*.

Storn, R. and K. Price (1997). Differential Evolution âĂŞ A Simple and Efficient Heuristic for global Optimization over Continuous Spaces. *Journal of Global Optimization 11*(4), 341–359.

Szucs, F. (2021). Discretion and Corruption in Public Procurement. *Mimeo*.

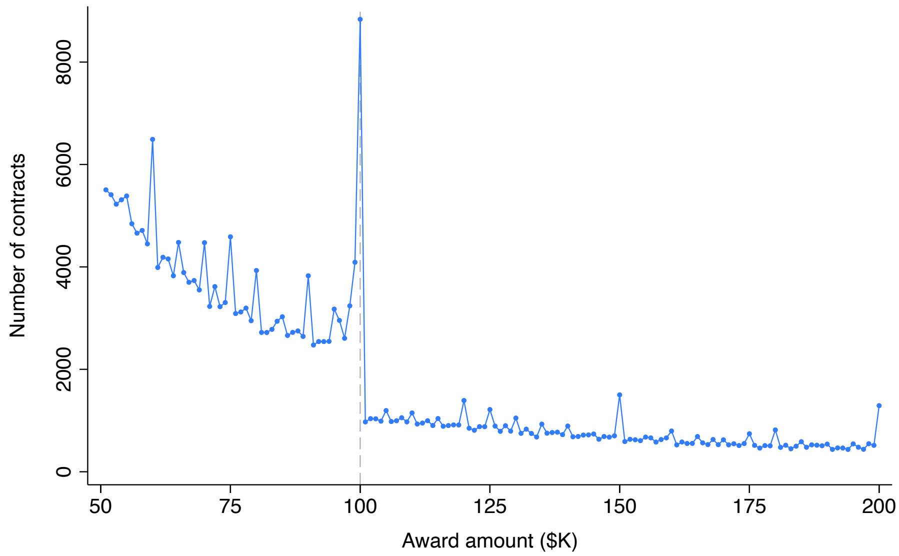
Warren, P. L. (2014). Contracting Officer Workload, Incomplete Contracting, and Contractual Terms. *RAND Journal of Economics 45*(2), 395–421.

Yang, D. (2008). Can Enforcement Backfire? Crime Displacement in the Context of Customs Reform in the Philippines. *The Review of Economics and Statistics 90*(1), 1–14.

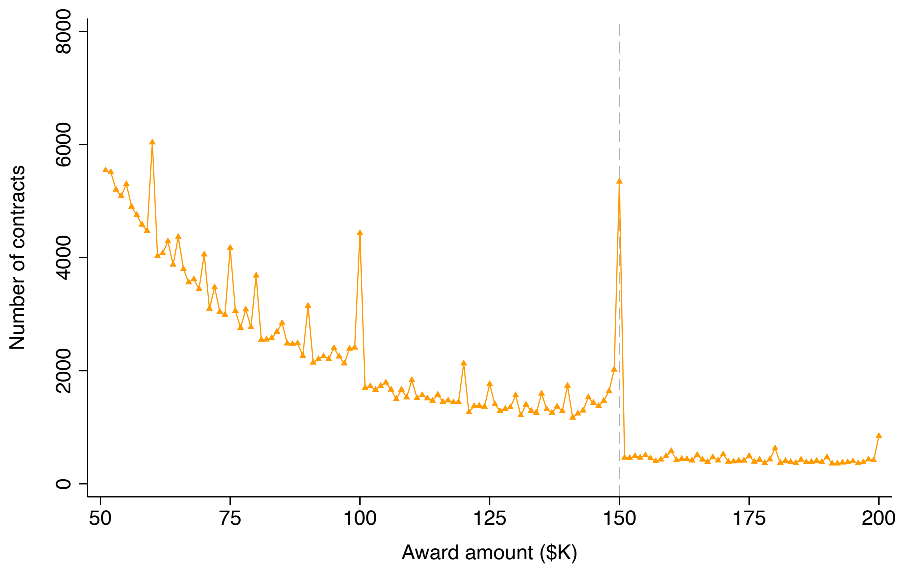
Figures

Figure 1: Distribution of contract awards

(a) "Pre" period: 2006-2010



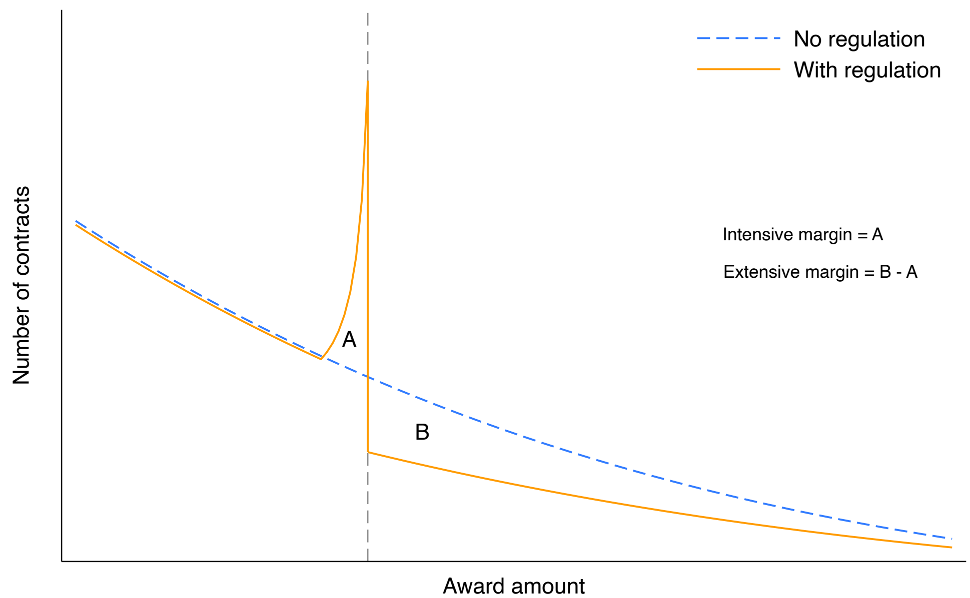
(b) "Post" period: 2012-2016



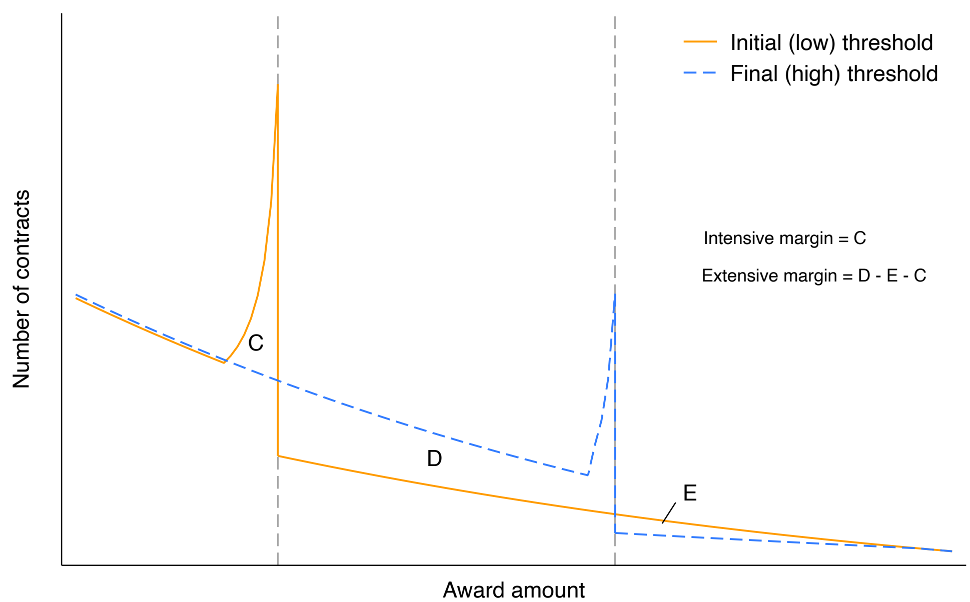
Notes: This figure shows contract awards frequency distributions in two periods, 2006-2010 and 2012-2016. The sample consists of non-R&D definitive contracts and purchase orders from the Federal Procurement Data System-Next Generation, with an award amount between $50,000 and $200,000. Award amounts are discretized into right-inclusive bins of one-thousand dollars length. A bin labeled "X" includes all contract awards in the range $(\$1000(X-1), \$1000X]$. Vertical dashed lines indicate the location of the simplified acquisition threshold in each period. Below the threshold, contracts can be awarded using (high-discretion) simplified acquisition procedures, whereas above the threshold non-exempt awards are subject to every acquisition law in the Federal Acquisition Regulation.

Figure 2: Potential responses to regulation

(a) Introducing regulation



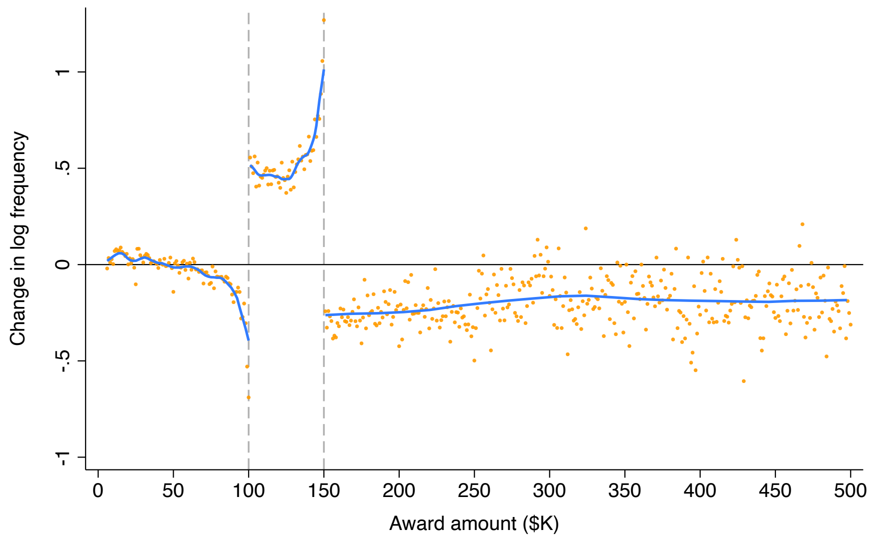
(b) Increasing regulatory threshold



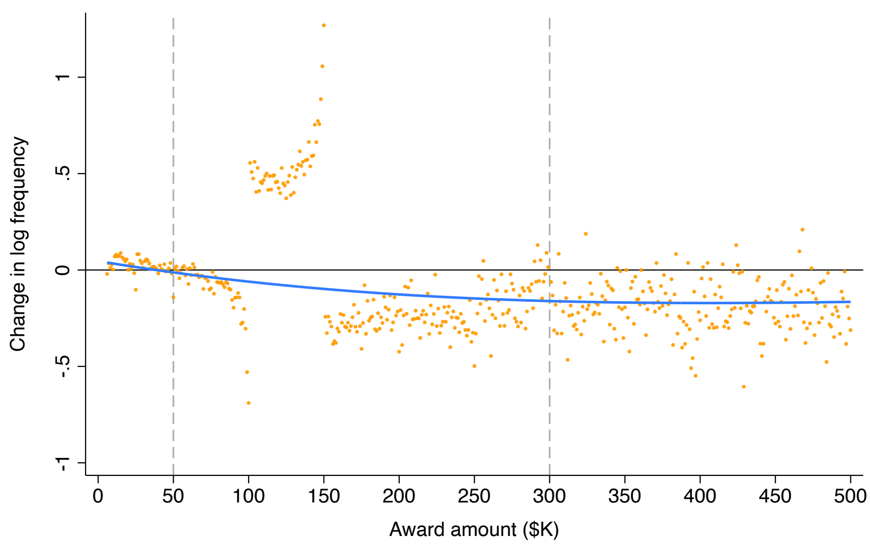
Notes: This figure shows potential changes in the distribution of contract awards in response to regulation. Panel (a) presents a comparison between a situation in which all contracts are allowed to use high-discretion simplified procedures (dashed line distribution), and a world in which only awards below a certain threshold are allowed to use them (solid line distribution). Panel (b) depicts a situation in which an existing threshold is raised, making the policy less stringent. The location of the thresholds is indicated with a vertical dashed line. Intensive margin effects are captured by the left-most "bunching" areas, (A and C, respectively). Extensive margin responses are given by the difference between the initial distribution and the final distribution (B-A and D-E-D, respectively).

Figure 3: Relative change in contracting and interpolation method

(a) Nonparametric smoothing

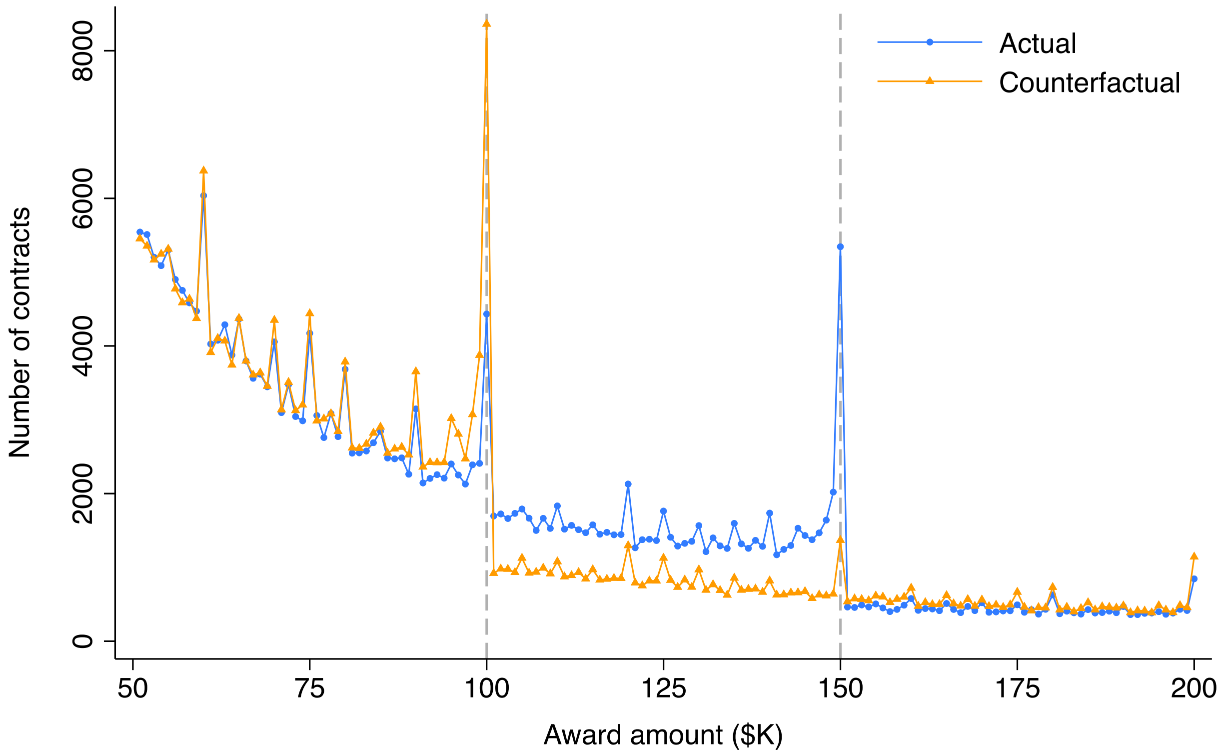


(b) Polynomial interpolation



Notes: This figure illustrates the interpolation method used to estimate the counterfactual distribution of Figure 4. Each orange dot depicts the change in log frequency between the "pre" (2006-2010) and "post" (2012-2016) periods for each award amount bin. The sample consists of non-R&D definitive contracts and purchase orders from the Federal Procurement Data System-Next Generation, with an award amount between $5,000 and $1,000,000. Award amounts are discretized into right-inclusive bins of one-thousand dollars length. A bin labeled "X" includes all contracts in the range $(\$1000(X-1), \$1000X]$. Panel (a) presents the frequency changes by award amount bin, along with a nonparametric kernel smoothing function (blue solid line) fitted over three segments: below the pre-period simplified threshold of $100,000; between this pre-period threshold and the post-period threshold of $150,000; and above this post-period threshold. The pre- and post- period thresholds are depicted by vertical dashed lines. In panel (b), the blue solid line is constructed by first regressing the log frequency changes by award bin on a third degree polynomial of award amounts and a set of dummies for each of the bins in the excluded region, defined as between $50,000 and $300,000. The boundaries of the excluded region are depicted as vertical dashed lines. The blue line is the fit of this regression, *excluding* the contribution of the dummies within the excluded region.

