Rules Versus Discretion in Public Procurement

Rodrigo Carril*

November 2026

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. Based on the threshold in force during the estimation period, the model implies that the benefits from waste prevention were modest relative to the compliance costs introduced by regulation. More recent regulatory changes that have increased the level of discretion are consistent with the model's optimal policy. 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 the Editor (Pierre Dubois), four anonymous referees, 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 grants MICIU/AEI/UE - PID2023-153318NB-I00 and RYC2024-050701-I, and through the Severo Ochoa Programme for Centres of Excellence in R&D (CEX2024-001476-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 remain relatively understudied areas of government activity.

In this paper, I analyze how the strength of procurement regulation affects the incentives of public buyers and private sellers and, ultimately, 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]

[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." Interviews with procurement officers echoed these concerns, with several buyers arguing that the additional scrutiny leads to a more passive management of contracts, where the effort is instead focused on complying with the different layers of review (legal, small-business office, policy, competition advocate, government accountability office, etc.). This leads to a "defensive" or "risk-averse" procurement practice, in which the incentive is to avoid procedural error at the cost of worse performance.

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 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 would have been observed after 2011 had the threshold remained at $100,000 rather than rising 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 4,095 additional contracts per year (equivalent to 7.3% of contracts between $50,000 and $300,000), which translated into $526 million in additional annual contract spending (8.7% 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 can be constructed across the main contract sample, subject to the data-availability and cleaning restrictions described below, but primarily capture execution quality rather than the quality of the goods and services obtained. To assess whether these proxies also reflect the quality of the purchased goods and services, I compare them with CIO project ratings in a matched sample of federal IT awards. The positive association in the baseline delivery-order and BPA-call sample supports their use, although it need not generalize to all procurement. Using the execution proxies in the main analysis, I find that contracts subject to tighter regulation exhibit worse ex-post performance, consistent 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. In the model, tight rules eliminate misaligned agents' wasteful spending but impose 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 can be awarded with 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.

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

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: the estimated distribution has a mean of $16.87 thousand, rationalizing the strong bunching in the data. I also estimate that 77.5% of procurement opportunities are led by aligned bureaucrats. This estimate reflects the quality drop when discretion is removed from contracting 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 because the analysis only allows me to recover public agencies' valuations for contracted goods and services, as opposed to citizens' valuations of such products, which would be a key input for a full welfare analysis. The estimates imply that the optimal simplified acquisition threshold should have been $233,000 in FY2010. Interestingly, adjusting this value by inflation lines up almost perfectly with the current simplified acquisition threshold of $350,000 as of FY2026.[7] Although this is a strong extrapolation, it suggests that recent threshold increases that have outpaced inflation may have closed the estimated gap between observed and optimal policy.

This paper relates to a rapidly growing literature that studies how institutional and market features affect contracting outcomes in public procurement.[8] In particular, a number of recent and concurrent studies have analyzed the role of buyer discretion in public procurement. My results are in line with evidence from Pertold and Palguta (2017) in the Czech Republic and Szucs (2024) in Hungary, showing that procurement rules that change at arbitrary thresholds can induce transaction responses that result in "bunching" of contracting activity. 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 increase bureaucratic discretion in institutional settings that feature a low degree of misalignment, which rationalizes the cross-country patterns presented by these authors. It also helps reconcile the results from different country-specific studies, where discretion is sometimes found to improve (e.g., Coviello et al., 2018; Bandiera et al., 2021; Decarolis et al., 2025)

[7] Based on the Consumer Price Index for All Urban Consumers (CPI-U), the inflation-adjusted value of $233,000 in FY2010 is $357,300 in FY2026.

[8] Examples include the effects of procurement centralization (Bandiera, Prat, and Valletti, 2009; Lotti et al., 2024; Allende et al., 2025), the use of electronic platforms (Lewis-Faupel et al., 2016), publicity requirements (Coviello and Mariniello, 2014; Carril, Gonzalez-Lira, and Walker, 2026), the use of different auction formats (Decarolis, 2014, 2018), expiring annual budgets (Liebman and Mahoney, 2017), bureaucrats' characteristics (Coviello and Gagliarducci, 2017; Decarolis, Giuffrida, Iossa, Mollisi, and Spagnolo, 2020; Best, Hjort, and Szakonyi, 2023), bureaucrats' workload (Warren, 2014), audits (Colonnelli and Prem, 2022; Gerardino, Litschig, and Pomeranz, 2024), industry consolidation (Carril and Duggan, 2020), and political connections (Baltrunaite, 2020; Brugués, Brugués, and Giambra, 2024; Szucs, 2024).

or worsen (e.g., Palguta and Pertold, 2017; Szucs, 2024) outcomes.

Two recent papers have also studied 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 (2022) 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.

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.[9] 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, 2020; Gelber, Jones, Sacks, and Song, 2021; Marx, 2024).

More broadly, this paper contributes to the literature that studies how the incentives of public officers inform policy design and affect public service delivery.[10] 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; Rasul and Rogger, 2018; Duflo, Greenstone, Pande, and Ryan, 2018; Gerardino, Litschig, and Pomeranz, 2024).

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 agencies obligated $554.1 billion to procurement contracts, accounting for about 40% of discretionary spending.[11]

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

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

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

[11] Discretionary spending excludes mandatory programs, such as Social Security and Medicare, and interest on debt. Source: US Government Accountability Office, *A Snapshot: Government-Wide Contracting, A 2018 Update*, https://www.gao.gov/assets/700/699330.pdf.

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). The main analysis focuses on non-IDV DCs and purchase orders. 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.[12] 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[13] 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.[14] 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, a change that we will leverage for identification. Since 2018—after our period of analysis—it was further increased to $250,000 in a non-uniform manner across agencies, before a FAR amendment made the higher threshold effective on August 31, 2020.[15]

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

[12] 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.

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

[14] The relevant regulation is FAR 13.003(a), which states that: *"Agencies shall use simplified acquisition procedures to the maximum extent practicable for all purchases of supplies or services not exceeding the simplified acquisition threshold"*. Note that the threshold is based on the actual award amount, rather than an ex-ante estimate. If a buyer planned to award a contract using simplified acquisition, and learned that the best quote would be above the threshold, they would have to adjust the procedure and follow the more stringent rules. This was confirmed by several procurement officers in interviews. They also suggested that their ability to verbally solicit quotes from potential suppliers gave them a very good sense of the market and the prices they can expect.

[15] See Section 805 of the National Defense Authorization Act for Fiscal Year 2018, CAAC Letter 2018-02 (https://www.acquisition.gov/caac/2018-02), and FAR Case 2018-004 (https://www.acquisition.gov/content/publication-fac-2020-07).

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 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. These requirements are intended to constrain favoritism and other abuses 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 acquisitions qualify for simplified procedures under a much higher ceiling, which is $9 million as of October 1, 2025.[16] 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. 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. Throughout the award-size analysis, the award amount is the expected contract value including options at the time of award. I distinguish this

[16] See FAR 13.500 and the acquisition-related threshold changes effective October 1, 2025: https://www.acquisition.gov/threshold-changes.

amount from realized total obligations, obtained by summing obligations across the contract's actions.

The main analysis sample consists of non-IDV definitive contracts and purchase orders with expected award amounts in $(\$5,000, \$5 \text{ million}]$, awarded between fiscal years 2006 and 2016 for products and services other than Research and Development (R&D).[17] Table 1 presents summary statistics for the data. The table presents statistics for the main sample with award amounts in $(\$5,000, \$5 \text{ million}]$ and for the narrower $(\$50,000, \$300,000]$ window used in parts of the analysis. We see that among federal agencies, the Department of Defense makes the most awards (54% overall and 61% in the restricted sample), followed by the Department of Veterans 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).

Table 2 describes the distribution of FY2016 contracts by realized total obligations, rather than by the expected award amount used in the threshold analysis. This table includes contracts with total obligations of at least $5,000 and imposes no upper limit. Two-thirds have total obligations below $25,000, 80% are below $50,000, and less than 1% exceed $5 million. Contracts above $5 million nevertheless account for more than 80% of spending, while the $50,000–$300,000 bracket accounts for 15% of contracts and 4% of 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

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

(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)).[18] To get a better sense of the dynamics of adjustment to the new regulation, 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. Furthermore, Appendix Figure A2 shows the dynamics of bunching at the quarterly level. It reveals that bunching at $100,000 was stable up until the end of fiscal year 2010, and then gradually dropped until converging to a new lower level by the beginning of 2012. The bunching at $150,000 reflects a mirror image, with a gradual increase from the beginning of 2011, until reaching a permanent higher level in 2012. This further reinforces the idea that the 2011 fiscal year was a transition period.[19] For this reason, in all the analysis that follows I drop 2011, distinguishing between "pre" (2006-2010) and "post" (2012-2016) periods, as in Figure 1.