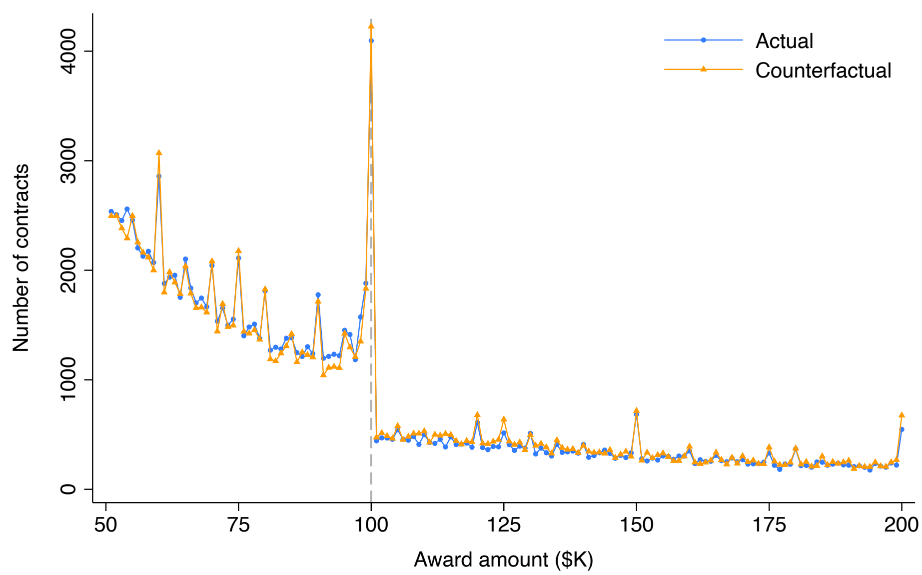
Figure 4: Counterfactual estimation: no threshold change



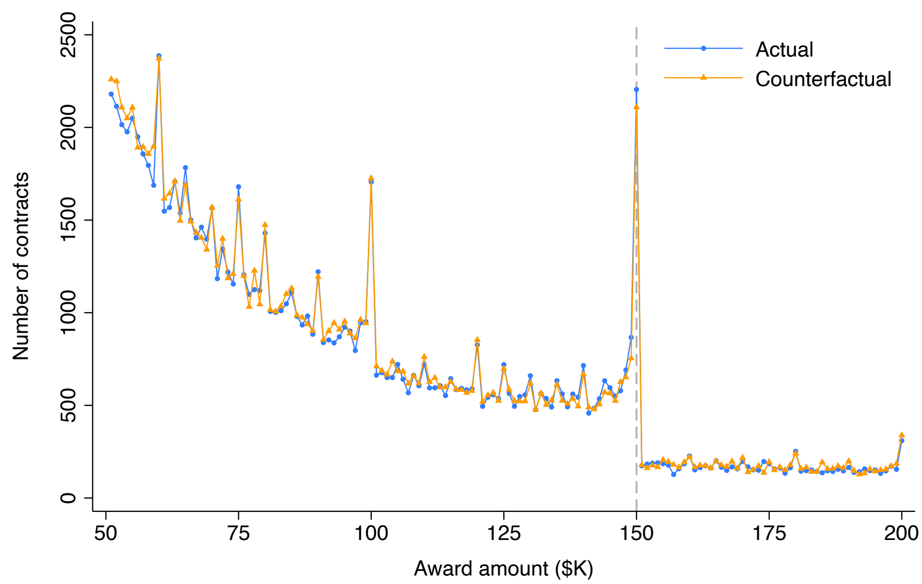
Notes: This figure shows the contract awards frequency distribution in the period 2012-2016 (blue dots) and the counterfactual distribution that we would have observed in that same period, had the simplified acquisition threshold been kept constant at $100,000 rather than raised to $150,000 (orange triangles). The sample consists of non-R&D definitive contracts and purchase orders from the Federal Procurement Data System-Next Generation, with an award amount between $50,000 and $200,000. Award amounts are discretized into right-inclusive bins of one-thousand dollars length. A bin labeled "X" includes all contracts in the range $(\$1000(X-1), \$1000X]$. The counterfactual distribution is computed by taking the pre-period distribution (2006-2010) and adjusting it for predicted frequency changes, based on an interpolation of frequency changes from contracts below $50,000 and above $300,000. The predicted frequency changes are shown in Figure 3(b).

Figure 5: Placebo estimates

(a) 2008-2009 predicted based on 2006-2007

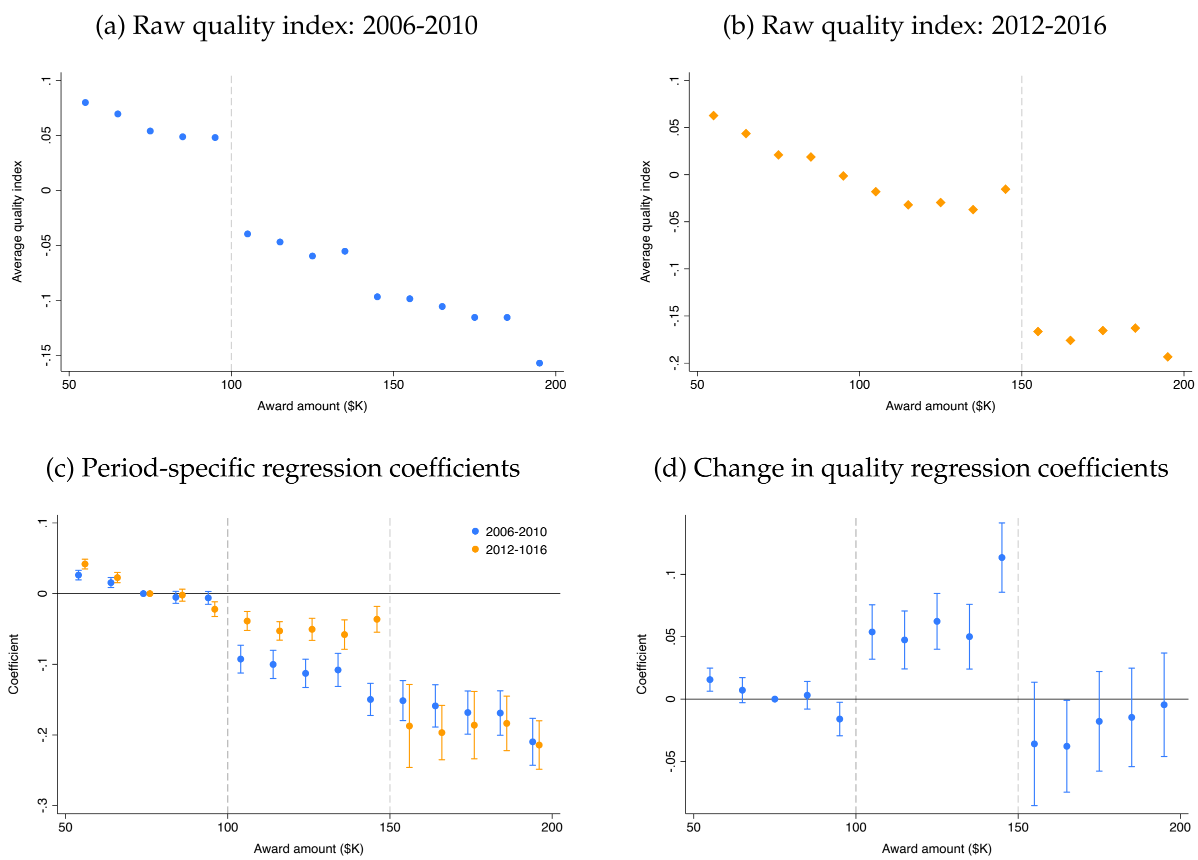


(b) 2015-2016 predicted based on 2013-2014



Notes: This figure shows the results from the placebo exercises for 2008-2009 (panel a) and 2015-2016 (panel b). The sample consists of non-R&D definitive contracts and purchase orders from the Federal Procurement Data System-Next Generation, with an award amount between $50,000 and $300,000. Award amounts are discretized into right-inclusive bins of one-thousand dollars length. A bin labeled "X" includes all contracts in the range $(\$1000(X-1), \$1000X]$. Both panel (a) and panel (b) present graphs that are constructed in the same way as Figure 4: they compare the actual distribution in some period, with a counterfactual distribution estimated by adjusting the distribution in the previous period to account for predicted frequency changes. These predicted frequency changes are based on an interpolation of frequency changes from contracts below $50,000 and above $300,000. The key difference between this figure and Figure 4 is that in these placebos no regulatory change occurs between the two consecutive periods. The vertical dashed lines on each panel indicate the simplified acquisition threshold in the relevant period.

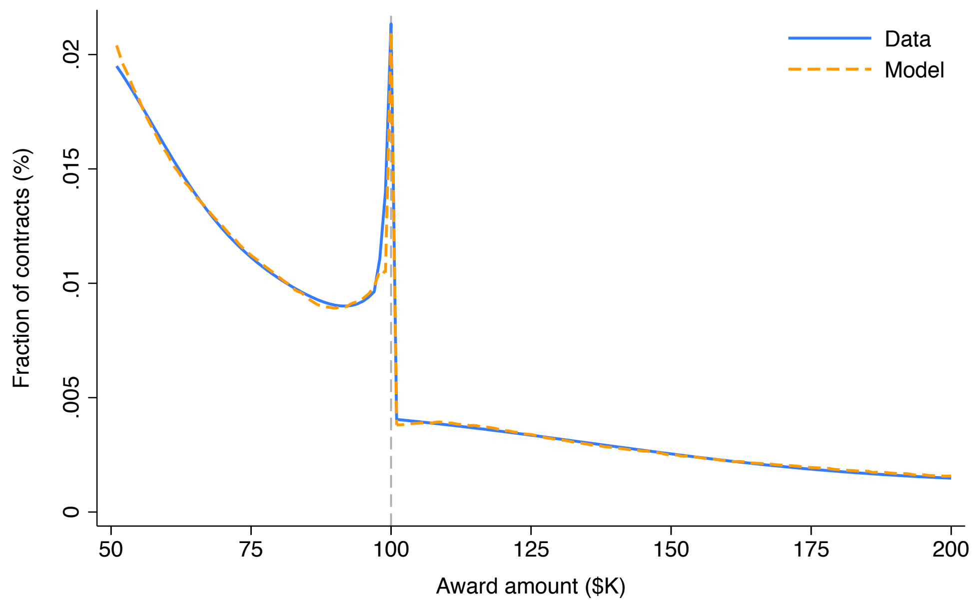
Figure 6: Contract quality



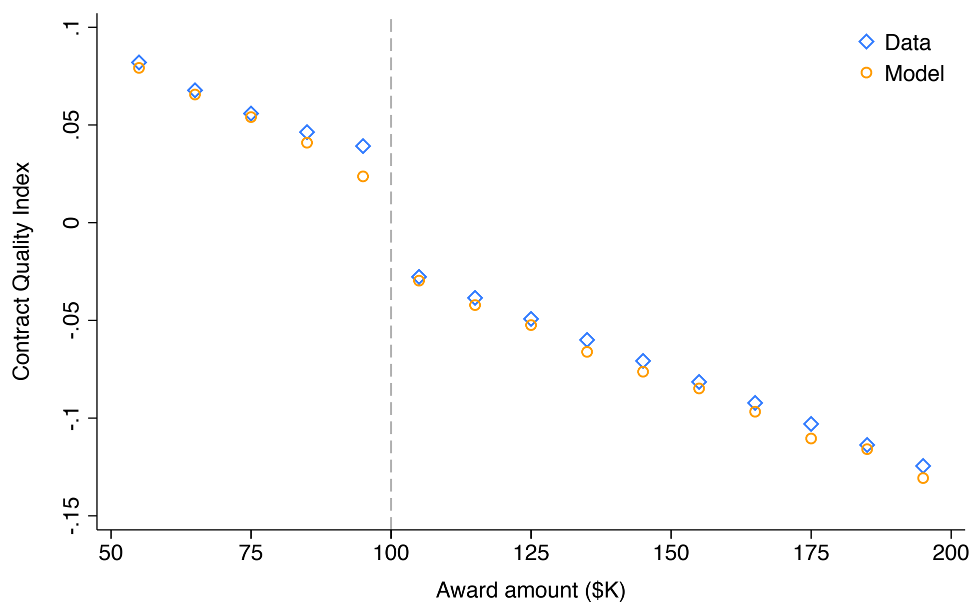
Notes: Panels (a) and (b) respectively show the average contract quality index by bins of award amounts, both in the "pre" (2006-2010) and "post" (2012-2016) periods. Panel (c) shows period-specific regression coefficients of specification (3) using contract quality index as dependent variable and no controls. Coefficients can be interpret as normalized average contract quality by period and by bin. Panel (d) presents regression coefficients from specification (4) using contract quality index as dependent variable and no controls. Coefficients can be interpreted as the change in the average contract quality index between the pre and post periods, by bin of award amount. The sample consists of non-R&D definitive contracts and purchase orders from the Federal Procurement Data System-Next Generation, with an award amount between $50,000 and $300,000. Award amounts are discretized into right-inclusive bins of ten-thousand dollars length. The contract quality index is constructed by averaging five normalized performance proxies: the number of within-scope modifications; within-scope cost overruns (in dollars); within-scope delays (in days); the probability of contract termination; and the probability of award in the last week of the fiscal year. Performance proxies are normalized by subtracting the pre-period mean and dividing them by the pre-period (2006-2010) standard deviation. The vertical dashed lines indicate the simplified acquisition threshold in each period. These correspond to $\$100,000$ in the pre-period (2006-2010) and $\$150,000$ in the post-period (2012-1016).

Figure 7: Model fit: in-sample (2006-2010)

(a) Award distribution



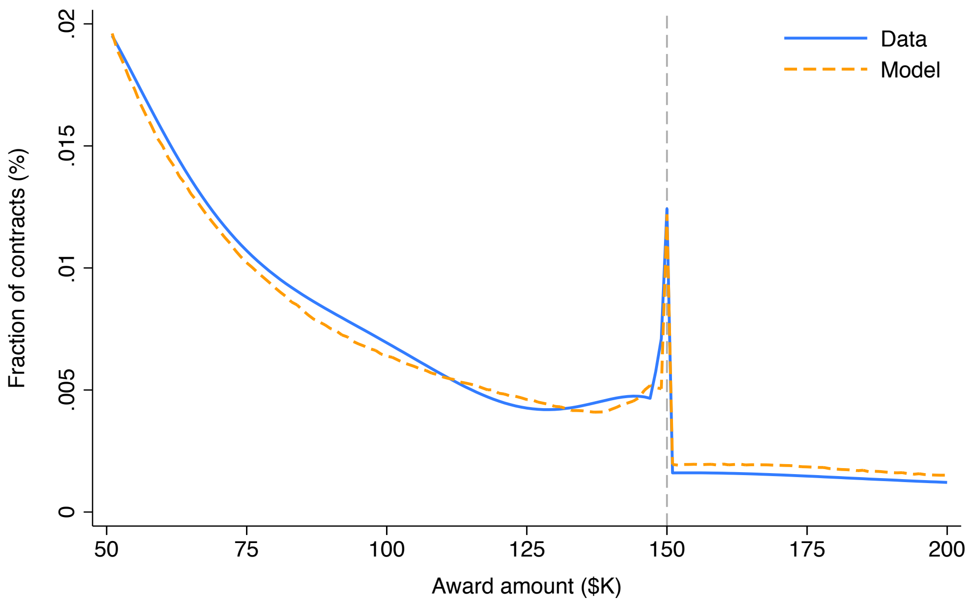
(b) Contract quality



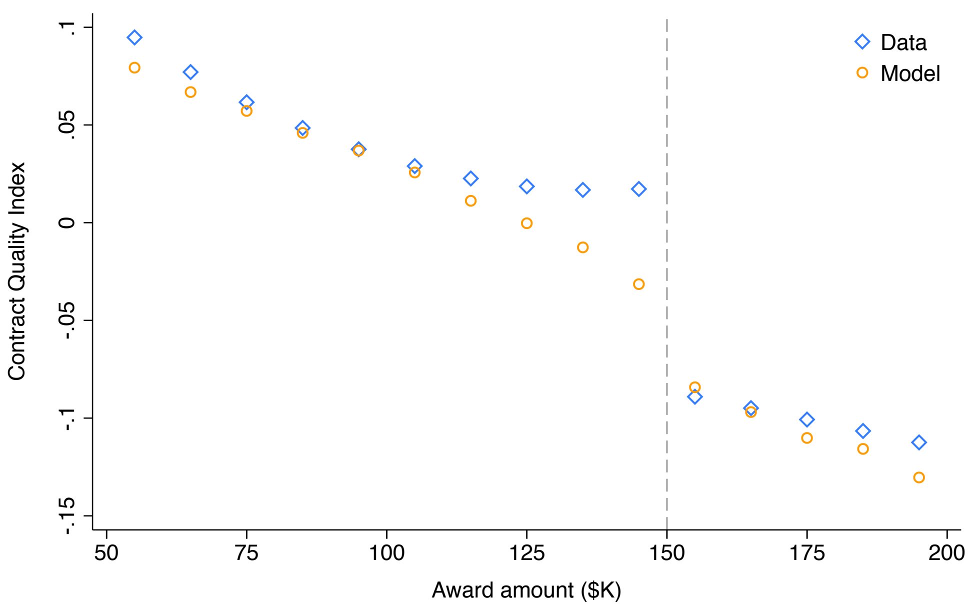
Notes: This figure shows the model fit within the estimation sample (2006-2010). The estimation sample consists of non-R&D definitive contracts and purchase orders from the Federal Procurement Data System-Next Generation, with an award amount between $50,000 and $300,000. Both panels compare smoothed moments from the estimation sample (blue solid line) with model-based simulated moments (orange dashed line). The moments used in Panel (a) are the smoothed fraction of contracts for each $1,000-wide award bin between $50,000 and $300,000. The moments used in Panel (b) are the smoothed average contract quality index for each $5,000-wide award bin between $50,000 and $300,000. Moments are smoothed according to the procedure described in Appendix G. The vertical dashed line at $100,000 in both panels indicates the simplified acquisition threshold.

Figure 8: Model fit: out-of-sample (2012-2016)

(a) Award distribution

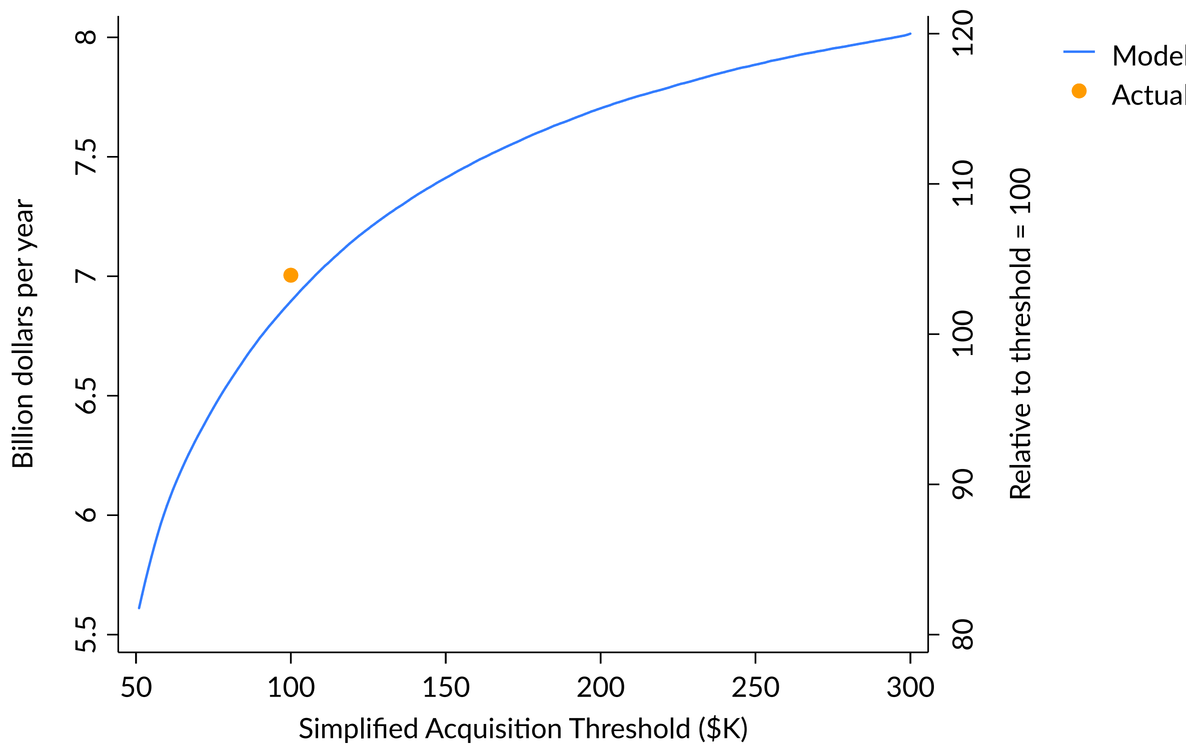


(b) Contract quality



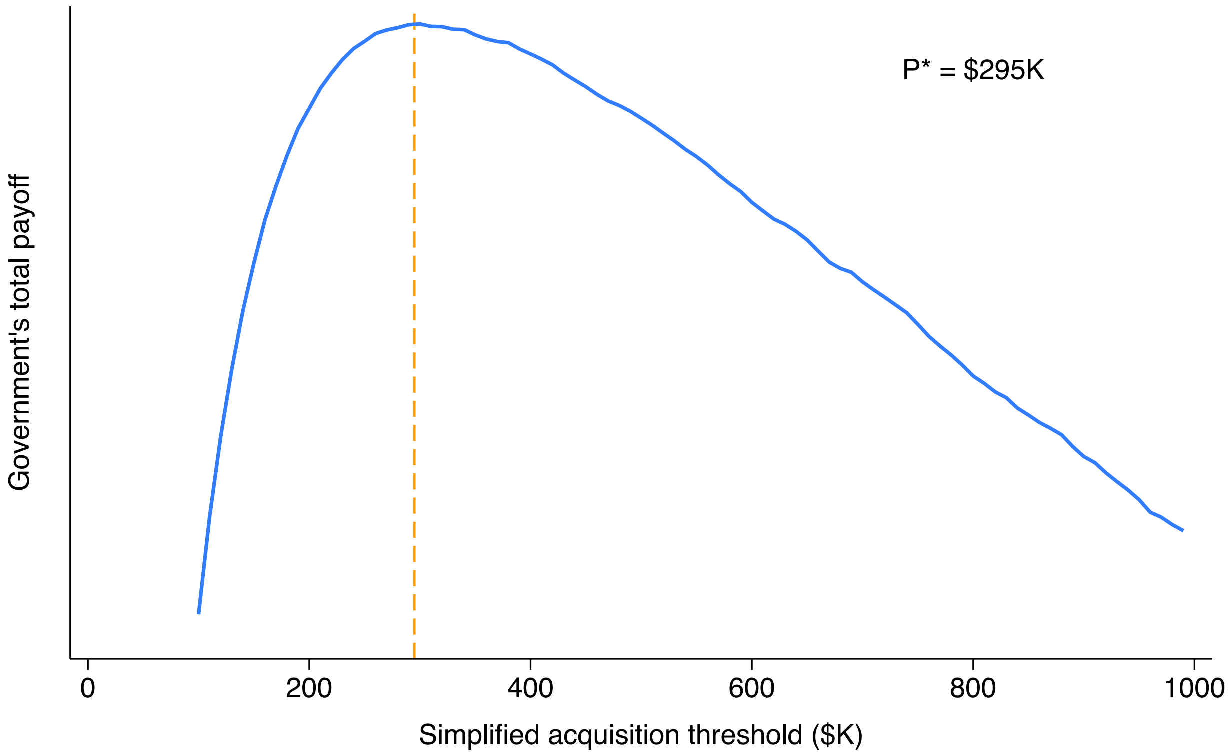
Notes: This figure shows the model fit outside the estimation sample (2012-2016). The estimation sample consists of non-R&D definitive contracts and purchase orders from the Federal Procurement Data System-Next Generation, with an award amount between $50,000 and $300,000. Both panels compare smoothed moments from 2012-2016 (blue solid line) with model-based simulated moments (orange dashed line). Since the model is estimated using data from 2006-2010, I refer to these figures as out-of-sample fit. The moments used in Panel (a) are the smoothed fraction of contracts for each $1,000-wide award bin between $50,000 and $300,000. The moments used in Panel (b) are the smoothed average contract quality index for each $5,000-wide award bin between $50,000 and $300,000. Moments are smoothed according to the procedure described in Appendix G. The vertical dashed line at $150,000 in both panels indicates the simplified acquisition threshold.

Figure 9: Counterfactual contract spending under different thresholds



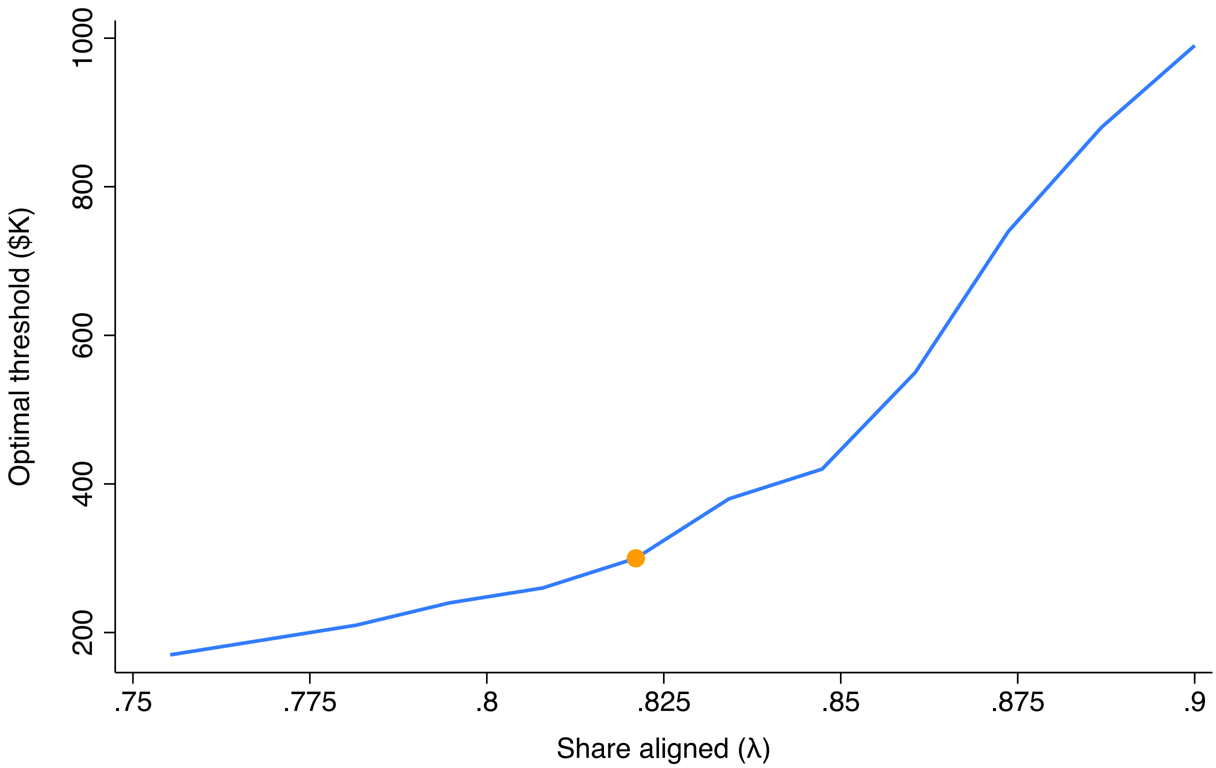
Notes: This figure shows estimated contract spending in contracts between $50,000 and $300,000 for different values of the simplified acquisition threshold. Each point in the blue line is computed by simulating a counterfactual, where all parameters remain fixed except for the value of the simplified acquisition threshold $(\bar{P})$. The model is estimated with the sample of non-R&D definitive contracts and purchase orders in 2006-2010 with an award amount between $50,000 and $300,000. The scale of the left axis is billions of current dollars per year, while the right axis is measured in percent, relative to the baseline case with a simplified threshold of $100,000. The orange dot depicts actual spending observed in 2006-2010, when the threshold was $100,000.

Figure 10: Optimal regulatory threshold



Notes: This figure shows the payoff obtained in the model by the principal, as a function of different simplified acquisition thresholds. Each point along the blue line is computed by simulating a counterfactual, where all parameters remain fixed except for the value of the simplified acquisition threshold $(\bar{P})$. The model is estimated with the sample of non-R&D definitive contracts and purchase orders in 2006-2010 with an award amount between $50,000 and $300,000. The vertical dashed line represents the value of the simplified acquisition threshold where the principal's payoff is maximized. The value of this threshold corresponds to $295,000. Since the model is estimated on 2006-2010 data, this threshold corresponds to roughly $345,000 in 2019 dollars.

Figure 11: Optimal threshold and a function of misalignment



Notes: This figure shows the optimal simplified acquisition threshold ($\bar{P}^*$) as a function of the share of aligned agents in the model (λ). Each point in the blue line is computed by finding the value of the threshold that maximizes the principal's payoff as in Figure 10 for a given level of the share aligned parameter λ. The model is estimated with the sample of non-R&D definitive contracts and purchase orders in 2006-2010 with an award amount between $50,000 and $300,000. The orange dot depicts the optimal threshold given the level of alignment that is estimated from the data $\lambda = 0.82$.

Tables

Table 1: Summary statistics of non-R&D stand-alone contracts, 2006-2016.

	Between \$5K and \$5M	Between \$50K and \$300K
Department		
DoD (all agencies)	0.5350	0.6076
DoD - Army	0.1279	0.1834
DoD - Navy	0.0978	0.1144
DoD - Air Force	0.0369	0.0603
DoD - Defense Logistics Agency	0.2598	0.2263
DoD - Other	0.0125	0.0231
Veteran Affairs	0.1234	0.0858
State	0.0634	0.0406
Justice	0.0520	0.0327
Agriculture	0.0383	0.0405
Solicitation and Award		
Competed	0.6682	0.6477
Fixed Price	0.9761	0.9779
Simplified Procedures	0.9137	0.7983
One offer	0.5057	0.5621
Last week of FY	0.0621	0.0886
Contract award (\$ K)	77.4	109.4
Number of offers received	2.9	2.9
Number of Contracts	4,268,746	664,038

Notes: This table presents summary statistics. The sample consists of non-R&D definitive contracts and purchase orders from the Federal Procurement Data System-Next Generation in fiscal years 2006 through 2016. An observation is a contract, defined by aggregating all contract *actions* (initial award, modification, termination, etc.) associated with the same contract ID. The left column presents means for each variable listed, considering all award amounts. The right column presents means restricting the sample to award amounts between \$50,000 and \$200.000.

Table 2: Distribution of contracts by amount brackets in FY2016

	Number of contracts (K)	Share of contracts (%)	Expenditure ($M)	Share of expenditure (%)
Between $5K & $25K	250.2	66.6	2,890	1.8
Between $25K & $50K	50.9	13.5	1,789	1.1
Between $50K & $300K	57.2	15.2	6,201	3.9
Between $300K & $1M	9.9	2.6	5,361	3.4
Between $1M & $5M	5.4	1.4	12,358	7.9
Above $5M	2.3	0.6	128,458	81.8
Total	376	100.0	157,057	100.0

Notes: This table presents the distribution of contract awards by amount brackets. The sample consists of non-R&D definitive contracts and purchase orders from the Federal Procurement Data System-Next Generation, in fiscal year 2016. An observation is a contract, defined by aggregating all contract *actions* (initial award, modification, termination, etc.) associated with the same contract ID. For each amount bracket, the table presents the number of contracts and expenditure, both in absolute levels and as a share of the total.

Table 3: Top Product and Service Categories between \$50K and \$300K in FY 2016

Code	Name	# Contracts
R	Support Services (Professional, Administrative, Management)	3927
59	Electrical and Electronic Equipment Compontents	3484
65	Medical, Dental and Veterinary Equipment and Supplies	3476
J	Maintenance, Repair and Rebuilding of Equipment	2984
66	Instruments And Laboratory Equipment	2751
70	ADP Equipment, Software and Supplies	2720
Z	Maintenance, Repair and Alteration of Real Property	2540
S	Utilities And Housekeeping	2101
D	ADP and Telecommunications	1824
16	Aircraft Components and Accessories	1733

Code	Name	# Contracts
6515	Medical And Surgical Instruments, Equipment, And Supplies	2585
7030	ADP Software	1343
6640	Laboratory Equipment And Supplies	1287
R499	Other Professional Services	952
1680	Miscellaneous Aircraft Accessories and Components	915
4820	Valves, Nonpowered	840
1560	Airframe Structural Components	814
5340	Hardware, Commercial	640
7110	Office Furniture	636
J065	Maintenance, Repair or Rebuilding of Medical Equipment	558

Notes: This table presents counts of contracts in the most common product and service categories. The sample consists of non-R&D definitive contracts and purchase orders from the Federal Procurement Data System-Next Generation in fiscal year 2016, and with award amounts between \$50,000 and \$300,000. An observation is a contract, defined by aggregating all contract actions (initial award, modification, termination, etc.) associated with the same contract ID. The bottom panel presents categories in 4-digit alphanumeric codes. The top panel aggregates these into 2-digit codes for products and 1-letter codes for services.