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 agents' 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. The simplified acquisition ceiling for commercial products is far above the $100,000 and $150,000 thresholds studied here, so we should not expect bunching at these thresholds for exempt purchases. 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 A3 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

[18] 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 (2013), who documented a similar phenomenon in the context of income taxation.

[19] The $150,000 threshold became effective under the Federal Acquisition Regulation (FAR) on the first day of fiscal year 2011. One possible explanation for this slow adjustment is that some contracts—especially those awarded in early FY2011—may have started their solicitation procedures 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.

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 A4 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.

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.[20] This prohibition is featured explicitly in most agencies' manuals and guidance documents,[21] and it is not uncommon for audit processes to be triggered by suspicions of split purchases.[22]

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 A5 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

[20] 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."

[21] 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).

[22] 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 Veterans Affairs (https://www.va.gov/oig/pubs/VAOIG-15-05519-377.pdf).

explained by offices with multiple purchases for the same detailed product code within the fiscal year. Yet as Appendix Figure A5 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 examine this implication with a triple-difference design (DDD), using variation across contracting offices, before and after the policy change, and across award sizes. The starting point is that under the null hypothesis of pure splitting, offices that 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 A7. The plotted coefficients compare changes in log contract counts between high- and low-bunching offices, across award ranges and relative to awards between $500,000 and $5 million, among office–bin–year cells with at least one contract. Panel (a) shows that the number of transactions dropped for contracts just below the initial threshold and increased to its right. This is 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 offsetting decrease elsewhere in the distribution, as we would expect if split purchases were driving the bunching results.

Taken together, these results suggest that splitting is unlikely to explain the observed bunching and that transaction responses are concentrated 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.[23] 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

[23] This argument is formalized in Section 4.

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. 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.[24] 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.[25] 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. The identifying assumptions are twofold. First, transaction responses to the threshold change are confined to a window around the initial and final thresholds. Second, absent the threshold change, proportional changes in contract frequencies would have varied smoothly with award size and would be well approximated within this window by the polynomial fitted outside it.

Like standard bunching designs (Saez, 2010; Chetty et al., 2011; Kleven and Waseem, 2013),

[24] 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.

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

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 denote the number 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,

$$\widehat{m} = \frac{\sum_{b=\underline{R}+1}^{100} (\hat{n}_b^{CF} - n_b^{POST})}{\sum_{b=\underline{R}+1}^{\overline{R}} \hat{n}_b^{CF} / (\overline{R} - \underline{R})} \quad ; \quad \widehat{x} = \frac{\sum_{b=101}^{\overline{R}} (n_b^{POST} - \hat{n}_b^{CF})}{\sum_{b=\underline{R}+1}^{\overline{R}} \hat{n}_b^{CF} / (\overline{R} - \underline{R})} \quad . \tag{2}$$

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 a graphical analogue of (1) after the excluded region has been "dummied out." For readability, the figure uses changes in log frequencies, whereas the baseline estimates use the proportional changes defined above. In both cases, the excluded region runs from \$50,000 to \$300,000. These limits correspond to half the size of the initial threshold and twice the size of the final threshold.[26]

Having estimated $\widehat{RFC}_b$ I can then compute the counterfactual distribution for the post-2011 period, in a world where 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

[26] This choice is obviously arbitrary, yet it is arguably a conservative one. Recall that the evidence from Appendix Figure A7 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.

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. These mechanisms could therefore lead to lower-quality contracts under more extensive regulation.

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.[27] 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 agreement.[28] 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 types of awards to be of lower quality (Liebman and Mahoney, 2017).[29]

I set negative cost overruns and delays to zero and treat negative recorded durations as missing. I also treat an overrun or delay as missing when realized within-scope payments or duration exceed ten times their expected values. The quality analysis uses a common sample of contracts with nonmissing overruns and delays; I set the remaining quality measures to missing for contracts that do not satisfy this restriction.

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 measure of poor performance k for contract i, with $k \in \{1, ..., 5\}$, and let $\bar{Y}^k$ and σ^k be its mean and standard deviation among contracts awarded in 2006–2010 with expected award amounts in $(\$50,000, \$300,000]$. The normalized poor-performance measure k for contract i is $\hat{Y}_i^k = (Y_i^k - \bar{Y}^k)/\sigma^k$, and the quality index reverses the sign before averaging across measures: $q_i = -\frac{1}{5}\sum_{k=1}^{5} \hat{Y}_i^k$. Higher values of q_i therefore correspond to better execution quality.

[27] 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.

[28] E.g. Lewis and Bajari, 2011; Decarolis et al., 2020; Giuffrida and Rovigatti, 2022; Kang and Miller, 2022. This is important since out-of-scope modifications may come mainly from the buyer and be orthogonal to suppliers' performance.

[29] One potential concern with this last measure is reverse causality, since it may be that the timing of the purchase affects the choice of awarding procedure, rather than the other way around. Reassuringly, results are not dependent on any particular measure.

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 on a discrete 1-to-5 scale by chief information officers (CIOs) at the responsible agency. CIO evaluations are ultimately subjective, but they must be informed by a series of objective quality criteria. The specific metrics vary by 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; here, I briefly summarize the main results. Appendix Table B2 presents summary statistics for 552 IT Dashboard projects matched to 5,141 distinct delivery orders and BPA calls.[31] I use this sample to study whether the quality proxies based on award implementation used in the main analysis correlate with CIO evaluations for the associated IT project.[32]

Results suggest a positive correlation between award implementation quality and CIO evaluations. Appendix Figure A6 (a) shows that the unconditional mean quality index is monotonically increasing across the five CIO ratings: $\bar{q} = -0.453$, -0.310, -0.012, -0.009, and 0.005 for ratings 1 through 5, respectively. Panels (b) and (c) show similar positive relationships after controlling for project-agency and year fixed effects, both with and without weighting by project size.

The positive association with CIO ratings suggests that the execution-based measures are reasonable proxies for product and service quality in the matched IT sample. Still, this evidence ought to be interpreted with caution. First, it does not directly establish whether changes in procurement procedures that improve execution performance also improve product or service quality. Second, the relationship may not hold across the whole population of awards, although

[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", and "percent of customer issues that are addressed and fully resolved within 24-hours".

[31] CIO ratings assess the IT project as a whole, whereas a project can involve several awards that contribute differently to its performance. Delivery orders and BPA calls meet specific requirements within an ongoing umbrella arrangement, so their execution may be more closely related to the project's implementation and operation than that of standalone purchases. I therefore use these orders and calls in the baseline validation. Appendix Table B6 shows that adding standalone definitive contracts and purchase orders yields smaller but still positive coefficients.

[32] An alternative would be to perform all the analyses of this paper in the sample of IT projects. However, the small sample makes the nonparametric exercises presented here infeasible.

execution measures should be particularly informative when the purchased goods and services are well defined. Fortunately, the main analysis focuses on definitive contracts and purchase orders with well-defined requirements and uses only modifications within the original scope.

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.[33]

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

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

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 a partition of the award amount space into right-inclusive bins of ten-thousand dollars length, i.e., $S = \{(\$50,000, \$60,000], (\$60,000, \$70,000], ..., (\$290,000, \$300,000]\}$.[34]

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 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

[33] Analogous patterns are observed for each of the five underlying quality measures. Appendix Figure A8 shows the average value of each of these measures across the award distribution, separately for the pre and post periods.

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

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 A9 replicates the plot in Figure 6 Panel (c) for each underlying measure.[35] 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 extensive margin 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.[36]

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 A11 and A12, and are essentially equivalent to those in Appendix Figures A9 and A10.

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

[36] For intensive margin selection, the direction of the bias is less clear, but its magnitude can be gauged by following the methodology of Diamond and Persson (2016) and applied to this context by Coviello et al. (2026). I come back to this point in the next section, where I show that selection on the intensive margin has a relatively small effect on the effects of regulation on contract quality.