Table 4: Quantification of contracting responses to threshold change

	Data		
Estimation period	2012-2016	2008-2009	2015-2016
(Base period)	(2006-2010)	(2006-2007)	(2013-2014)
Threshold change	100K→150K	100K→100K	150K→150K
Estimates			
Missing mass ($\widehat{m}$)	9.58	-3.42	1.99
(s.e.)	(0.94)	(7.53)	(3.54)
Excess mass ($\widehat{x}$)	27.83	-4.46	-0.62
(s.e.)	(1.48)	(3.73)	(1.35)
Net excess mass ($\widehat{x} - \widehat{m}$)	18.24	-1.04	-2.61
(s.e.)	(2.31)	(10.38)	(4.80)
Aggregate implications (per year)			
Additional contracts	4,095	-	-
[as % of counterfactual]	[7.3%]		
Additional spending ($M)	526	-	-
[as % of counterfactual]	[8.7%]		

Notes: This table presents the estimates from the bunching analysis. The sample consists of non-R&D definitive contracts and purchase orders from the Federal Procurement Data System-Next Generation, with award amounts between $50,000 and $300,000. Missing and excess mass estimates are obtained by comparing observed award frequency distributions in different periods with a counterfactual distribution. This counterfactual is computed by adjusting a base period distribution for predicted frequency changes, using an interpolation method based on contracts below $50,000 and above $300,000. These estimates correspond to the missing and excess mass areas observed in Figure 4 and Figure 5, normalized by the average number of contracts between $50,000 and $300,000. The formula for calculating these areas is given in Equation (2). The first column shows the main estimates (that correspond to Figure 4), while the second and third columns show the placebo exercises (shown in Figure 5). Standard errors are computed via bootstrap, sampling contract observations in the sample with replacement. The lower panel shows the additional number of contracts and contract spending per year implied by the net excess mass estimates. This is presented both in absolute levels and as a share of the estimated counterfactual distribution.

Table 5: Estimates of changes in quality by award range

	(1) Quality Index	(2) No. of mods	(3) Cost overruns	(4) Delays	(5) Terminated	(6) Last Week
$\in [\$75K, \$100K) \times$ Post	-0.0061*	0.0374***	0.7215*	0.2966	0.0111	0.0017
	(0.0034)	(0.0124)	(0.4041)	(1.6289)	(0.0585)	(0.1940)
$\in [\$100K, \$150K) \times$ Post	0.0275***	-0.0728***	-2.7578**	-5.6948**	-0.0987	-0.8837***
	(0.0084)	(0.0232)	(1.1521)	(2.7415)	(0.0657)	(0.2271)
$\in [\$150K, \$300K) \times$ Post	-0.0026	0.0255	-2.2920	3.9720	0.0485	0.4619
	(0.0159)	(0.0496)	(2.6441)	(4.9177)	(0.1272)	(0.3265)
N	578,765	578,765	578,765	578,765	578,765	578,765
R^2	0.2121	0.2460	0.0916	0.2084	0.0539	0.1052
Mean D.V.	-0.0098	0.9767	8.6098	81.2586	1.3159	8.8575

Standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Notes: This table presents coefficient estimates from a series of regressions. The sample consists of non-R&D definitive contracts and purchase orders from the Federal Procurement Data System-Next Generation, in fiscal years 2006 through 2016, with award amounts between $50,000 and $300,000. An observation is a contract, defined by aggregating all contract *actions* (initial award, modification, termination, etc.) associated with the same contract ID. Each column represents a separate regression, where the dependent variable is indicated in the column header. The regressors include a full set of 4-digit PSC codes fixed effects, awarding agency fixed effects, fiscal year fixed effects, and $1,000-wide award bin fixed effects. The Post indicator is equal to 0 in 2010 or earlier and equal to 1 in 2012 or later. Fiscal year 2011 is excluded from the regression. Standard errors are clustered by awarding agency.

Table 6: Model Estimates

Parameter	Estimate (s.e.)
Mean valuation (μ_v)	-0.1653 (0.0043)
S.d. of valuation (σ_v)	3.0341 (0.0027)
Mean red tape cost (μ_κ)	12.7643 (0.0286)
S.d. of red tape cost (σ_κ)	10.2497 (0.0262)
Govt's bargaining weight (ϕ)	0.4024 (0.0055)
Share aligned (λ)	0.8192 (0.0030)
Govt's share of red tape costs (γ)	0.2572 (0.0055)
Mean aligned implementation shock (Δ_A)	0.0778 (0.0008)
Mean misaligned implementation shock (Δ_M)	-0.0265 (0.0002)
Misaligned reservation payoff ($\bar{b}$)	-0.6380 (0.0048)
Adjustment frictions (η)	0.8790 (0.0055)

Notes: This table presents model parameter estimates obtained via simulated method of moments. Standard errors are in parentheses and calculated using the asymptotic formula in (8). The estimation sample consists of non-R&D definitive contracts and purchase orders from the Federal Procurement Data System-Next Generation, with an award amount between \$50,000 and \$300,000. The moments used in the estimation correspond to the smoothed fraction of contracts for each \$1,000-wide award bin between \$50,000 and \$300,000, and the smoothed average contract quality index for each \$5,000-wide award bin between \$50,000 and \$300,000. The procedures to smooth the estimation moments and to construct the weighting matrix are described in Appendix G.

Appendix (For Online Publication)

Contents:

A. Additional Figures

B. Additional Tables

C. Office-Level Evidence on Number of Contracts (DDD)

D. Sensitivity to Bunching Estimation Parameters

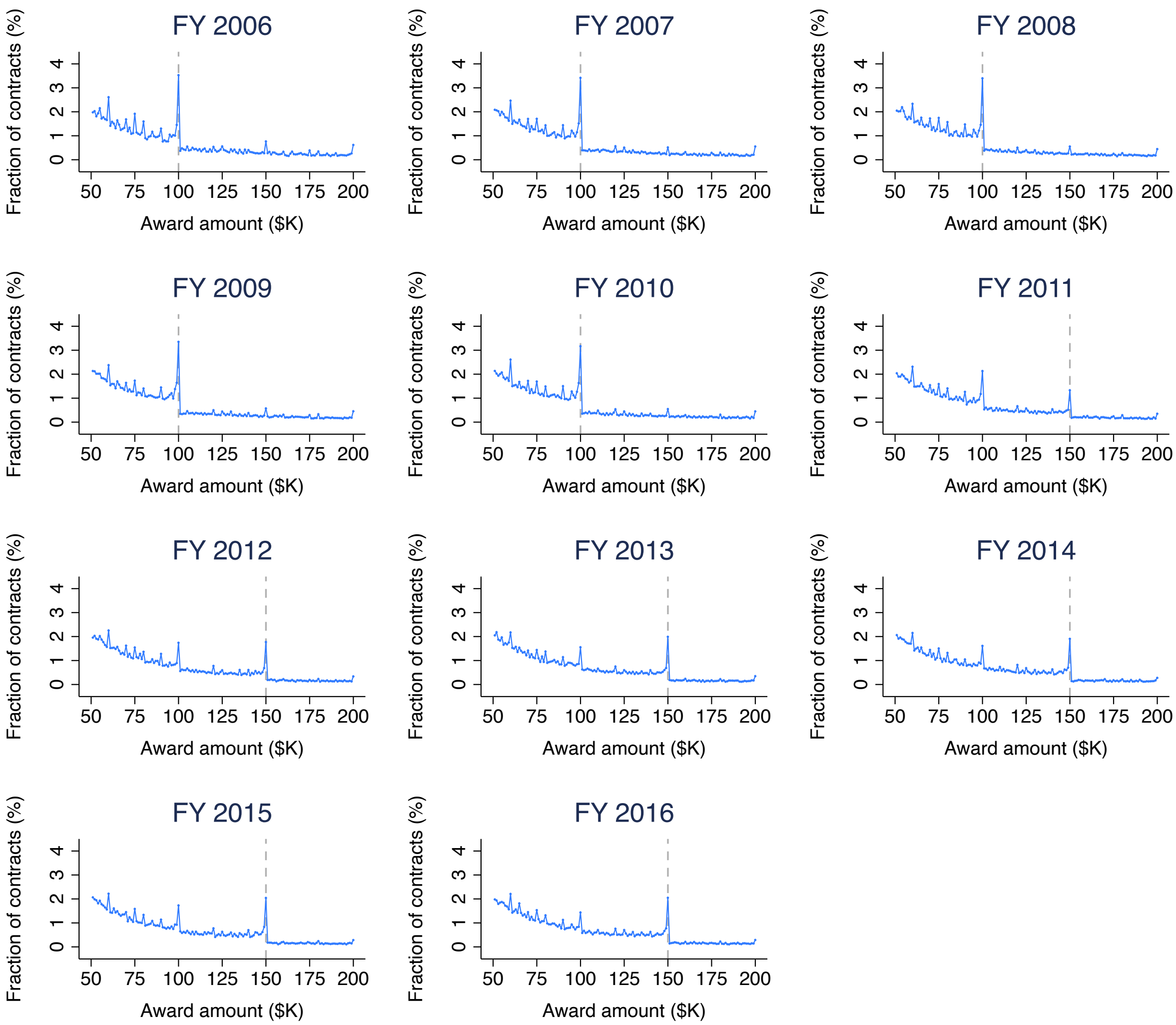
E. Validating Quality Measures Using IT Dashboard

F. Model Details

G. Model Estimation Details

A. Additional Figures

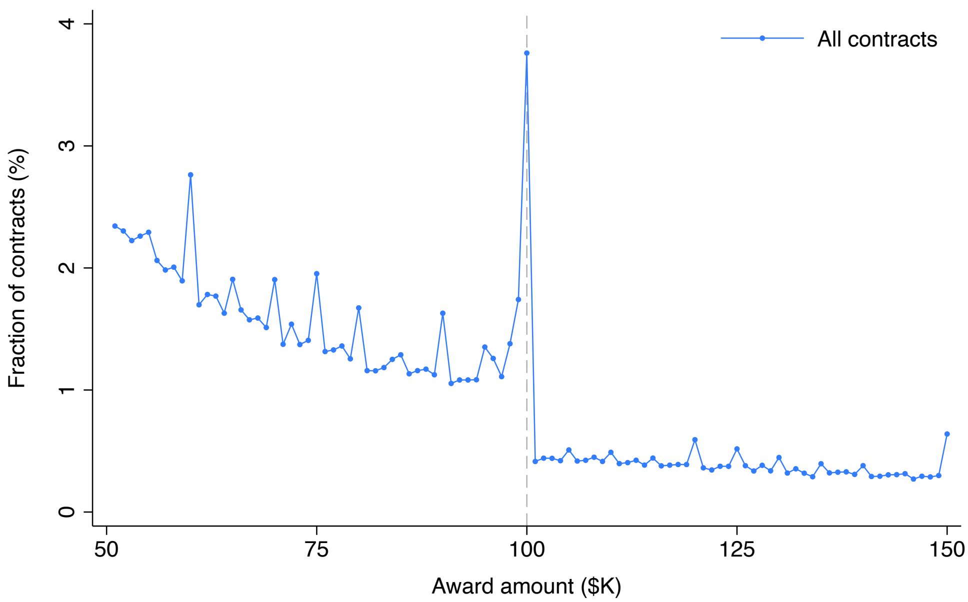
Figure A1: Year-by-year distributions of federal contracts



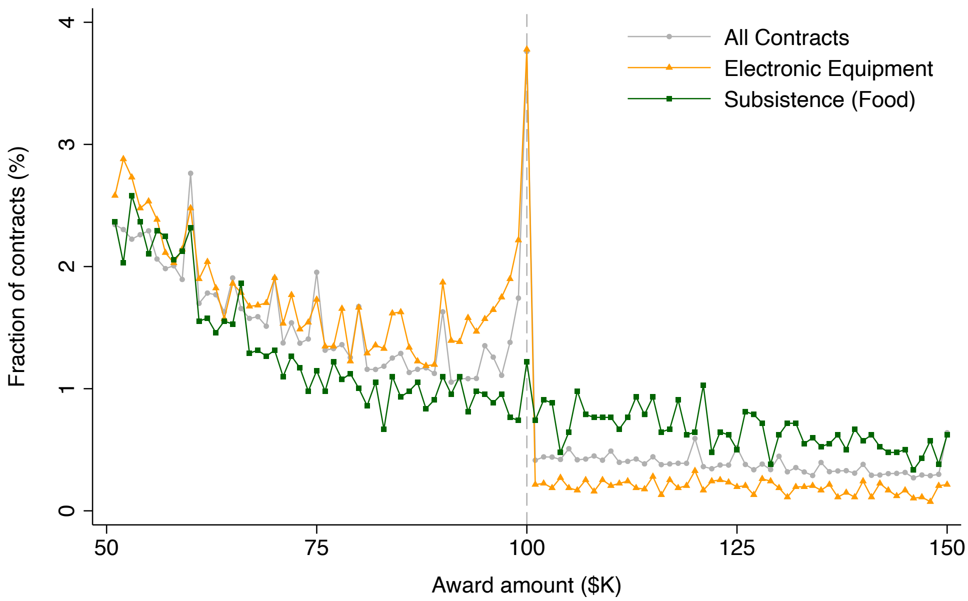
Notes: This figure shows contract awards frequency distributions in each fiscal year between 2006 and 2016. The sample is non-R&D definitive contracts and purchase orders from the Federal Procurement Data System-Next Generation, with an award amount between \$50,000 and \$200,000. Award amounts are discretized into right-inclusive bins of one-thousand dollars length. A bin labeled "X" includes all contract awards in the range $(\$1000(X-1), \$1000X]$. Vertical dashed lines indicate the location of the simplified acquisition threshold in each period. Below the threshold, contracts can be awarded using (high-discretion) simplified acquisition procedures, whereas above the threshold non-exempt awards are subject to every acquisition law in the Federal Acquisition Regulation.

Figure A2: Distribution of contract awards by product categories, 2006-2010

(a) All contracts



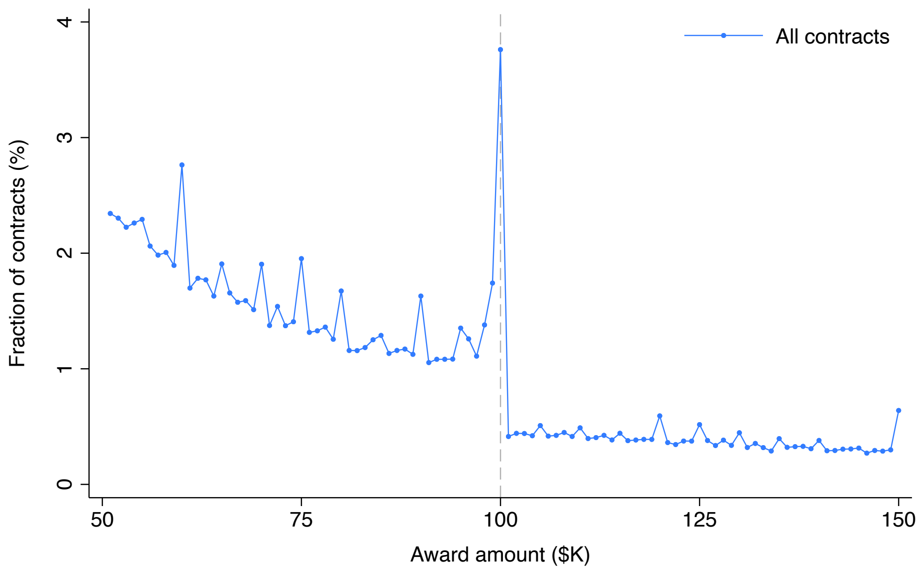
(b) Selected product categories



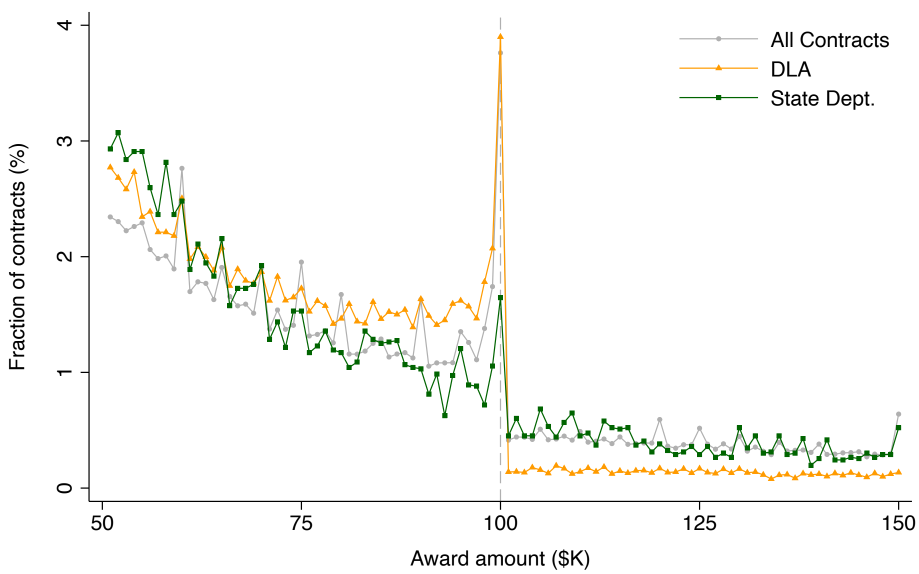
Notes: This figure shows contract awards frequency distributions in 2006-2010, by selected product categories. The sample is non-R&D definitive contracts and purchase orders from the Federal Procurement Data System-Next Generation, with an award amount between \$50,000 and \$150,000. Panel (a) features the distribution of all awards in the sample. In panel (b), this distribution is compared to selected categories: electrical and electronic equipment components (federal product and service code PSC 59), and subsistence (food, federal product and service code PSC 89). Award amounts are discretized into right-inclusive bins of one-thousand dollars length. A bin labeled "X" includes all contract awards in the range $(\$1000(X-1), \$1000X]$. Vertical dashed lines indicate the location of the simplified acquisition threshold in each period. Below the threshold, contracts can be awarded using (high-discretion) simplified acquisition procedures, whereas above the threshold non-exempt awards are subject to every acquisition law in the Federal Acquisition Regulation.

Figure A3: Distribution of contract awards by awarding agency, 2006-2010

(a) All contracts



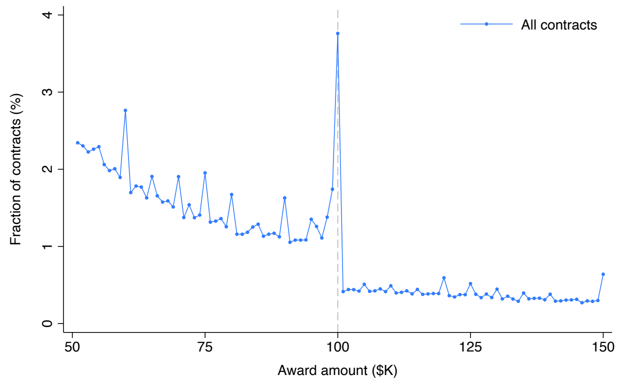
(b) Selected agencies



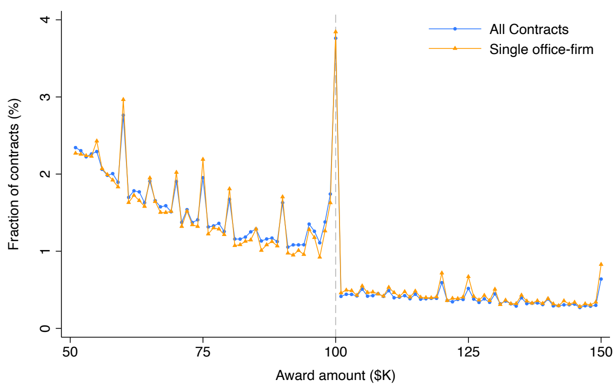
Notes: This figure shows contract awards frequency distributions in 2006-2010, by selected awarding agencies. The sample is non-R&D definitive contracts and purchase orders from the Federal Procurement Data System-Next Generation, with an award amount between $50,000 and $150,000. Panel (a) features the distribution of all awards in the sample. In panel (b), this distribution is compared to selected awarding agencies: the Defense Logistics Agency (DLA), and the State Department. Award amounts are discretized into right-inclusive bins of one-thousand dollars length. A bin labeled "X" includes all contract awards in the range $(\$1000(X-1), \$1000X]$. Vertical dashed lines indicate the location of the simplified acquisition threshold in each period. Below the threshold, contracts can be awarded using (high-discretion) simplified acquisition procedures, whereas above the threshold non-exempt awards are subject to every acquisition law in the Federal Acquisition Regulation.

Figure A4: Distribution by contract awards in selected samples, 2006-2016

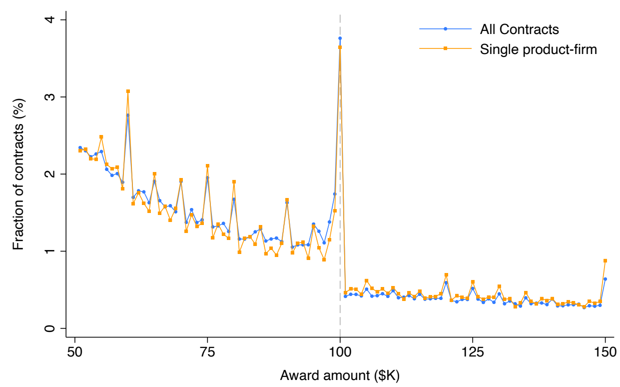
(a) All contracts in FY2006-2016



(b) Versus single-office-firm contacts



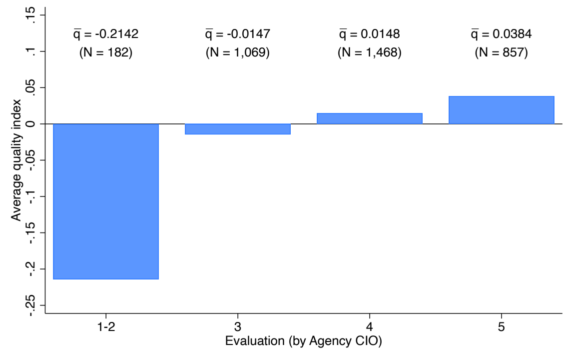
(c) Versus single-office-product contacts



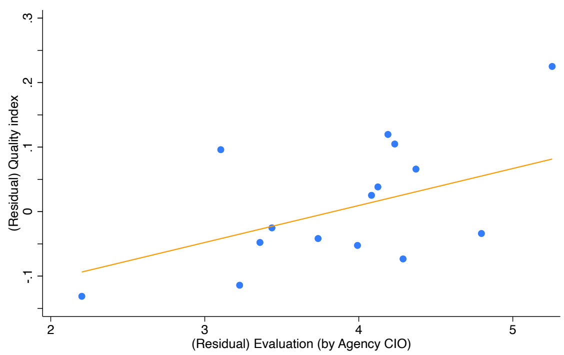
Notes: This figure shows contract awards frequency distributions in 2006-2010, in selected samples. The full sample is non-R&D definitive contracts and purchase orders from the Federal Procurement Data System-Next Generation, with an award amount between $50,000 and $150,000. Panel (a) features the distribution of all awards in the sample. In panel (b), this distribution is compared to the restricted sample of office-firm pairs with a single award within the fiscal year. In panel (c), the full distribution is compared to the restricted sample of office-product pairs with a single award within the fiscal year. Products are defined at the 4-digit product or service codes (PSC). Award amounts are discretized into right-inclusive bins of one-thousand dollars length. A bin labeled "X" includes all contract awards in the range $(\$1000(X-1), \$1000X]$. Vertical dashed lines indicate the location of the simplified acquisition threshold. Below the threshold, contracts can be awarded using (high-discretion) simplified acquisition procedures, whereas above the threshold non-exempt awards are subject to every acquisition law in the Federal Acquisition Regulation. The goal of the figure is to show that split purchases are an unlikely key driver of the bunching patterns. Bunching is observed with the same strength even among contracting offices that have a single transaction with a firm over the budget cycle, or that have a single purchase for a particular product during this period.

Figure A5: Validating quality measures

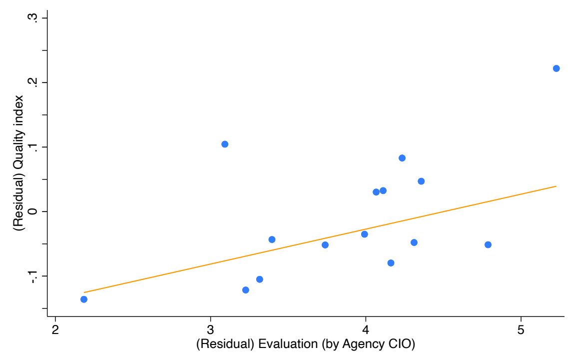
(a) Mean quality index by CIO evaluation



(b) Controlling for agency



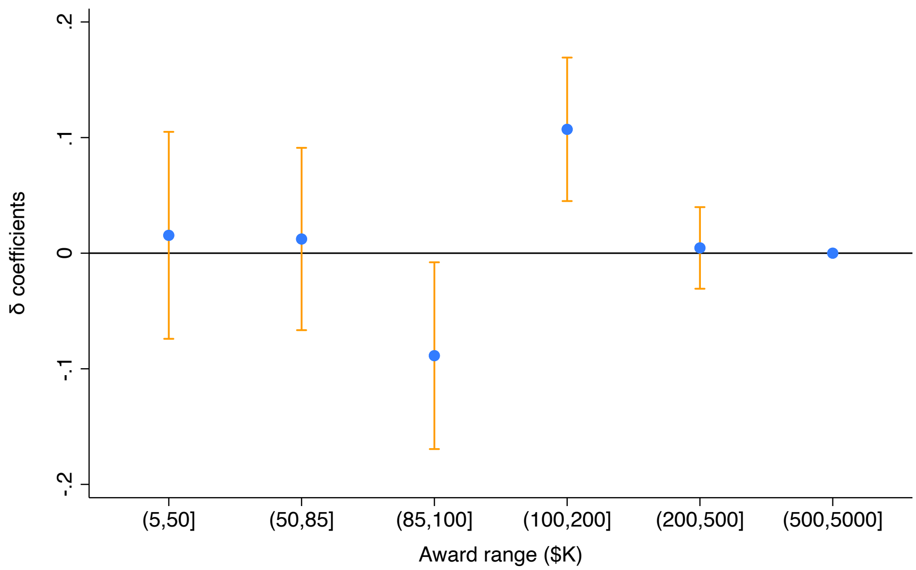
(c) Weighting by the size of the project



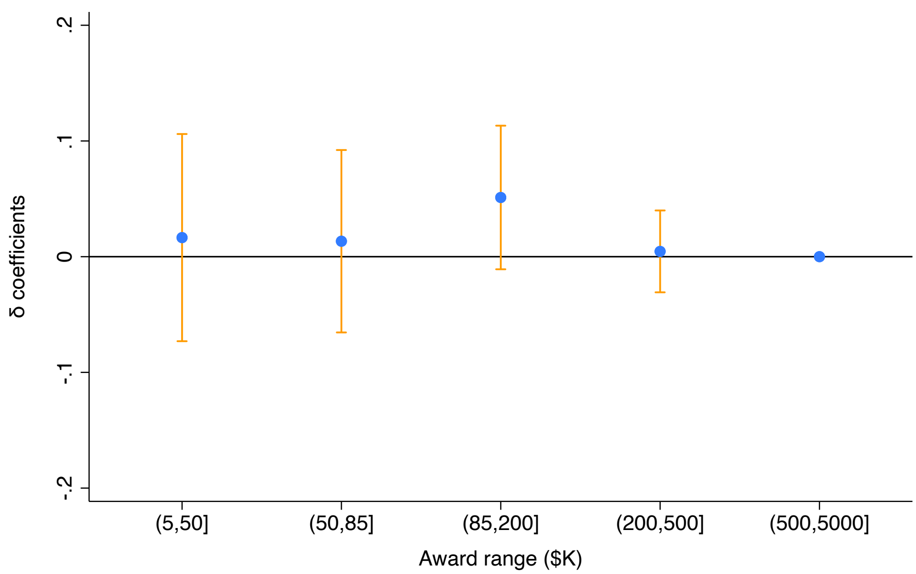
Notes: This figure shows the relationship between the contract execution quality proxies used in the main analysis and the main quality measure available in the IT Dashboard: Chief Information Officers (CIOs) evaluations. The sample consists of federal IT projects in the IT Dashboard (`itdashboard.gov`) annual files 2013-2018 that can be matched to contracts in our main analysis sample, that is, non-R&D definitive contracts and purchase orders from the Federal Procurement Data System-Next Generation in fiscal years 2006-2016. Panel (a) presents the unconditional mean of the quality index computed from FPDS (see Section 3), by CIO evaluation rating. Panel (b) presents a binned scatter plot of the same two variables, controlling for a set of agency or department fixed-effects. Panel (c) presents the same exercise from (b), but weighting observations by the natural logarithm of the project's total contract spending in the sample.

Figure A6: Change in number of contracts in high vs. low bunching offices, by award size

(a) Baseline



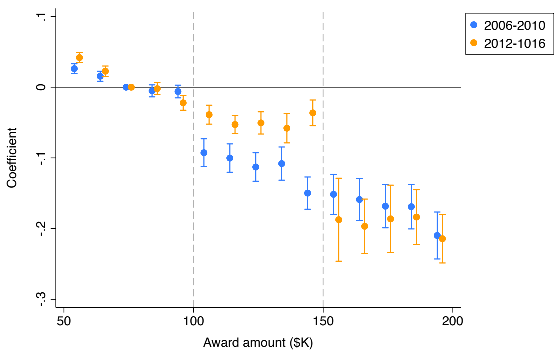
(b) Combining award ranges around the simplified acquisition threshold



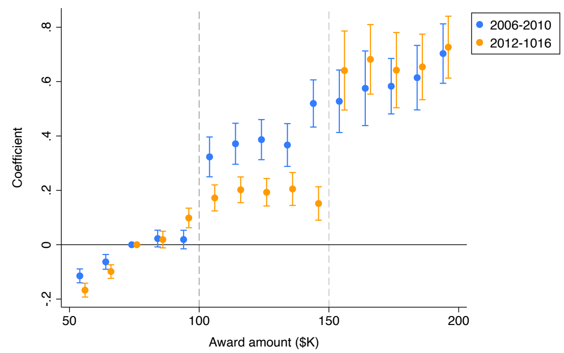
Notes: This figure presents coefficient estimates and their 95% confidence intervals for two triple-difference (DDD) specifications. The sample is non-R&D definitive contracts and purchase orders from the Federal Procurement Data System-Next Generation, in fiscal years 2006 through 2016, with award amounts between $5,000 and $5,000,000. This sample is aggregated at the contracting office - fiscal year - $1,000-wide award amount bin. The dependent variable is number of contracts by office-year-bin. The regressors include three full sets of fixed effects by office, fiscal year and award amount bin. The plotted estimates are the coefficients of the triple interaction between a High-bunching indicator, a Post indicator, and an indicator for ranges of award amounts. The High-bunching indicator is equal to one for offices with above-median pre-2011 bunching, as defined by the share of total awards with amounts between $99,000 and $100,000. The Post indicator is equal to 0 in 2010 or earlier and equal to 1 in 2012 or later. Fiscal year 2011 is excluded from the regression. The award amount ranges are indicated on the horizontal axes. Standard errors are clustered by contracting office. More details can be found in Appendix C.