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 acquisition pre and post), $(100, 150]$ (regulation pre, simplified acquisition post), and $(150, 300]$ (tight regulation both pre and post).[37] Consistent with the raw data and less parametric specifications discussed 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 and are also reflected in the aggregate quality index.[38] The middle row of Table 5 indicates that contracts newly eligible for simplified procedures experience, on average: (i) 0.09 fewer modifications, a 9% decline relative to the mean; (ii) \$3.6 thousand lower cost overruns, a 42% decline; (iii) 5.4 fewer days of delay, a 7% decline; (iv) a 0.16-percentage-point decline in termination probability, a 12% decline; and (v) a 1.00-percentage-point decline in the probability of an award during the last week of the fiscal year, an 11% decline. Together, these changes raise the normalized quality index by 0.033 points.

The effect of "manipulation" on contract outcomes

Suppose that a buyer strategically adjusts the award amount of a contract to make it eligible for simplified acquisition. Indeed, this is exactly what the intensive margin responses documented above suggest. Do these bunching responses to circumvent regulation lead to increased or reduced contract quality? This question is distinct from the effect of regulation on contract quality and is of interest in its own right. Additionally, quantifying the effect of bunching on contract quality gives us a formal measure of the extent to which the quality effects described above are driven by selection on the intensive margin.

I closely follow Coviello et al. (2026), who study these effects in the context of Italian public procurement. In turn, they build upon the methodology developed by Diamond and Persson (2016) in the context of grade manipulation in Sweden. To be consistent with this literature, in this section I refer to intensive margin responses in award amounts as "manipulation", even if no actual wrongdoing is involved. The identifying assumption is the same as for most bunching estimators, but applies to both count distributions and outcomes. That is, both counterfactual distributions and counterfactual outcomes in the absence of manipulation are well approximated by fitting a polynomial from outside a window around the threshold.

The methodology consists of three steps. First, we use a standard bunching estimator (e.g., Kleven and Waseem, 2013) to construct a counterfactual distribution of contracts in the absence of bunching. This is equivalent to estimating Equation (1), but with bin counts of a given period (e.g., pre-2011) on the left-hand side, and then predicting counterfactual counts without the contribution of the excluded window dummies. Second, a similar specification is estimated on contract *outcomes*, but allowing for a level change at the threshold:

[37] All numbers in thousand dollars.

[38] 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.

$$y_b = \sum_{x=0}^{p} \beta_x \cdot b^x + \delta \cdot \mathbf{1}[b > Threshold] + \sum_{j \in R} \gamma_j \cdot \mathbf{1}[b = j] + \nu_b \quad . \tag{5}$$

Counterfactual outcomes in the absence of manipulation are estimated by:

$$\widehat{y_b} = \sum_{x=0}^{p'} \widehat{\beta_x} \cdot b^x + \widehat{\delta} \cdot \mathbf{1}[b > Threshold] \quad . \tag{6}$$

The final step combines the counterfactual counts and outcomes to estimate the intent-to-treat (ITT) effect of manipulation. This estimate compares the observed mean outcome in the excluded region with its predicted mean in the absence of manipulation:

$$\Delta\widehat{y} = \frac{\sum_{b \in R}(y_b \cdot n_b)}{\sum_{b \in R} n_b} - \frac{\sum_{b \in R}(\widehat{y}_b \cdot \widehat{n}_b)}{\sum_{b \in R} \widehat{n}_b} \quad . \tag{7}$$

One can then divide the ITT effect on contract outcomes by the ITT effect on the use of simplified procedures, to derive a local average treatment effect (LATE) of manipulation-driven discretion on contract outcomes. We compute both estimates, but the main focus is on the ITT effect on contract quality, as the LATE relies on the stronger exclusion restriction that manipulation only affects contract outcomes through its effect on the use of simplified procedures.

Appendix Table B3 reports the estimated effect of manipulation on the use of simplified procedures: 6 percentage points in the pre-2011 period and 10 percentage points in the post-2011 period. Appendix Table B4 reports ITT increases in the quality index of 0.027 points before the reform and 0.031 points afterward. The point estimates indicate fewer modifications, lower cost overruns, fewer delays, and a lower probability of last-week awards in both periods, although their precision varies. The termination estimates have opposite signs: a decline of 0.3 percentage points before the reform and an increase of 0.3 percentage points afterward. The LATE estimates divide the outcome ITTs by the first-stage effects on simplified procedures.

Overall, this suggests that responses on the intensive margin generate positive selection into simplified acquisitions. This means that the effects of regulation on contract quality that I presented may be overestimated, as the quality drops that we observed at the threshold are partially driven by high quality contracts that bunch just below the threshold. As we will see, however, the magnitude of this bias is quantitatively minor. To assess the magnitude of this selection, note that when estimating Equation (6), we obtain counterfactual outcomes in the absence of manipulation, which allows us to compute directly the drop in quality that we would have observed at the threshold if there were no manipulation. The logic is similar to a "donut-RD" (Barreca et al., 2011; Cattaneo and Titiunik, 2022), in which observations that are close to the threshold are excluded from the estimation.

Appendix Figure A13 presents these results visually for the aggregate quality index pre-2011. To assess the role of the bias directly, the graph includes a polynomial fit using all observations (dashed line), as well as the same fit excluding the window around the threshold (solid line). The latter is an estimate of the counterfactual quality in the absence of manipulation. We see that both fits are very similar, which indicates that the bias introduced by manipulation is small.

Appendix Table B5 directly compares the gaps at the threshold with and without the exclusion of the manipulation window. The estimated quality change at the threshold using all observations is -0.070, while the estimate excluding the manipulation window is -0.072. The difference between them is not statistically significant. This suggests that the quality effects of regulation documented above are not driven by selection on the intensive margin, but rather reflect an effect of regulation on contract quality. A similar pattern is observed for all quality sub-measures. Manipulation makes us overestimate the effect on delays, but underestimate the effect on overruns and terminations, while modifications and last week awards are minimally affected.

Discussion

In sum, across a variety of measures, contracts subject to higher levels of regulation perform worse than lightly-regulated simplified contracts. This is consistent with other work finding that delays and overruns increase above the simplified acquisition threshold (Calvo et al., 2019; Giuffrida and Rovigatti, 2022). Furthermore, results appear to be driven by the effect of regulation itself rather than by the selection of transactions into different acquisition procedures. I show this by leveraging the time variation in the location of the threshold, and finding that outcome changes concentrate in the parts of the distribution where regulation changes, as opposed to contract values for which regulation was kept constant, but that were subject to significant compositional changes due to intensive margin responses. These results are remarkably robust to the inclusion of controls, also suggesting a limited role for selection. Finally, this conclusion is further supported by a more formal test of the effect of manipulation on contract quality, which suggests that counterfactual outcomes in the absence of manipulation are very similar to actual outcomes.

Still, the results ought to be interpreted with caution. The outcome measures describe contract execution and need not capture every dimension of the quality of the purchased good or service. Their positive association with CIO ratings in the baseline IT sample is reassuring, but the relationship may differ across procurement settings.

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).[39] 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, the firm implements the contract, and a shock ϵ is realized, which affects the actual value that the government obtains ex-post, $w(\epsilon)$.[40] 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 $R \in \{0, 1\}$.[41] 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:

$$p_A = \arg\max_p (v_A - p)^\phi (p - c)^{1-\phi} \tag{8}$$

[39] 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.

[40] 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.

[41] I do not model the precise mechanism by which regulation leads to a different distribution of execution shocks. Kelman (1990) argues that this operates through a different selection of suppliers, and through different performance of vendors conditional on selection. The model shuts down the former channel to focus on the latter mechanism. In Section 5.4, I come back to the implications of this assumption.

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 $p_M = v_A$.[42]

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 compensates the bureaucrat by paying a kickback equal to b_M. Another interpretation is that the bureaucrat lacks either ability or motivation. For example, suppose that bargaining requires effort, but 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). Note that these two interpretations coincide with the distinction between *active waste* and *passive waste* proposed by Bandiera, Prat, and Valletti (2009). Since the misalignment in the model could reflect either of the two interpretations, the waste generated can also be interpreted as either active or passive waste.