Figure A7: Contract quality by each performance proxy

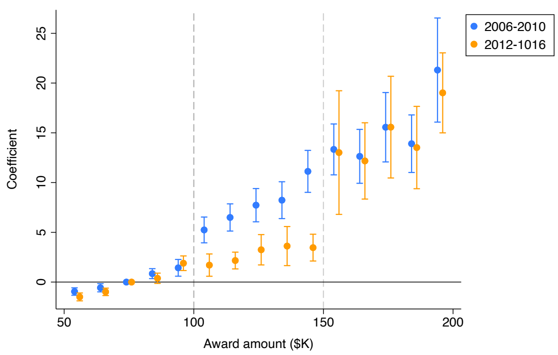
(a) Quality index



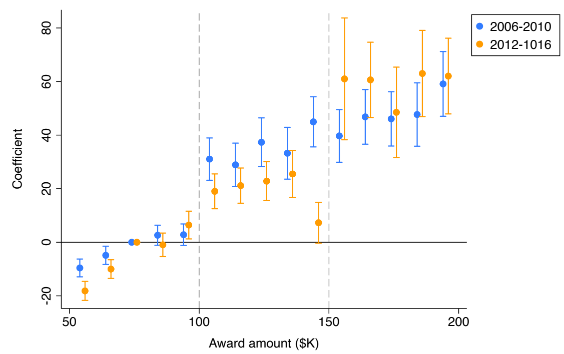
(b) Number of modifications



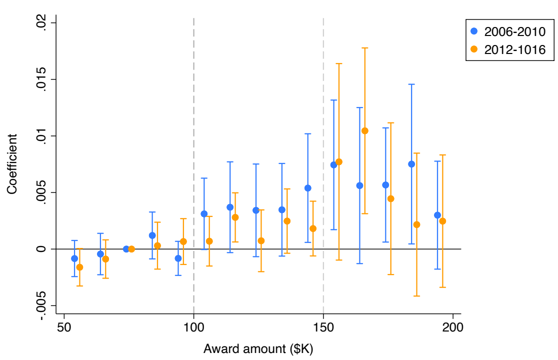
(c) Cost overruns



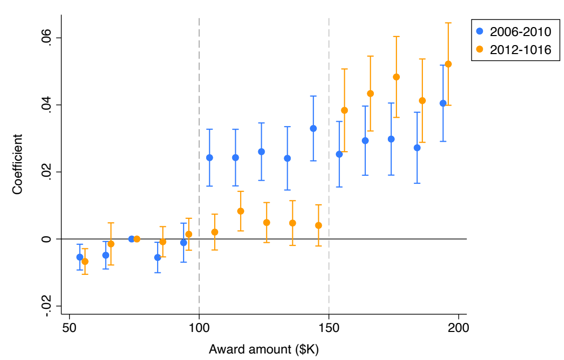
(d) Delays



(e) Termination

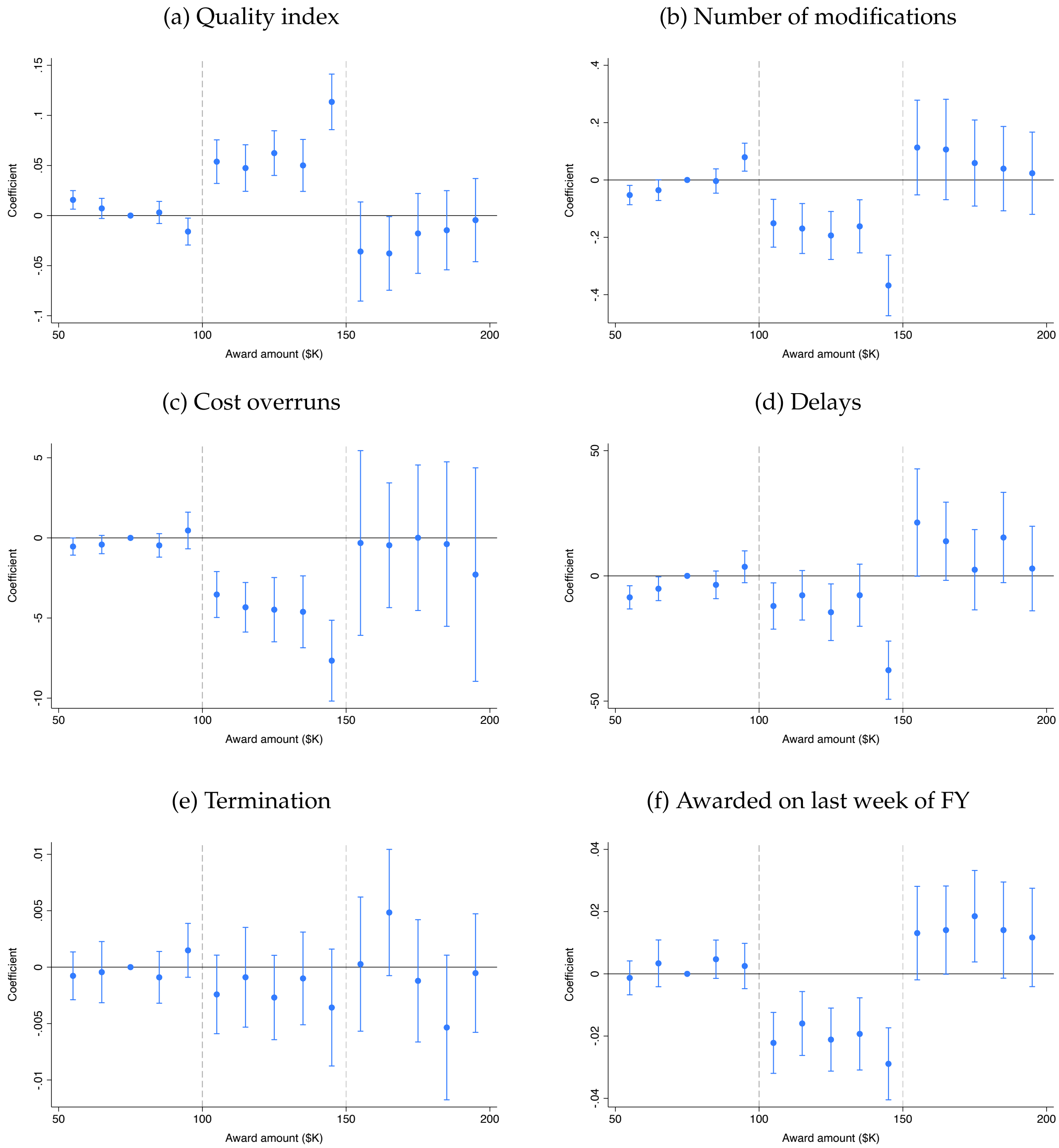


(f) Awarded on last week of FY



Notes: This figure shows period-specific regression coefficients of specification (3) using different measures of contract performance as dependent variables and no controls. Coefficients can be interpret as normalized average outcomes by period and by bin. The sample is non-R&D definitive contracts and purchase orders from the Federal Procurement Data System-Next Generation, with an award amount between \$50,000 and \$300,000. The dependent variable in each panel is as follows: panel (a) shows the number of within-scope modifications, panel (b) shows cost overruns (difference, in dollars, between expected obligations at the time of award and total ex-post obligations), panel (c) shows delays (difference, in days, between expected completion day and actual completion day), panel (d) shows the fraction of terminated contracts, and panel (e) shows the fraction of contracts awarded in the last week of the fiscal year. Award amounts are discretized into right-inclusive bins of ten-thousand dollars length. Award amounts are discretized into right-inclusive bins of ten-thousand dollars length. The vertical dashed lines indicate the simplified acquisition threshold in each period. These correspond to $\$100,000$ in the pre-period (2006-2010) and $\$150,000$ in the post-period (2012-1016).

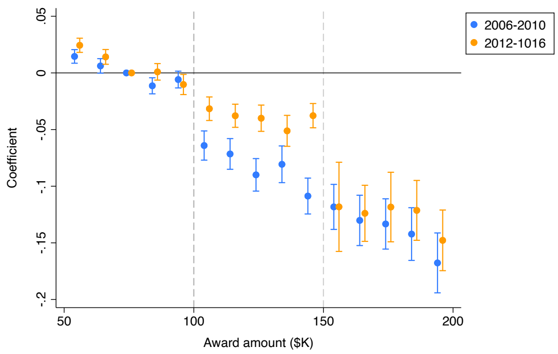
Figure A8: Change in contract quality by each performance proxy



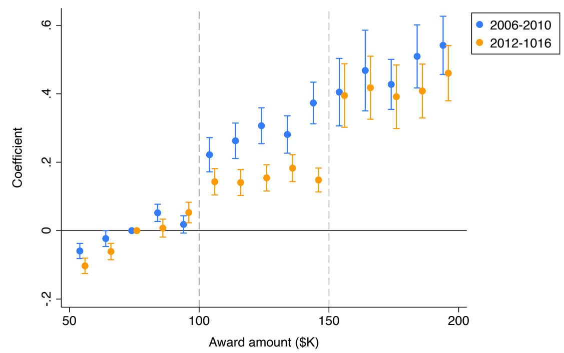
Notes: This figure shows regression coefficients from specification (4) using different measures of contract performance as dependent variables and no controls. Coefficients can be interpreted as the change in the average performance measure between the pre and post periods, by bin of award amount. The sample is non-R&D definitive contracts and purchase orders from the Federal Procurement Data System-Next Generation, with an award amount between $50,000 and $300,000. The dependent variable in each panel is as follows: panel (a) shows the number of within-scope modifications, panel (b) shows cost overruns (difference, in dollars, between expected obligations at the time of award and total ex-post obligations), panel (c) shows delays (difference, in days, between expected completion day and actual completion day), panel (d) shows the fraction of terminated contracts, and panel (e) shows the fraction of contracts awarded in the last week of the fiscal year. Award amounts are discretized into right-inclusive bins of ten-thousand dollars length. Award amounts are discretized into right-inclusive bins of ten-thousand dollars length. The vertical dashed lines indicate the simplified acquisition threshold in each period. These correspond to $100,000 in the pre-period (2006-2010) and $150,000 in the post-period (2012-1016).

Figure A9: Contract quality by each performance proxies, with controls

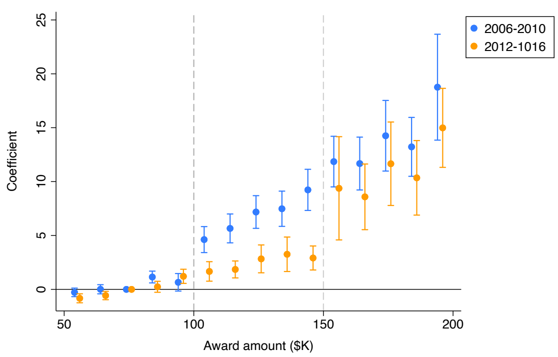
(a) Quality index



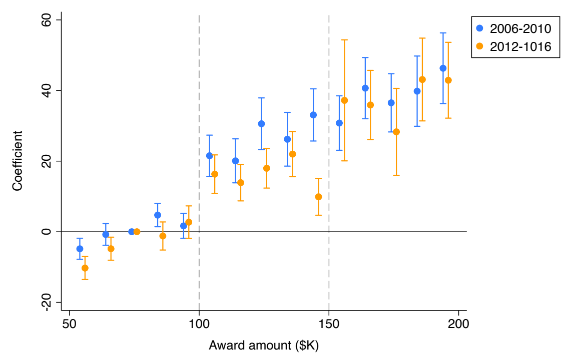
(b) Number of modifications



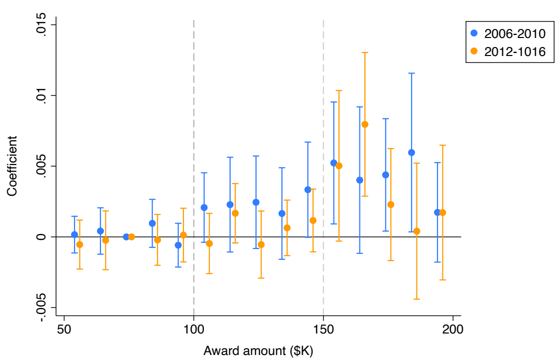
(c) Cost overruns



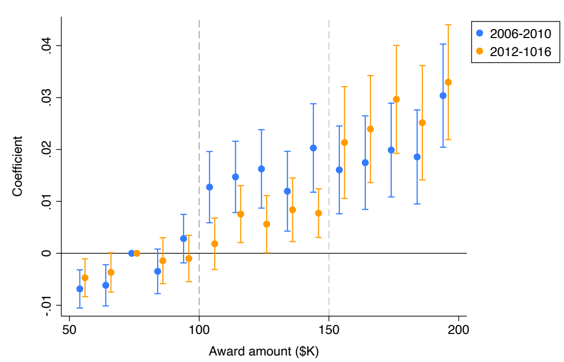
(d) Delays



(e) Termination



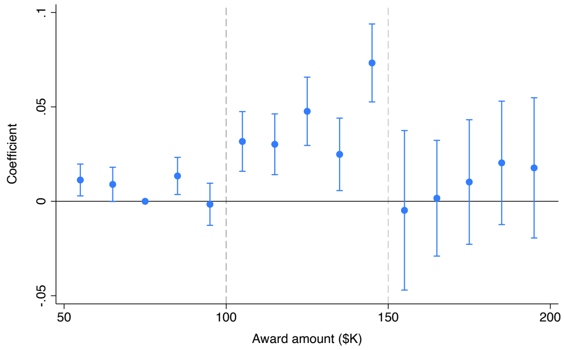
(f) Awarded on last week of FY



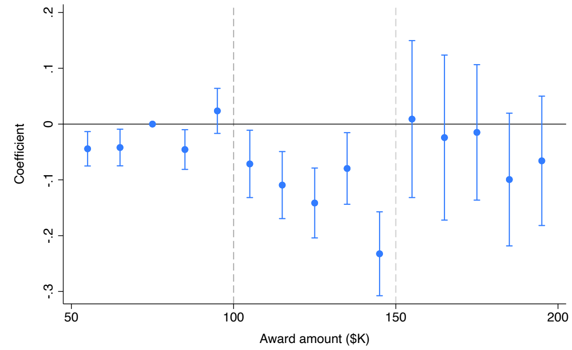
Notes: This figure replicates Figure A7 with the addition of controls. It shows period-specific regression coefficients of specification (3) using different measures of contract performance as dependent variables and controlling for a full set of product code and awarding agency fixed-effects. Coefficients can be interpret as (conditional) normalized average outcomes by period and by bin. The sample is non-R&D definitive contracts and purchase orders from the Federal Procurement Data System-Next Generation, with an award amount between $50,000 and $300,000. The dependent variable in each panel is as follows: panel (a) shows the number of within-scope modifications, panel (b) shows cost overruns (difference, in dollars, between expected obligations at the time of award and total ex-post obligations), panel (c) shows delays (difference, in days, between expected completion day and actual completion day), panel (d) shows the fraction of terminated contracts, and panel (e) shows the fraction of contracts awarded in the last week of the fiscal year. Award amounts are discretized into right-inclusive bins of ten-thousand dollars length. Award amounts are discretized into right-inclusive bins of ten-thousand dollars length. The vertical dashed lines indicate the simplified acquisition threshold in each period. These correspond to $100,000 in the pre-period (2006-2010) and $150,000 in the post-period (2012-1016).