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 that maximizes the Nash

[42] 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.

bargaining objective is:

$$p_R = \arg\max_p (v_R - p - \gamma\kappa)^{\phi}(p - c - (1-\gamma)\kappa)^{1-\phi} \tag{9}$$

Aligned bureaucrats obtain an expected payoff identical to that of the government (i.e., a share ϕ of this reduced surplus). When a discretionary outcome is unavailable, a misaligned bureaucrat uses regulated procurement whenever the regulated transaction is feasible; the model does not assign a separate private reservation payoff to regulation.

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 chooses a feasible procurement procedure, anticipating the award conditional on that procedure. Simplified procurement is available only for awards that do not exceed $\bar{P}$, except for commercial purchases; regulated procurement may result in an award below the threshold. Conditional on regulation, or on simplified procurement by an aligned agent, the award is determined by bargaining. Third, 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:	$p_A^* = \bar{P} - s$;	$\tilde{U}(p_A^*) = (v_A - \bar{P} + s)$
Case C:	$p_A^* = \phi c + (1-\phi)v_R + (\phi - \gamma)\kappa$;	$\tilde{U}(p_A^*) = \phi(v_R - c - \kappa)$
Case D:	$p_A^* = \emptyset$;	$\tilde{U}(p_A^*) = 0$

Note that bunched transactions are awarded at or slightly below $\bar{P}$, where $s \geq 0$ is a transaction-specific distance between the award price and the statutory threshold.

On the other hand, let p_M^* be the equilibrium award made by a misaligned bureaucrat who faces a simplified acquisition threshold $\bar{P}$. Its possible outcomes are:

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

The bunching outcome requires the absence of an adjustment friction and a nonnegative payoff for the contractor. An aligned buyer chooses bunching when its net payoff is at least as high as under regulation; a misaligned buyer can bunch only if the contractor weakly prefers it to regulation. If bunching is unavailable, a feasible simplified transaction remains the fallback below the threshold; otherwise the parties use regulated procurement whenever it generates nonnegative surplus.

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 B and B′)—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 overpayment by misaligned agents but introduces compliance costs for everyone. 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 would occur without regulation and without incurring compliance costs. By contrast, extensive margin responses represent a 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 would generate 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 $w_i = \bar{v}_i(1 + \epsilon_i)$. The reference valuation 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,

$(\epsilon_i|R=1) \sim N(0,1)$, $(\epsilon_i|A,R=0) \sim N(d,1)$, and $(\epsilon_i|M,R=0) \sim N(-d,1)$, where $d > 0$.[43] Finally, costs lie between one-half and all of the reference valuation:

$$\frac{c_i}{\bar{v}_i} = \frac{1}{2} + \frac{1}{2}U_i^{1/a}, \qquad U_i \sim U(0,1), \quad a = \exp(\ell_a).$$

The distribution of available surplus depends jointly on the cost-ratio distribution and the log-normal distribution of reference valuations. If $a = 1$, costs and reference surplus $\bar{v}_i - c_i$ are uniformly distributed conditional on $\bar{v}_i$.

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, $q_i = \delta_0 + \delta_1 p_i + \epsilon_i + u_q + v_q(p_i - 100)/100$, with award values measured in thousands of dollars. The implementation quality index is allowed to depend linearly on award size to capture the fact that, for reasons unrelated to the agent's actions or the regulatory framework, 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. The pre-period reduced-form relationship above the threshold provides a reference mapping, with $\delta_1 = \alpha_1$ and $\delta_0 = \alpha_0 - \psi(2\lambda - 1)d$. I estimate the additional level and award-size corrections (u_q, v_q) jointly with the structural parameters. Both corrections and the pre-period reduced-form coefficients are held fixed in the counterfactuals. Note that these coefficients help map performance into the observed quality index, but do not directly affect choices or welfare.

Alignment may vary with the reference value:

$$\Pr(A_i = 1|\bar{v}_i) = \text{logit}^{-1}(a_A + b_A z_i), \qquad z_i = \Phi^{-1}(F_{\bar{v}}(\bar{v}_i)),$$

where $F_{\bar{v}}$ is the assumed log-normal distribution of reference valuations. I set a_A so that the mean alignment probability equals λ.

Red tape costs are independent of reference valuations, 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. The estimated specification imposes $\gamma = 0$: the contractor pays the direct compliance cost, while bargaining determines its economic incidence.

Inspired by the commercial-item exception, I allow exempt products to use simplified acquisition regardless of the threshold. Both agent types pay the aligned-reference bargaining price $p_A = \phi c + (1-\phi)(1+d)\bar{v}$, while their expected performance effects remain $+d$ and $-d$. I fix

[43] The symmetry assumes that the entire effect of discretion on execution operates through bureaucratic alignment. This rules out a common effect on mean performance, i.e., an improvement or deterioration that would occur regardless of whether the agent is aligned or misaligned. But note that the average effect across contracts can still be positive or negative, depending on the composition of agents exercising discretion.

the unconditional eligibility probability at $\psi = 0.25$, but allow eligibility to vary with valuation:

$$\Pr(C_i = 1|\bar{v}_i) = \text{logit}^{-1}(a_C + b_C \log(\bar{v}_i/100)) .$$

For every candidate parameter vector, a_C is chosen deterministically so that the unconditional eligibility probability remains exactly ψ.[44]

The model contains two adjustment frictions. First, s_i is an independent exponential buffer with mean s_0, so a bunching transaction is awarded at $\bar{P} - s_i$ rather than at a single point. Second, transactions cannot bunch with baseline probability η. For noncommercial transactions led by a misaligned agent, the effective probability is $\eta + (1 - \eta)\rho_{BM}$. Thus, ρ_{BM} allows misaligned agents to face additional constraints on adjusting noncommercial awards below the threshold.

I estimate the model by the method of simulated moments (see McFadden, 1989, Pakes and Pollard, 1989). To reduce simulation error, I exploit the conditional structure of the model when computing the simulated moments. For each draw of the continuous primitives, I compute equilibrium outcomes under every possible realization of commercial status, bureaucratic alignment, and bunching friction, and average them using their model-implied probabilities. I replace post-award execution shocks by their conditional means and evaluate the remaining expectations over continuous heterogeneity using scrambled Sobol sequences. The distance function is evaluated on two independently scrambled Sobol panels of 200,000 draws each, and the parameter estimates are obtained by minimizing the average of the two panels.

Formally, consider a vector of J moments from the data, given a sample size of n. Denote these *target moments* by m_n. The corresponding simulated moments obtained using the Sobol panel k are denoted by $m_{s,k}(\theta)$.

The estimated parameters are chosen to minimize:

$$\hat{\theta} = \arg\min_{\theta} \ \frac{1}{2} \sum_{k=1}^{2} \left(m_n - m_{s,k}(\theta)\right)' W_n \left(m_n - m_{s,k}(\theta)\right) \tag{10}$$

where W_n is a fixed weighting matrix.

Under the standard SMM regularity conditions, the estimator is asymptotically normal. I use the corresponding normal approximation for inference and compute standard errors using the sandwich formula:

$$\widehat{V}(\hat{\theta}) = (G'WG)^{-1} G'W\Sigma_g WG(G'WG)^{-1}. \tag{11}$$

where G is the Jacobian of the simulated moments with respect to the estimated parameters, and Σ_g is the covariance matrix of the moment discrepancies.[45]

The estimated vector is

$$\theta = (\mu_v, \sigma_v, d, \phi, \lambda, \eta, \mu_\kappa, \sigma_\kappa, s_0, b_C, \rho_{BM}, \ell_a, b_A, u_q, v_q),$$

[44]The calibrated share follows the US Government Accountability Office, "Contracting Data Analysis: Assessment of Government-wide Trends", http://www.gao.gov/products/GAO-17-244SP.

[45]The moments are stacked across the two Sobol panels, with $W = \text{diag}(W_n/2, W_n/2)$. Appendix G describes the construction of Σ_g.

There are thirteen free structural parameters and two corrections to the quality mapping. I profile out the corrections because they affect quality moments without changing equilibrium awards. For each candidate vector of structural parameters, the algorithm chooses (u_q, v_q) to minimize the same distance function, leaving a thirteen-dimensional structural search.

I use 172 moments divided into three categories. 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. I use a contract-level piecewise-linear regression over awards between \$50,000 and \$300,000, allowing for changes at the threshold and fiscal-year effects. I then extract 15 right-inclusive ten-thousand-dollar quality-bin means with upper endpoints $60, 70, \ldots, 200$ (see Appendix G).