Figure A10: Change in contract quality by each performance proxy, with controls

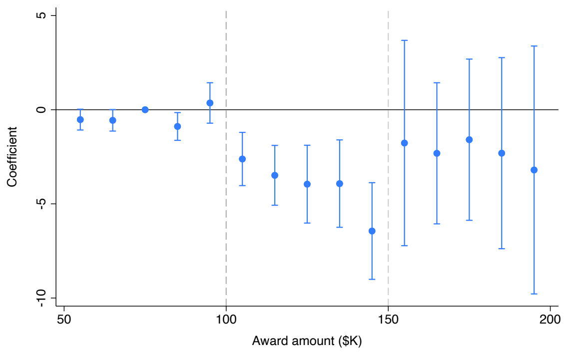
(a) Quality index



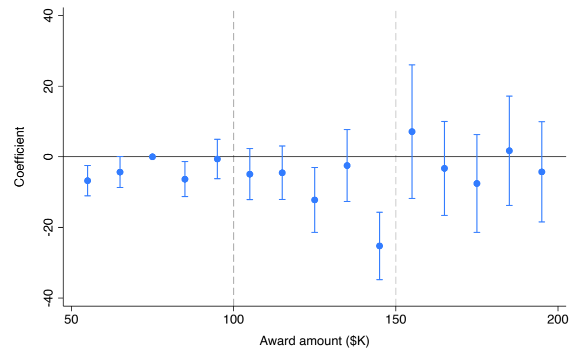
(b) Number of modifications



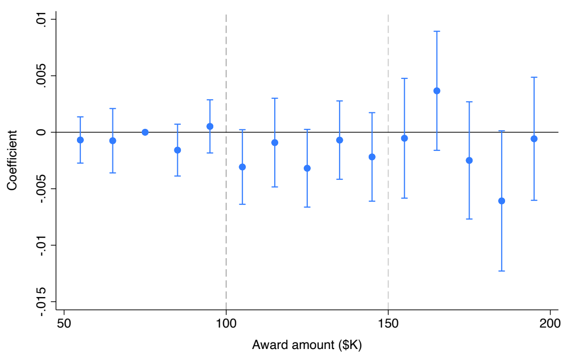
(c) Cost overruns



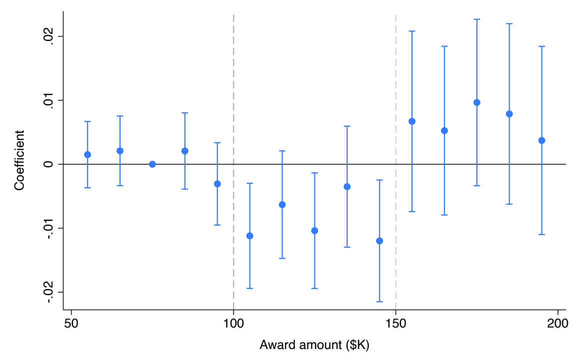
(d) Delays



(e) Termination

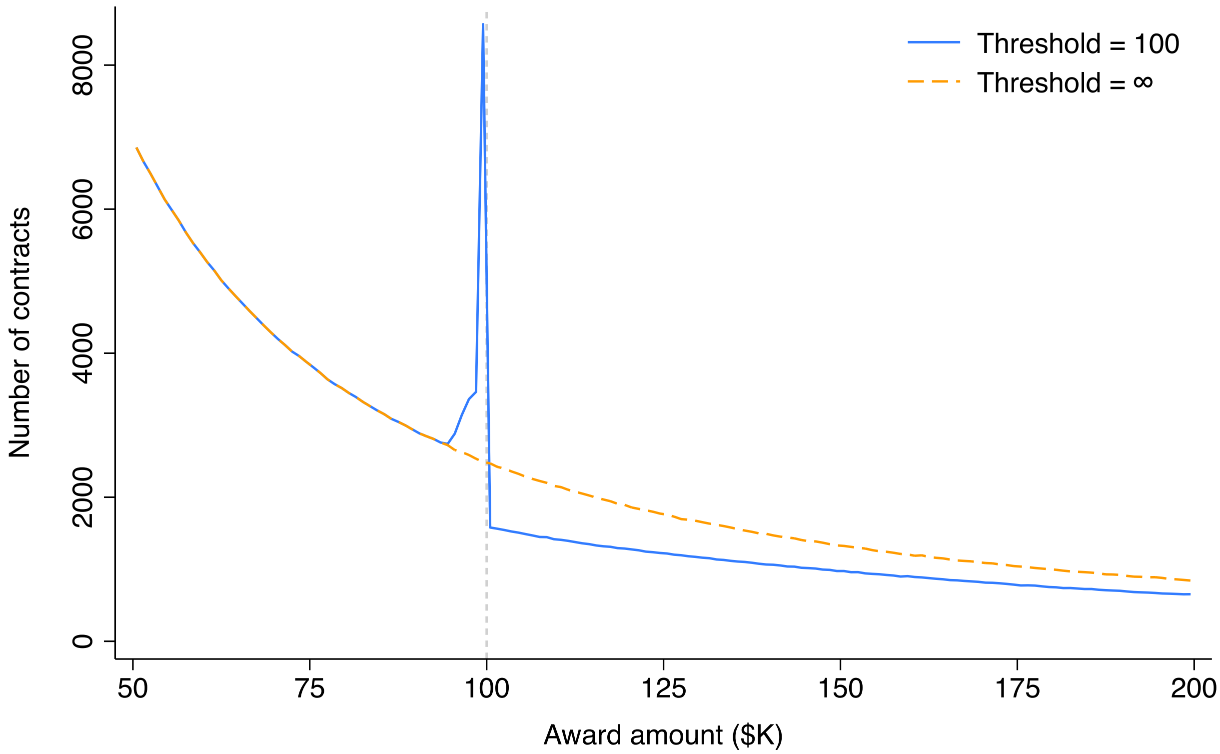


(f) Awarded on last week of FY



Notes: This figure replicates Figure A8 with the addition of controls. It shows regression coefficients from specification (4) using different measures of contract performance as dependent variables and controlling for a full set of product code and awarding agency fixed-effects. Coefficients can be interpreted as the (conditional) change in the average performance measure between the pre and post periods, by bin of award amount. The sample is non-R&D definitive contracts and purchase orders from the Federal Procurement Data System-Next Generation, with an award amount between $50,000 and $300,000. The dependent variable in each panel is as follows: panel (a) shows the number of within-scope modifications, panel (b) shows cost overruns (difference, in dollars, between expected obligations at the time of award and total ex-post obligations), panel (c) shows delays (difference, in days, between expected completion day and actual completion day), panel (d) shows the fraction of terminated contracts, and panel (e) shows the fraction of contracts awarded in the last week of the fiscal year. Award amounts are discretized into right-inclusive bins of ten-thousand dollars length. Award amounts are discretized into right-inclusive bins of ten-thousand dollars length. The vertical dashed lines indicate the simplified acquisition threshold in each period. These correspond to $100,000 in the pre-period (2006-2010) and $150,000 in the post-period (2012-1016).

Figure A11: Contract distribution with simplified acquisition for all purchases



Notes: This figure shows two simulated contract award distributions using the estimated model. The solid blue line shows the baseline case with a simplified acquisition threshold of $100,000. The dashed orange line presents a counterfactual where the threshold is assumed to be infinity. This means that all purchases are allowed to use simplified procedures, regardless of the size of award. The counterfactual assumes that all estimated parameters remain fixed, except for the value of the simplified acquisition threshold ($\bar{P}$). The model is estimated with the sample of non-R&D definitive contracts and purchase orders in 2006-2010 with an award amount between $50,000 and $300,000.

B. Additional Tables

Table B1: Sensitivity of net excess mass estimates $(\hat{m} - \hat{x})$

Panel (a): $p = 2$

		$\overline{R}$				
		200	250	300	350	400
	30	24.91	18.41	19.56	15.44	9.87
	40	20.63	15.15	15.81	12.49	7.98
$\underline{R}$	50	20.13	15.81	16.84	14.53	11.07
	60	20.13	16.65	18.02	16.41	13.66
	70	20.93	18.18	19.99	19.04	16.91

Panel (b): $p = 3$

		$\overline{R}$				
		200	250	300	350	400
	30	14.71	10.68	13.70	13.72	11.97
	40	14.44	11.37	14.45	14.66	12.88
$\underline{R}$	50	16.60	14.51	18.24	19.10	17.77
	60	18.37	16.95	21.19	22.53	21.51
	70	20.41	19.60	24.39	26.23	25.60

Panel (c): $p = 4$

		$\overline{R}$				
		200	250	300	350	400
	30	15.99	15.17	27.32	41.39	64.45
	40	17.90	17.44	30.29	44.89	67.64
$\underline{R}$	50	21.29	21.51	35.46	50.82	72.93
	60	23.66	24.34	38.68	53.61	72.84
	70	25.95	27.08	41.56	55.69	71.69

Notes: This table presents a series of net excess mass estimates $(\hat{m} - \hat{x})$, as a function of key parameters of the bunching estimation procedure. The estimation procedure is described in Section 3.2. Each of the numbers in the table is an estimate of $(\hat{m} - \hat{x})$ as a function of the degree of the polynomial p used to fit Equation (1), and the lower and upper bounds of the excluded region $[\underline{R}, \overline{R}]$. Both bounds are measured in thousand dollars. The baseline specification considers values of $p = 3$, $\underline{R} = 50$, and $\overline{R} = 300$. The estimate for this baseline case (= 18.76) is presented in bold at the center of panel (b), and coincides with the estimate presented in Table 4.

Table B2: Summary statistics of IT dashboard - FPDS sample

(a) Number of projects, contracts, and total obligated by agency

Department or Agency	Number of IT Projects	Number of Contracts	Total Obligated ($M)
Agriculture	36	220	1,739.3
Commerce	19	448	1,127
Defense	38	287	3,004.4
Health and Human Services	57	232	4,974.1
Interior	23	78	261.9
State	10	123	9,863.1
Treasury	15	22	261.6
General Services Administration	13	38	1,540.5
Homeland Security	69	856	8,369.7
Veterans Affairs	24	514	6,686
All Others	86	758	4,773.6
Total	390	3576	42,601.2

(b) Summary statistics of quality measures

Quality measure	Mean	s.d.	p10	p90
Contract-level (Source: FPDS; N=3576)				
Number of modifications	5.47	9.92	0	14
Cost overruns ($M)	4.72	21.30	0	8.38
Delays (days)	260	438	0	893
Termination (0-1)	0.007	0.083	0	0
Last week of FY (0-1)	0.12	0.32	0	1
Quality index	0.00	0.55	-0.52	0.65
Project-level (Source: IT Dashboard; N=390)				
CIO evaluation (1-5)	3.83	0.86	3	5

Notes: This table presents summary statistics on a sample of federal IT projects that can be matched to contracts in our main analysis sample. The sample of IT projects comes from the IT Dashboard (`itdashboard.gov`) annual files in fiscal years 2013 through 2018. The sample of contracts consists of non-R&D definitive contracts and purchase orders from the Federal Procurement Data System-Next Generation in fiscal years 2006-2016. Panel (a) presents counts and spending totals (computed as total obligated dollars in the sample) by department or agency. Panel (b) presents mean, standard deviation, and percentiles 10 and 90, for the main quality measures. The top part presents statistics at the contract level for the execution-based quality proxies, while the lower part shows statistics at the IT project level for the Chief Information Officer's evaluations.

C. Office-Level Evidence on Number of Contracts

This appendix describes the triple-difference (DDD) analysis discussed at the end of Section 3.1. The sample of analysis is non-R&D definitive contracts and purchase orders from the Federal Procurement Data System-Next Generation, with an award amount between \$5,000 and \$5,000,000.

The goal of this empirical strategy is to analyze how the *number of awards* that contracting offices made responded to the 2011 change in the simplified acquisition threshold. Under a null hypothesis that the bunching observed in Figure 1 is entirely explained by split purchases, then the number of awards should have —all else equal— *decreased* following the increase in the simplified acquisition threshold. The proposed DDD specification leverages variation across three dimensions: before and after the threshold, across contracting offices, and across contract award sizes.

Formally, I estimate the following specification:

$$n_{kbt} = \alpha_k + \alpha_b + \alpha_t + \sum_{q \in Q} \beta_q \cdot \mathbf{1}[b \in q] \times HighBunch_k + \sum_{q \in Q} \gamma_q \cdot \mathbf{1}[b \in q] \times Post_t$$
$$+ \sum_{q \in Q} \delta_s \cdot \mathbf{1}[b \in q] \times HighBunch_k \times Post_t + \epsilon_{kbt} \quad (9)$$

where n_{kbt} be the number of awards made by a contracting office k, for an award amount in bin b, in fiscal year $t \in \{2006, ..., 2010, 2012, ..., 2016\}$. Bins are right-inclusive and have \$1,000 length, so that bin $b = 6$ includes all awards in $(\$5,000, \$6,000]$, and so forth. $Post_t$ is an indicator equal to 1 for years 2012 and later. $HighBunch_k$ is an indicator for whether an office is above the median on a rough proxy of pre-period bunching. In particular, for each office k with at least 500 awards in the pre-period, I compute the share of these contracts with value between \$99,000 and \$100,000, and then define the dummy $HighBunch_k$ as equal to 1 for offices above the median for this measure. Finally, Q is some partition of the award amount space. We are interested in the δ_q coefficients on the triple interaction between award size range, high bunching offices and post. Standard errors are clustered by awarding office. The number of observations is 629,804.

Appendix Figure A6 shows estimated δ_q coefficients and their 95% confidence intervals. Panel (a) presents the results for the following partition of the award amount space (all numbers represent thousand dollars): $Q = \{\ (5, 50], (50, 85], (85, 100], (100, 200], (200, 500], (500, 5000]\}$. Panel (b) repeats the exercise by collapsing the $(85, 100]$ and $(100, 200]$ indicators into a single variable. The excluded range in both specifications is $(500, 5000]$, since these are awards that are not plausibly being affected by the change in the simplified acquisition threshold.

The δ_q coefficients indicate the change in number of contracts, for offices with high pre-period bunching relative to those with low pre-period bunching, for award amounts in range q, relative to the difference for awards between \$500,000 and \$5,000,000. The results indicate that any differential change between low-bunching and high-bunching offices occurred only within a window around the old and new simplified acquisition threshold. When combined into a single indicator (panel (b)), we see that high-bunching offices on net increased their number of contracts relative to low-bunching offices. Further, this increase was not compensated elsewhere in the distribution, as there are no differences between the two types of offices on how transactions changed below

$85,000 and above $200,000. This evidence is at odds with split purchases being the main driver of the observed bunching.

D. Sensitivity to Bunching Estimation Parameters

This appendix explores the sensitivity of the bunching estimates with respect to key estimation parameters. Throughout, I focus on one measure of the bunching estimation: the net excess mass $(\hat{m} - \hat{x})$. This corresponds to a normalized measure of the estimated extensive margin effects that arose in response to the 2011 change in the simplified acquisition threshold. It is computed as the cumulative difference between: (i) the actual award frequency distribution observed in 2011-2016, and (ii) the estimated counterfactual distribution that we would have observed in the absence of a change in the simplified threshold. This difference is normalized by the average frequency over some window around the simplified acquisition threshold.

Section 3.2 describes the estimation procedure in detail. From equations (1) and (2), we see that there are three key parameters that the estimation depends upon. These correspond to the degree of the polynomial used to fit the relative frequency changes (p), and the lower and upper bounds of the excluded region around the simplified acquisition threshold ($\underline{R}$ and $\overline{R}$, respectively).

The goal of this exercise is to assess how sensitive the estimate of net excess mass $(\hat{m} - \hat{x})$ is with respect to p, $\underline{R}$, and $\overline{R}$. In the baseline estimates, the chosen parameters are $p = 3$, $\underline{R} = 50$, $\overline{R} = 300$, where $\underline{R}$, and $\overline{R}$ are represented in thousand dollars. The baseline estimate of net excess mass is presented in Table 4, and is equal to $(\hat{m} - \hat{x}) = 18.76$, with a standard error 4.50. Therefore, a 95% confidence interval based on asymptotic normality corresponds to $[9.94, 27.58]$

Appendix Table B1 presents 75 different estimates of $(\hat{m} - \hat{x})$, computed for each possible combination of $(p, \underline{R}, \overline{R}) \in \{2, 3, 4\} \times \{30, 40, 50, 60, 70\} \times \{200, 250, 300, 350, 400\}$. The baseline estimate can be found at the center of panel (b). Starting from this value, and fixing the polynomial degree $p = 3$, we see that the estimate is remarkably robust to changes in the definition of the excluded region. The estimates range between a minimum of 10.48 and a maximum of 26.93, and therefore all 25 values lie within the 95% confidence interval of the baseline estimate.

The results are also robust to considering a second-degree polynomial fit ($p = 2$), as shown in panel (a). The estimate using the baseline excluded region decreases marginally to 17.39. Moreover, the estimates obtained when varying the excluded region definition again are relatively stable. 23 out of 25 of the estimates lie within the 95% confidence interval of the baseline estimate. The coefficient is remarkably stable when we only consider changes to the lower bound of the region of ± 10 thousand dollars, and of ± 50 thousand dollars to the upper bound.

Finally, the estimate is *not* stable when we consider a fourth-degree polynomial ($p = 4$), as presented in panel (c). The point estimate doubles in magnitude, and can yield very large coefficients (up to 80.99) when we consider alternative excluded region definitions. This final result is, perhaps, not surprising. Recall that the interpolation method seeks to fit a smooth function over the relative frequency changes across the award distribution (see Figure 3). This intends to capture the level changes that would occur even in the absence of a regulatory change. And while the polynomial specification allows for some degree of curvature in this fit, we expect these changes to be relatively uniform across the distribution. This means that the polynomial degree that would work well on this interpolation is relatively low. Higher degree polynomials will quickly start over-fitting the observations outside of the excluded region, rapidly decreasing the accuracy of the prediction within the excluded region. Note that this is in contrast with standard bunching methods that use these interpolations to predict distribution *levels* as opposed to *changes*. Because distribution levels typically feature much more curvature, baseline choices of polynomials tend to be of significantly higher order (e.g. Chetty et al. (2011) use a seventh-degree polynomial and Kleven and Waseem (2013) use a fifth-degree polynomial).

E. Validating Quality Measures Using the IT Dashboard

This appendix explores the empirical relationship between the quality proxies used in this paper based on contract execution variables, and credible quality assessment measures available for a sample of large federal IT projects. The latter are obtained from the IT Dashboard dataset, which combines assessments from agency chief information officers (CIOs) based on objective performance criteria with information on cost and timeliness.

E.1. The IT Dashboard

The federal IT Dashboard (www.itdashboard.gov) tracks the performance of the most important federal IT projects. The IT Dashboard was launched in June 2009 and provides Federal agencies, industry, the general public and other stakeholders with measures of the overall performance of major IT projects. CIOs are responsible for evaluating and updating select data on a regular basis.

The overall rating of the project is based on three components: cost, schedule, and performance. In the same spirit of our measures of cost overruns and delays, the first two components compute deviations between planned and actual cost and tardiness of the projects. The third component is a CIO rating "based on his or her best judgment, using a set of pre-established criteria." The evaluation should incorporate the following factors: risk management, requirements management, contractor oversight, historical performance, and human capital. Evaluation ratings are based on a categorical five-point scale, where 1 is the lowest rating (high-risk) and 5 is the highest rating (low-risk).

Importantly, the CIO's rating decision must take as an input a series of objective performance measures. These measures intend to capture with a high degree of accuracy the objective quality of the service being provided, which may be unrelated to cost and schedule performance. While the metrics vary depending on the specific project, examples include "percent of the time that the system is available", "percent of servers reduced as a result of virtualization", "number of applicants using ePermits", "number of repeat customers using system", "reduction in number of field office data storage devices", "percent of customer issues that are addressed and fully resolved within 24-hours", etc. Furthermore, as Liebman and Mahoney (2017) note, there are widespread incentives for the CIO's rate projects accurately.