The final 7 moments consist of shares of observed simplified procedures in the following award-amount bins: $(\$50,000, \$80,000]$, $(\$80,000, \$90,000]$, $(\$90,000, \$99,000]$, $(\$99,000, \$100,000]$, $(\$100,000, \$110,000]$, $(\$110,000, \$130,000]$, and $(\$130,000, \$200,000]$.

Because much of the identification comes from the behavior close to the threshold, I use a non-optimal weighting matrix that emphasizes moments in the vicinity of the threshold and the size of the quality drop.

I perform the optimization using 20 independent searches of a genetic algorithm, starting from different initial populations and using two independently scrambled 200,000-draw panels. The solution is polished locally using pattern search and tested for local stability. I re-evaluate all the seeds that passed the local stability check using a two-million-draw panel to ensure that the ranking is preserved.

Appendix G presents all remaining estimation details.

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 largely by the shape of the award distribution well below the simplified threshold. The cost-shape parameter ℓ_a and the bargaining weight ϕ also affect this distribution by determining the available surplus and how it is divided. The *magnitude* of the spike in awards at the threshold is directly informative about the level of compliance costs (μ_κ): larger costs make avoiding regulation more valuable and increase the incentive to bunch. The *shape* of the excess mass below the threshold—i.e., how diffuse bunching is—helps identify the mean buffer s_0. The area just to the right of the threshold—and,

in particular, the absence of a hole in the distribution—helps identify the dispersion in red tape costs (σ_κ) as well as the adjustment frictions (η, ρ_{BM}).[46] The shares of simplified procedures across award bins provide additional information about these responses. Above the threshold, simplified purchases reflect the commercial exception, so their shares are particularly informative about the commercial-size slope b_C.

The alignment and performance parameters (λ, d, b_A) are closely related to the quality moments below the simplified acquisition threshold. Given the remaining parameters, the model determines how aligned and misaligned agents differentially sort into different parts of the award distribution and into simplified or regulated procurement. At any bin b, let π_b be the share of simplified transactions and let $\tau_b(\lambda, b_A)$ be the share of aligned agents among those transactions, conditional on the other parameters. Mean post-award performance shocks at that bin are then

$$\bar{\epsilon}_b = \pi_b \left[\tau_b(\lambda, b_A)d + [1 - \tau_b(\lambda, b_A)](-d)\right] = \pi_b[2\tau_b(\lambda, b_A) - 1]d,$$

since regulated transactions have mean-zero shocks. Thus, differences in quality across bins with different compositions of agents and procedures help identify (λ, d), while the way this composition changes with valuation provides information about b_A.

Quality moments above the threshold also contribute to identification because commercial purchases remain simplified and their mean performance shocks need not be zero. These moments provide a reference for the quality mapping, while the jointly estimated corrections (u_q, v_q) absorb smooth differences in its level and slope; they cannot by themselves generate a discontinuity at the threshold. The composition equation helps to see intuitively where the identification of λ comes from. Given that $d > 0$, positive average performance shocks below the threshold require aligned agents to predominate among simplified purchases, whereas negative average shocks require misaligned agents to predominate. This is why quality changes at the threshold are informative about λ: a drop (jump) is consistent with a higher (lower) degree of alignment, after accounting for the model's predictions for commercial purchases and sorting across procedures.[47]

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 mean compliance cost is \$16.87 thousand, equivalent to about 16.9% of a \$100,000 award. The implied standard deviation of compliance costs after censoring at zero is \$10.24 thousand.

Second, the estimated probability that a procurement opportunity is led by an aligned bureaucrat is 77.5%. Under simplified acquisition, the expected performance effect is $+0.253$

[46] 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.

[47] Note that a limitation of this assumption is that, if misaligned agents overpay but achieve better execution on average than the model assumes, the model could attribute some of that performance to a higher degree of alignment.

for aligned agents and -0.253 for misaligned agents, relative to the regulated normalization. As a result, the average performance under simplified acquisition is predicted to be better than above the threshold, consistent with the results from Figure 6.

Third, the government's bargaining weight is estimated to be 0.469. I fix the statutory incidence parameter at $\gamma = 0$, while the bargaining weight determines the economic incidence of compliance costs. 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.

The estimates also indicate substantial heterogeneity in available surplus and frequent constraints on bunching. The point estimate allows alignment to rise with reference valuation, but this slope is small and imprecisely estimated.

Overall, the model is able to closely replicate the key empirical patterns in the estimation sample. Figure 7 presents the smoothed 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.

For example, the model predicts that raising the threshold from \$100,000 to \$150,000 would increase spending on awards between \$50,000 and \$300,000 by 9.9% relative to its \$100,000 baseline. The bunching exercise in Section 3 estimates an increase of 8.7% relative to counterfactual spending had the threshold remained at \$100,000. The model therefore predicts an increase of a similar magnitude, although the two exercises construct their counterfactual baselines differently.

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 A15 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 is high ($\lambda = 0.78$), the model implies that the government would benefit from reducing the stringency of regulation relative to the estimation-period policy.

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 increase to \$150,000 is estimated to have improved —all else equal— the overall benefit for the federal government. The point-estimate buyer-utility curve is maximized at a threshold of \$233,000. 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.

Interestingly, after our sample period ended, the US federal government has continued to increase the simplified acquisition threshold at a rate that has outpaced inflation. In fact, the current value of the simplified acquisition threshold is \$350,000 as of FY2026. This value is roughly equivalent to the inflation-adjusted value of the optimal threshold estimated above. This suggests that—with the caveats involved in extrapolating the model over a 15-year period assuming all other factors remain constant—the US federal government has increased discretion to the approximately

optimal level.

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. Estimating the same structure on data from a different institutional environment—say, one in which corruption or other forms of preference misalignment are much more pervasive—might lead to the conclusion 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 from reduced red tape and quality changes. This feature of the model is consistent with the framework proposed 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. Additionally, data limitations preclude me from capturing certain margins that may be relevant in practice. Here I discuss some of these considerations which could motivate further extensions in future work.

The model 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 characteristics. This assumption follows from a key data limitation in the US federal procurement data: unit prices and quantities are not observed separately. 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. Relaxing this assumption could change the estimated parameters, particularly the compliance costs inferred from bunching. For purchases that currently bunch, higher awards may buy additional quantity or scope, so the model's fixed-product assumption can understate the benefit of raising the threshold.

Relatedly, the same data limitation precludes me from directly measuring the extent of overpricing, which is one of the key motivations for regulation. As a result, the model's assessment of this benefit depends on its pricing assumptions. Note, however, that the model assumes a particularly severe form of overpayment when misaligned agents have full discretion over noncommercial purchases: they pay more than the government's expected value of the purchase. If overpayment were less severe, the benefits of regulation would be smaller. These considerations suggest that, holding the other mechanisms and parameter estimates fixed, the estimated optimal threshold may be a lower bound.

By focusing on the incentives of the buyer, the model remains largely silent about the supply side, assuming a one-to-one bargaining at both sides of the threshold. This may be problematic given abundant evidence that procurement procedures may affect the choice of suppliers (Coviello and Gagliarducci, 2017; Palguta and Pertold, 2017; Decarolis et al., 2025; Baltrunaite et al., 2020). In fact, Carril, Gonzalez-Lira, and Walker (2026) find strong effects on the number of bidders at a lower regulatory threshold in the context of US defense procurement, which leads them to study in detail the supply side responses to changes in procurement procedures. The model presented here can be seen as a complement to this line of work, where instead the focus is on the demand side and the possible misalignment between the government and its agents. Appendix Figure A14 shows that the competitive environment and the characteristics of selected contractors are largely unchanged by regulation in this context, alleviating some of the concerns with abstracting from the supply side. The median number of bids at both sides of the discontinuity is 1, and the mean number of bids is continuous at the simplified acquisition threshold. Moreover, the shares of awards won by small businesses and contractors located abroad show no large changes at the threshold.

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.[48] 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.

Relatedly, the concept of misalignment in this model can encompass both corruption on the one hand, and incompetence or lack of proper incentives on the other. Misalignment can therefore generate a combination of active and passive waste (Bandiera, Prat, and Valletti, 2009). While a richer framework could distinguish these two sources of waste, the setting and data do not provide a source of variation to identify them. For the assumed government objective, this distinction is not crucial. However, if the government placed a higher weight on active waste above and beyond its monetary costs—because of electoral incentives, for example, or because corruption scandals erode public trust—there would be an additional force toward less discretion and a lower regulatory threshold.