Below we ask whether and to what extent these credible measures of project quality captured by the CIO ratings correlate with the quality proxies based on execution performance used in the main analysis of this paper.

E.2. CIO evaluation vs contract quality proxies

Using data from 2013 through 2018, we are able to match 390 major IT projects in the Dashboard to at least one contract in our main sample, giving rise to 3,567 different project-contract pairs. Table B2 presents summary statistics of the matched IT Dashboard - FPDS sample. Panel (a) presents counts of number of projects and contracts, as well as spending totals, by agency. Panel (b) presents summary statistics of the main quality measures. The quality proxies used in this paper are computed at the contract level, while the CIO evaluation is computed at the project level.

With this matched sample, we can then analyze the extent to which the quality proxies based on contract implementation correlate with CIO evaluations at the contract level. Figure A5 shows the result. Panel (a) presents the value of the unconditional mean quality index by values of CIO evaluation. We see that there is a clear monotonic relationship, so that projects with higher CIO evaluation have associated contracts with a higher value of the quality proxy index based on contract execution measures. This result is robust to the inclusion of agency fixed effects (Panel (b)) and weighting by the size of the project (Panel (c)). The coefficient of the slope of the linear fit in Panel (b) is 0.057, with a standard error (clustered at the IT project level) of 0.027. The analogous coefficient in Panel (c) (weighting by the natural logarithm of the project's total obligated dollars) is 0.054 with a clustered standard error of 0.027.

Taken together, these results indicate that, for this sample of matched IT projects, CIO evaluations based on actual quality metrics correlate positively with the quality proxies computed in this paper based on contract execution variables.

F. Model Details

This appendix complements the main exposition of the model in Section 4. Here I provide additional details on how equilibrium awards are determined for both aligned and misaligned bureaucrats, for the case of a threshold regulation.

F.1. Aligned bureaucrat

Aligned bureaucrats engage in Nash bargaining with the firm. The expected payoff for the agency is: $\tilde{U}(p) = v_A - p - \gamma\kappa \cdot \mathbf{1}[p > \bar{P}]$. The payoff for the firm is: $\Pi(p) = p - c - (1-\gamma)\,\kappa \cdot \mathbf{1}[p > \bar{P}]$. The Nash bargaining (NB) objective ($\zeta(p)$) is, therefore:

$$\zeta(p) = (v_A - p - \gamma\kappa \cdot \mathbf{1}[p > \bar{P}])^{\phi} \cdot (p - c - (1-\gamma)\,\kappa \cdot \mathbf{1}[p > \bar{P}])^{(1-\phi)}$$

The four cases (A, B, C and D) outlined in Section 4 represent the four possible awards p_A^* that maximize $\zeta(p)$. The four cases represent, respectively: an interior solution subject to an award below the threshold (case A); a corner solution with an award equal to the threshold (case B); an interior solution subject to an award above the threshold (case C); or no solution (case D), i.e. no award gives nonnegative payoffs to both parties. I proceed to describe each of these cases in detail.

Case A: The first candidate solution occurs when the objective is maximized below the threshold $\bar{P}$. Regulation is essentially irrelevant, and parties divide the surplus as if the κ costs did not exist. The equilibrium award is obtained by maximizing $\zeta(p)$ subject to the condition that the award does not exceed the threshold, as in Equation (5). Taking first order conditions, it is easy to check that the equilibrium award is:

$$p_A^{*CaseA} = \phi c + (1-\phi) v_A \ .$$

Payoffs obtained in this case are:

$$\tilde{U}(p_A^{*CaseA}) = \phi(v_A - c) \qquad ; \qquad \Pi(p_A^{*CaseA}) = (1-\phi)(v_A - c) \ .$$

Finally, the NB objective evaluated at this solution is:

$$\zeta(p_A^{*CaseA}) = [\phi(v_A - c)]^{\phi} \cdot [(1-\phi)(v_A - c)]^{1-\phi} \ .$$

Case B: A second candidate award is the corner solution $\bar{P}$. This solution arise when the NB is optimized by "bunching" at the simplified acquisition threshold. The award is:

$$p_A^{*CaseB} = \bar{P}$$

Payoffs are:

$$\tilde{U}(p_A^{*CaseB}) = v_A - \bar{P} \qquad ; \qquad \Pi(p_A^{*CaseB}) = \bar{P} - c \ .$$

The NB objective is:

$$\zeta(p_A^{*CaseB}) = (v_A - \bar{P})^{\phi} \cdot (\bar{P} - c)^{1-\phi} \ .$$

Case C: The third candidate solution occurs when the objective is maximized strictly above the threshold. Regulation is enforced, and parties divide the surplus taken the κ costs into account. The equilibrium award is obtained by maximizing $\zeta(p)$ subject to the condition that the award exceeds the threshold, as in Equation (6). Taking first order conditions, it is easy to check that the equilibrium award is:

$$p_A^{*CaseC} = \phi c + (1-\phi)v_R + (\phi - \gamma)\kappa \ .$$

Payoffs obtained in this case are:

$$\tilde{U}(p_A^{*CaseA}) = \phi(v_R - c - \kappa) \qquad ; \qquad \Pi(p_A^{*CaseA}) = (1-\phi)(v_R - c - \kappa) \ .$$

The NB objective evaluated at this solution is:

$$\zeta(p_A^{*CaseC}) = [\phi(v_R - c - \kappa)]^{\phi} \cdot [(1-\phi)(v_R - c - \kappa)]^{1-\phi} \ .$$

Case D: The fourth candidate solution is to not transact at all. This gives a payoff of zero to the transacting parties. Awards, payoff and the value of the NB objective are given by:

$$p_A^{*CaseD} = \emptyset \ ,$$

$$\tilde{U}(p_A^{*CaseD}) = 0 \qquad , \qquad \Pi(p_A^{*CaseA}) = 0 \ ,$$

$$\zeta(p_A^{*CaseD}) = 0.$$

Equilibrium: Given the above, it follows that the equilibrium award for an aligned bureaucrat will be the candidate solution that yields a higher NB objective:

$$p_A^* = \arg\max_{p \in \mathbb{P}} \zeta(p),$$

where $\mathbb{P} = \{p_A^{*CaseA}, p_A^{*CaseB}, p_A^{*CaseC}, p_A^{*CaseD}\}$.

F.2. Misaligned bureaucrat

In the case of the misaligned bureaucrat, the possible levels of private payoff are given by $b_M(p)$ for awards at or below the simplified threshold. For awards above the threshold, the private payoff is the reservation value b_R, whereas with no transaction the misaligned bureaucrat obtains 0.

Case A': Because $b_M(\cdot)$ is increasing and awards up to v_A are undetected, the first candidate solution is to award $p_M^{*CaseA'} = v_A$ (as long as $v_A < \bar{P}$), and obtain the private payoff $B_M(p_M^{*CaseA'}) =$

$b_M(v_A)$. The firm obtains $\Pi(p_M^{*CaseA'}) = (v_A - c)$ and the government obtains $\tilde{U}(p_M^{*CaseA'}) = v_M - v_A$.

Case B': The second candidate solution is reached at the simplified acquisition threshold. In this case, $p_M^{*CaseB'} = \bar{P}$, with the bureaucrat getting a private payoff of $B_M(p_M^{*CaseB'}) = b_M(\bar{P})$. The firm obtains $\Pi(p_M^{*CaseB'}) = (\bar{P} - c)$ and the government obtains $\tilde{U}(p_M^{*CaseB'}) = v_M - \bar{P}$.

Case C': The third candidate solution is going through the regulated procurement process. The solution is identical to that with an aligned bureaucrat (case C), except for the fact that the misaligned bureaucrat gets a different private payoff. The candidate award is $p_M^{*CaseC'} = \phi c + (1 - \phi)v_R + (\phi - \gamma)\kappa$, and the bureaucrat obtains a private payoff of $B_M(p_M^{*CaseC'}) = b_R$. The firm obtains $\Pi(p_M^{*CaseC'}) = (1 - \phi)(v_R - c - \kappa)$ and the government obtains $\tilde{U}(p_M^{*CaseB'}) = \phi(v_R - c - \kappa)$.

Case D': The final candidate solution is to not transact at all and, therefore, payoffs for all parties are equal to zero. That is, $p_M^{*CaseD'} = \emptyset$, $B_M(p_M^{*CaseD'}) = 0$, $\Pi(p_M^{*CaseD'}) = 0$, and $\tilde{U}(p_M^{*CaseD'}) = 0$.

Equilibrium: I assume that, unlike the aligned agent, the misaligned agent makes a take-it-or-leave-it offer to the firm. Because these offers will typically maximize the payoff of the firm —recall that the misaligned bureaucrat pays as much as possible—, this assumption does not matter substantially, yet simplifies the analysis significantly. Additionally, I assume that, whenever the agent makes an offer that is inconsistent with Nash-bargaining (cases A' or B'), the firm will accept such offer only if the payoff it obtains is as least as high as the one under regulation. This is a no-whistle-blowing condition: the firm will not reveal the identity of the misaligned agent as long as it benefits from the misaligned behavior. This condition may only bind in case B', since the firm payoff is strictly better under A' than C'.

All the above implies that misaligned agents simply compare the payoff obtained between the four cases A', B', C' and D', subject to the no whistle-blowing condition. That is:

$$p_M^* = \arg\max_{p \in \mathbb{P}'} B_M(p),$$

where:

$$\mathbb{P}' = \begin{cases} \{p_M^{*CaseA'}, p_M^{*CaseB'}, p_M^{*CaseC'}, p_M^{*CaseD'}\} & \text{if } \Pi(p_M^{*CaseB'}) \geq \Pi(p_M^{*CaseC'}) \\ \{p_M^{*CaseA'}, p_M^{*CaseC'}, p_M^{*CaseD'}\} & \text{if } \Pi(p_M^{*CaseB'}) < \Pi(p_M^{*CaseC'}) \end{cases}$$

G. Model Estimation Details

Section 4.3 describes how I take the model to the data. In this appendix, I provide additional details of the model estimation process. Below I discuss the smoothing of the estimation moments, the choice of the weighting matrix, and the computation of standard errors

G.1. Smoothing of award distribution moments

The issue with directly using the empirical moments from the award distribution is that the model will not be able to replicate the spikes at every round number observed in Figure 1. Given the smooth distribution assumed for $\bar{v}$ and c, the award distribution will also be smooth everywhere, except for the simplified acquisition threshold. Therefore, rather than complicating the model by incorporating a feature that generates round-numbers bias, I simply modify the estimation moments by smoothing them. I do so by using a correction that follows the logic in Kleven and Waseem (2013) to control for round-number effects.

In particular, I estimate two separate regressions, one above and one below the simplified acquisition threshold, using data from the pre-reform period (2006 through 2010, when the threshold was equal to \$100,000). I first classify all awards between \$50,000 and \$200,000 into bins of \$1,000 length, where bin $b = 51$ includes all awards in $(\$50,000, \$51,000]$, and so forth. Let n_b the number of contracts in bin $b \in \{51, ..., 200\}$. And let R be a vector of round-numbers multiples for which we think there is a special bias. I estimate the following regressions:

$$n_b = \sum_{p=1}^{P} \beta_p \cdot b^p + \sum_{r \in R} \rho_r \cdot \mathbf{1}\left[\frac{b}{r} \in \mathbb{N}\right] + \sum_{k=0}^{2} \gamma_k \cdot \mathbf{1}[b = 100 - k] + \epsilon_b, \qquad \text{for } b = 51, ..., 100.$$

$$n_b = \sum_{p=1}^{P} \delta_p \cdot b^p + \sum_{r \in R} \lambda_r \cdot \mathbf{1}\left[\frac{b}{r} \in \mathbb{N}\right] + \nu_b, \qquad \text{for } b = 101, ..., 200.$$

Both specifications regress the frequency counts on a P-order polynomial of award values and a series of round-number dummies. I choose $P = 6$ and $R = \{5, 10, 25\}$, so that there are special round number effects for awards in bins that are multiples of \$5,000, \$10,000, and \$25,000. The specification below the threshold also has three "bunching dummies" to capture the non-smooth spike right below the threshold.

The smoothed distribution is then obtained in two steps. I first take the prediction from these regressions, *ignoring* the contribution from the round-number dummies. I then adjust the frequency counts above and below by a fixed factor, so that the total number of observations in the smoothed distribution matches that of the actual data. That is, first I compute:

$$\hat{n}_b = \begin{cases} \sum_{p=1}^{P} \hat{\beta}_p \cdot b^p + \sum_{k=0}^{2} \hat{\gamma}_k \cdot \mathbf{1}[b = 100 - k] & \text{if } b = 51, ..., 100 \\ \sum_{p=1}^{P} \hat{\delta}_p \cdot b^p & \text{if } b = 101, ..., 200 \end{cases}$$

And then obtain the smoothed frequencies by proportionally adjusting the counts to make the total number of observations match, i.e.:

$$\hat{n}_b^{smoothed} = \begin{cases} \hat{n}_b \cdot \frac{\sum_{b=51}^{100} n_b}{\sum_{b=51}^{100} \hat{n}_b} & \text{if } b = 51, ..., 100 \\ \hat{n}_b \cdot \frac{\sum_{b=101}^{200} n_b}{\sum_{b=101}^{200} \hat{n}_b} & \text{if } b = 101, ..., 200 \end{cases}$$

Finally, the vector of 151 award distribution moments is given by the *normalized* frequencies, i.e.:

$$\hat{f}_b = \frac{\hat{n}_b^{smoothed}}{\sum_{b=51}^{200} \hat{n}_b^{smoothed}}, \qquad \text{for } b = 51, ..., 200.$$

G.2. Smoothing of quality moments

For similar reasons, I also smooth the quality distribution moments. Because there are fewer quality moments, and because the empirical patterns are already relatively smooth (see Figure 6), I use a more parametric approach than with the distribution moments. Again, I fit separate regressions above and below the simplified acquisition threshold, yet for the quality moments I use contract-level observations for awards between \$50,000 and \$200,000.

Below the threshold, I estimate the following quadratic fit:

$$q_i = \alpha_0 + \alpha_1 \cdot b_i + \alpha_2 \cdot b_i^2 + \epsilon_i$$

for awards i between \$50,000 and \$100,000, and where b_i are defined as above (\$1000-wide right-inclusive award value bins). The vector of 5 quality moments below the threshold is simply:

$$\hat{q}_b = \hat{\alpha}_0 + \hat{\alpha}_1 \cdot b + \hat{\alpha}_2 b^2, \qquad \text{for } b = 55, 65, 75, 85, 95.$$

As described in Section 4.3, above the threshold I simply fit a linear function of the quality indices on award values. With this I obtain the estimates of δ_0 and δ_1 that I use to translate post-award performance shocks into quality moments. I estimate: $q_i = \delta_0 + \delta_1 \cdot b_i + \nu_i$ using awards between \$100,000 and \$200,000, and then take the predicted values at 10 bins above the threshold: $b = 105, 115, 125, 135, 145, 155, 165, 175, 185, 195$.

G.3. Weighting matrix

The starting point for the weighting matrix W is an estimate of the inverse of the asymptotic variance of the estimation moments, $\hat{\Omega}^{-1}$. I compute this estimate via bootstrap. I re-sample contracts with replacement from the original data, and recompute the smoothed vector of moments, repeating this process $B = 500$ times. I then compute $\hat{\Omega}^{-1}$ as the sample variance of these 500 vectors.

Following Einav et al. (2018), I then modify the weighting matrix to increase the weight given to the award distribution moments that are closer to the simplified acquisition threshold. I do this by adding an identity matrix to $\hat{\Omega}^{-1}$, and then subtracting a fixed amount of 0.01 for each \$1,000 away from the threshold.

In particular, consider the following diagonal matrix:

$$D = diag(\underbrace{0.51, 0.52, ..., 0.99, 1.00, 0.99, ..., 0.01, 0.00}_{\text{150 distribution moments}}, \overbrace{1.00, 1.00, ..., 1.00}^{\text{15 quality moments}})$$

The weighting matrix is given by:

$$W = \hat{\Omega}^{-1} + D$$

G.4. Standard errors

I compute standard errors using the asymptotic variance formula given by Equation (8). That is, the variance covariance matrix of the estimates $\hat{\theta}$ is:

$$\hat{V}(\hat{\theta}) = \frac{1}{n}\left(1 + \frac{1}{s}\right)\left(\hat{M}'W\hat{M}\right)^{-1}\hat{M}'W\hat{\Omega}W\hat{M}\left(\hat{M}'W\hat{M}\right)^{-1}$$

I compute $\hat{M}$ as the Jacobian matrix of the SMM objective function (Equation (7)), evaluated at $\hat{\theta}$. I compute this Jacobian numerically.