The results would also change under different government objectives. One reasonable extension would introduce risk aversion, which would call for more stringent regulation: Congress might be willing to pay a risk premium to eliminate expected losses generated by misaligned agents.

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

[48] 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.

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,[49] translating such a diverse set of contracts 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. The estimates imply that the US federal government would have obtained more value from procurement by allowing greater discretion than the $100,000 threshold permitted during the estimation period. 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. Within the model, transaction responses are informative about compliance costs, while contract-performance differences help identify alignment after accounting for the predicted composition of agents and procedures. The estimated buyer-utility curve is maximized at a simplified acquisition threshold of $233,000. Although this precise value is sensitive to the estimated degree of alignment and depends on parametric assumptions, the structural analysis provides a quantification of the key trade-off

[49] 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.

generated by procurement regulation. This illustrates how the benefits of controlling misaligned agents must be weighed against compliance costs and changes in contract performance.

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 for taxpayers.

References

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

Allende, C., J. P. Atal, R. Carril, J. I. Cuesta, and A. González-Lira (2025). Centralizing procurement: The roles of scale, selection, and variety. Manuscript.

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.

Baltrunaite, A. (2020). Political contributions and public procurement: evidence from lithuania. *Journal of the European Economic Association 18*(2), 541–582.

Baltrunaite, A., C. Giorgiantonio, S. Mocetti, and T. Orlando (2020, 05). Discretion and supplier selection in public procurement. *The Journal of Law, Economics, and Organization 37*(1), 134–166.

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.

Barreca, A. I., M. Guldi, J. M. Lindo, and G. R. Waddell (2011, 10). Saving Babies? Revisiting the effect of very low birth weight classification. *The Quarterly Journal of Economics 126*(4), 2117–2123.

Best, M. C., J. Hjort, and D. Szakonyi (2023). Individuals and organizations as sources of state effectiveness. *American Economic Review 113*(8), 2121–2167.

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 112*(4), 1091–1117.

Brugués, F., J. Brugués, and S. Giambra (2024). Political connections and misallocation of procurement contracts: Evidence from Ecuador. *Journal of Development Economics 170*, 103296.

Calvo, E., R. Cui, and J. C. Serpa (2019). Oversight and efficiency in public projects: A regression discontinuity analysis. *Management Science 65*(12), 5651–5675.

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-Lira, and M. S. Walker (2026). Competition under incomplete contracts and the design of procurement policies. *American Economic Review 116*(2), 535–581.

Cattaneo, M. D. and R. Titiunik (2022). Regression discontinuity designs. *Annual Review of Economics 14*(1), 821–851.

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. *The Review of Economic Studies 89*(2), 695–732.

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, C. Lotti, and G. Spagnolo (2026). Procurement with manipulation. *Journal of Public Economics*. Forthcoming.

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 (2025). Rules, discretion, and corruption in procurement: Evidence from italian government contracting. *Journal of Political Economy Microeconomics 3*(2), 213–254.

Decarolis, F., L. M. Giuffrida, E. Iossa, V. Mollisi, and G. Spagnolo (2020). Bureaucratic competence and procurement outcomes. *The Journal of Law, Economics, and Organization 36*(3), 537–597.

Diamond, R. and P. Persson (2016). The long-term consequences of teacher discretion in grading of high-stakes tests. Technical report, National Bureau of Economic Research.

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.

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. Report to Congressional Addressees GAO-17-244SP, Government Accountability Office.

Gelber, A. M., D. Jones, D. W. Sacks, and J. Song (2021, October). Using nonlinear budget sets to estimate extensive margin responses: Method and evidence from the earnings test. *American Economic Journal: Applied Economics 13*(4), 150–93.

Gerardino, M. P., S. Litschig, and D. Pomeranz (2024). Distortion by audit: Evidence from public procurement. *American Economic Journal: Applied Economics 16*(4), 71–108.

Giuffrida, L. M. and G. Rovigatti (2022). Supplier selection and contract enforcement: Evidence from performance bonding. *Journal of Economics & Management Strategy 31*(4), 980–1019.

Gutman, J. (2014). Is There Room for Discretion? Reforming Public Procurement in a Compliance-Oriented World. Global Economy & Development Working Paper 74, The Brookings Institution.

Kang, K. and R. A. Miller (2022). Winning by default: why is there so little competition in government procurement? *The Review of Economic Studies 89*(3), 1495–1556.

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.

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.

Lotti, C., A. Muço, G. Spagnolo, and T. Valletti (2024). Indirect savings from public procurement centralization. *American Economic Journal: Economic Policy 16*(3), 347–366.

Marx, B. M. (2024). Dynamic bunching estimation with panel data. *Journal of Econometric Methods 13*(2), 225–249.

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.

Newey, W. K. and D. McFadden (1994). Chapter 36 large sample estimation and hypothesis testing. In *Handbook of Econometrics*, Volume 4, pp. 2111–2245. Elsevier.

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.

Palguta, J. and F. Pertold (2017). Manipulation of procurement contracts: Evidence from the introduction of discretionary thresholds. *American Economic Journal: Economic Policy 9*(2), 293–315.

Persson, P. (2020). Social insurance and the marriage market. *Journal of Political Economy 128*(1), 252–300.

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. Working Paper 14152, National Bureau of Economic Research.

Szucs, F. (2024). Discretion and favoritism in public procurement. *Journal of the European Economic Association 22*(1), 117–160.

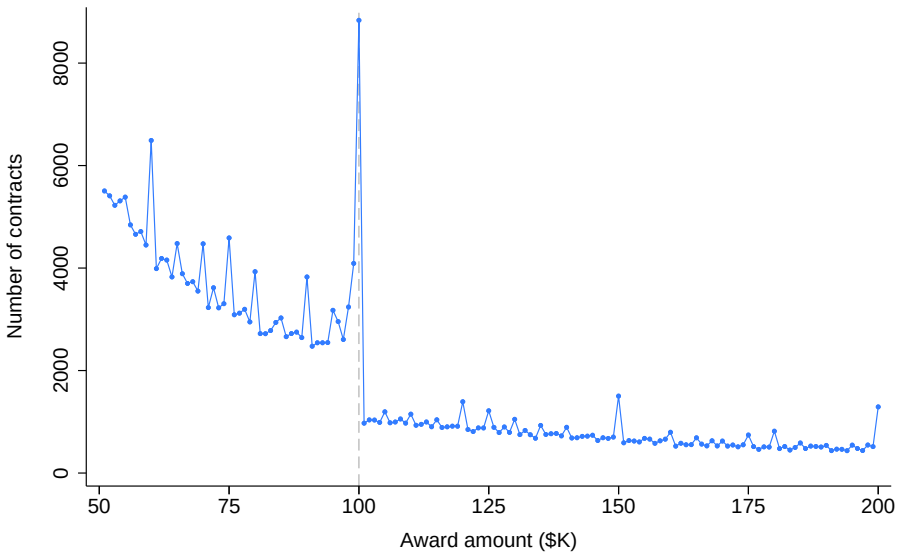
Warren, P. L. (2014). Contracting Officer Workload, Incomplete Contracting, and Contractual Terms. *RAND Journal of Economics 45*(2), 395–421.

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.

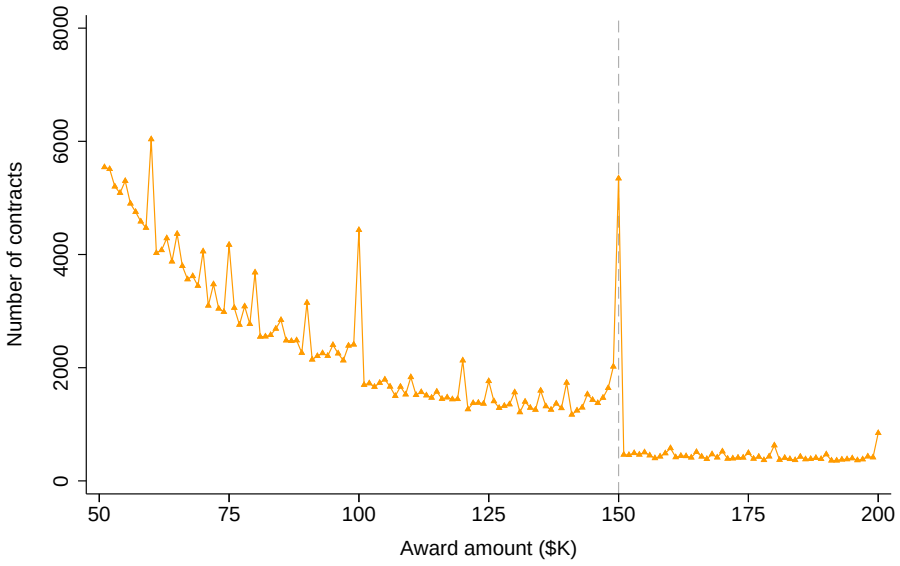
Figures

Figure 1: Distribution of contract awards

(a) "Pre" period: 2006-2010



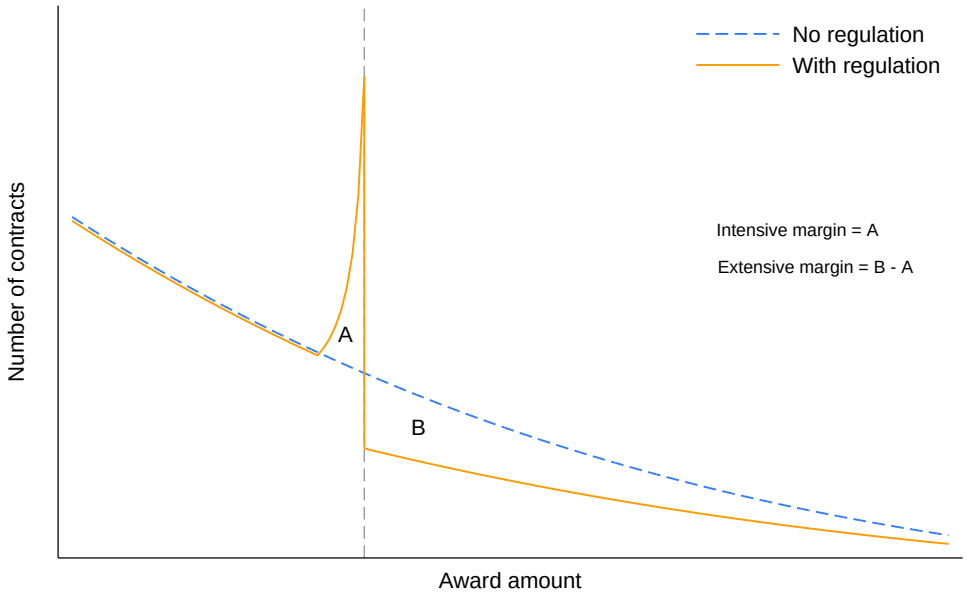
(b) "Post" period: 2012-2016



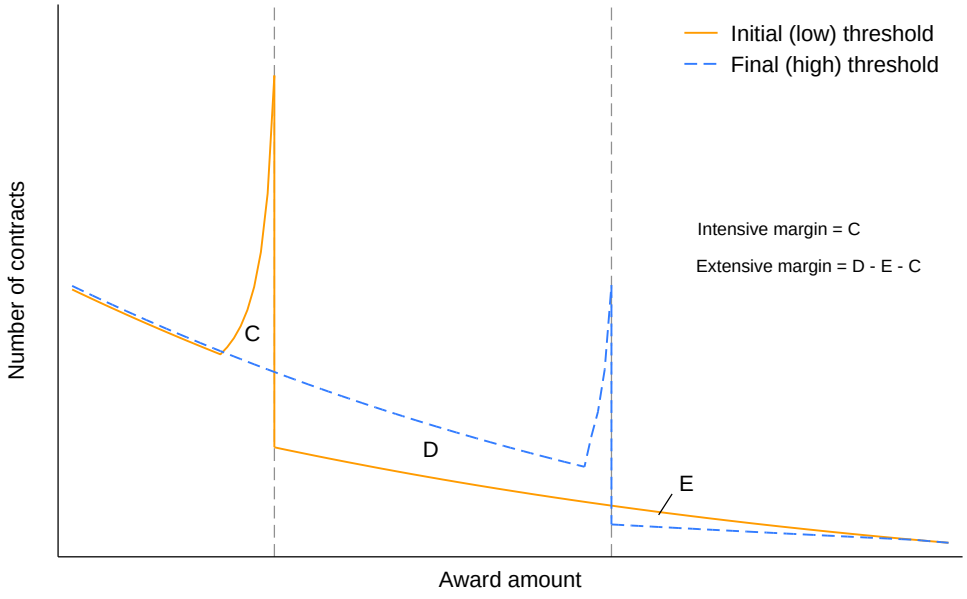
Notes: This figure shows contract awards frequency distributions in two periods, 2006-2010 and 2012-2016. The sample consists of non-R&D definitive contracts and purchase orders from the Federal Procurement Data System-Next Generation, with an award amount between $50,000 and $200,000. Award amounts 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 the more extensive requirements of nonsimplified acquisition.

Figure 2: Potential responses to regulation

(a) Introducing regulation



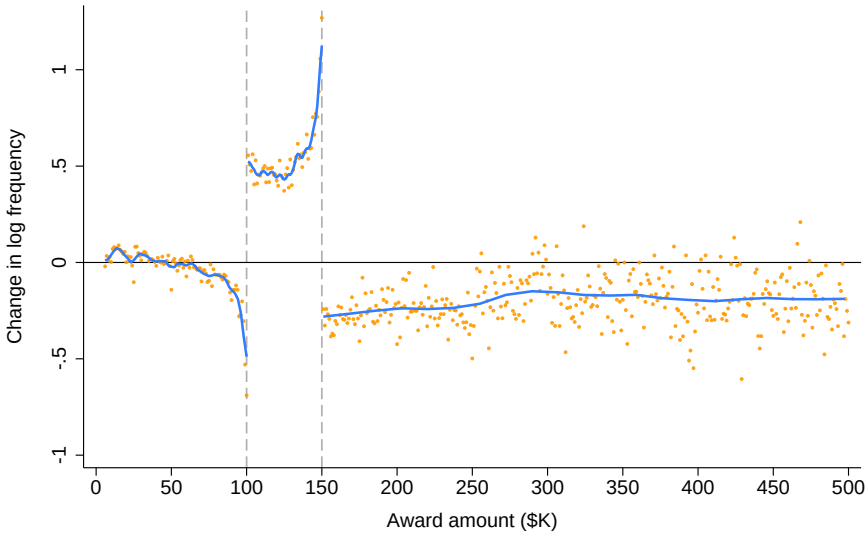
(b) Increasing regulatory threshold



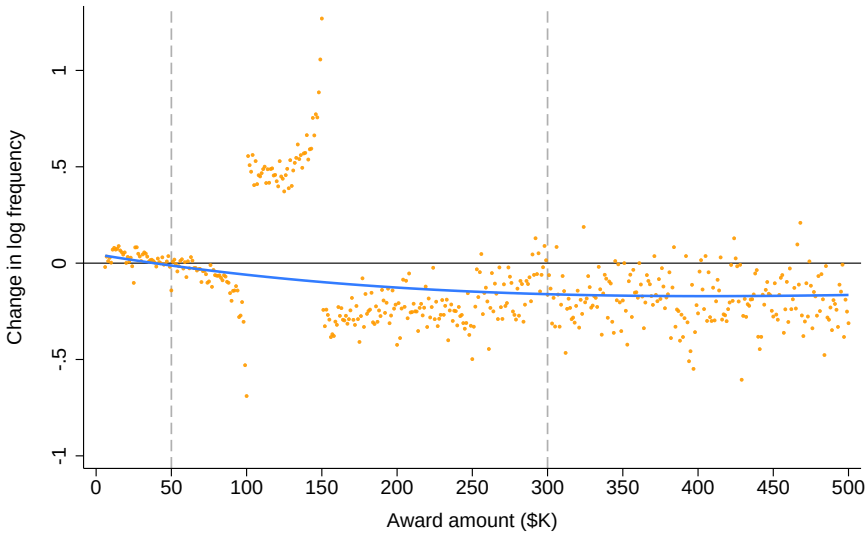
Notes: This figure shows potential changes in the distribution of contract awards in response to regulation. Panel (a) presents a comparison between a situation in which all contracts are allowed to use high-discretion simplified procedures (dashed line distribution), and a world in which only awards 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 and final distributions (B-A and D-E-C, respectively).

Figure 3: Relative change in contracting and interpolation method

(a) Nonparametric smoothing

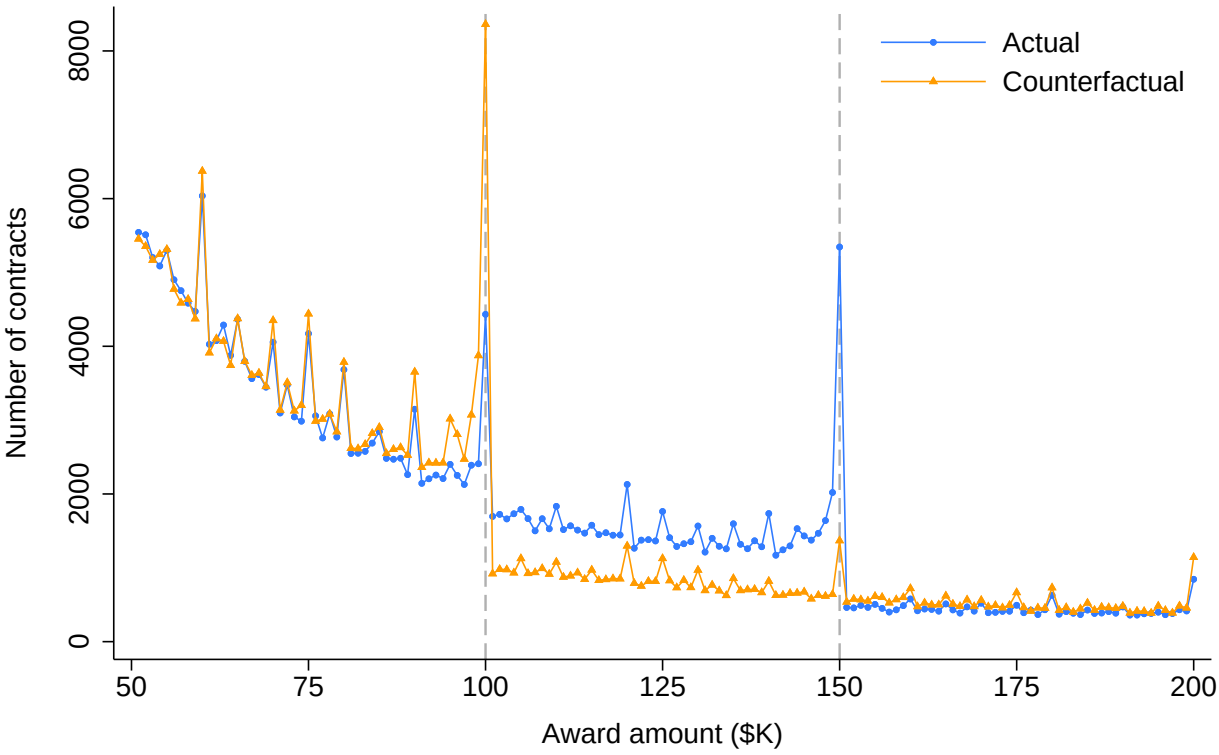


(b) Polynomial interpolation



Notes: This figure illustrates the interpolation method used to estimate the counterfactual distribution in Figure 4. Each orange dot depicts the change in log frequency between the "pre" (2006-2010) and "post" (2012-2016) periods for an award-amount bin. The log transformation is used here for readability; the baseline counterfactual estimates use the proportional frequency changes defined in Equation (1). The sample consists of non-R&D definitive contracts and purchase orders from the Federal Procurement Data System-Next Generation, with award amounts 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 contracts in $(\$1000(X-1), \$1000X]$. Panel (a) adds a nonparametric kernel smoother (blue solid line), fitted separately below the pre-period threshold of \$100,000, between the pre- and post-period thresholds, and above the post-period threshold of \$150,000. Panel (b) regresses the log frequency changes on a third-degree polynomial of award amounts and indicators for bins in the excluded region, which runs from \$50,000 to \$300,000. The blue line excludes the contribution of those indicators and is therefore the log-change analogue of the baseline interpolation.

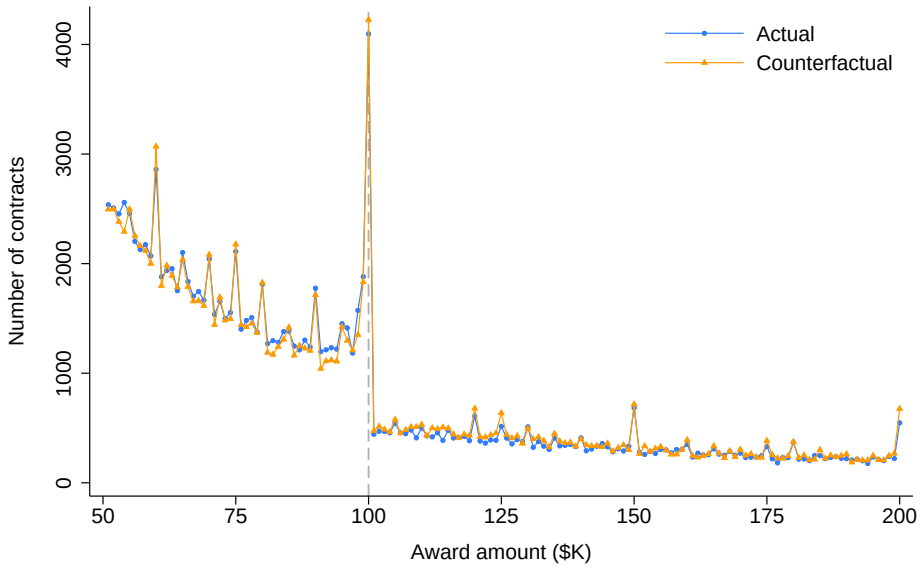
Figure 4: Counterfactual estimation: no threshold change



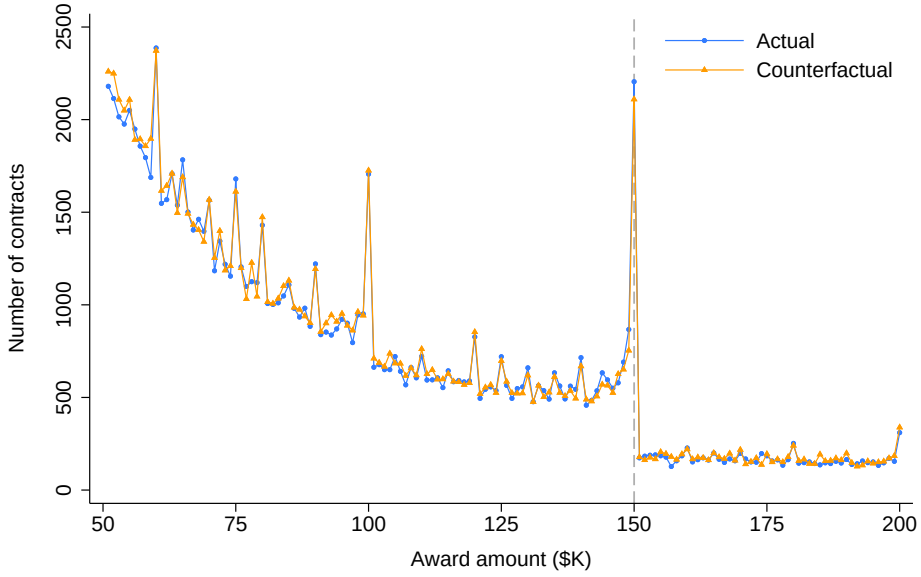
Notes: This figure shows the contract awards frequency distribution in the period 2012-2016 (blue dots) and the counterfactual distribution that would have been observed in that period had the simplified acquisition threshold remained at $100,000 rather than increasing to $150,000 (orange triangles). The sample consists of non-R&D definitive contracts and purchase orders from the Federal Procurement Data System-Next Generation, with award amounts between $50,000 and $200,000. Award amounts are discretized into right-inclusive bins of one-thousand dollars length. A bin labeled "X" includes contracts in $(\$1000(X-1), \$1000X]$. The counterfactual is computed by applying predicted proportional frequency changes to the pre-period distribution (2006-2010). Those changes are interpolated from bins below $50,000 and above $300,000; Figure 3(b) illustrates the analogous interpolation using log-frequency changes.

Figure 5: Placebo estimates

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

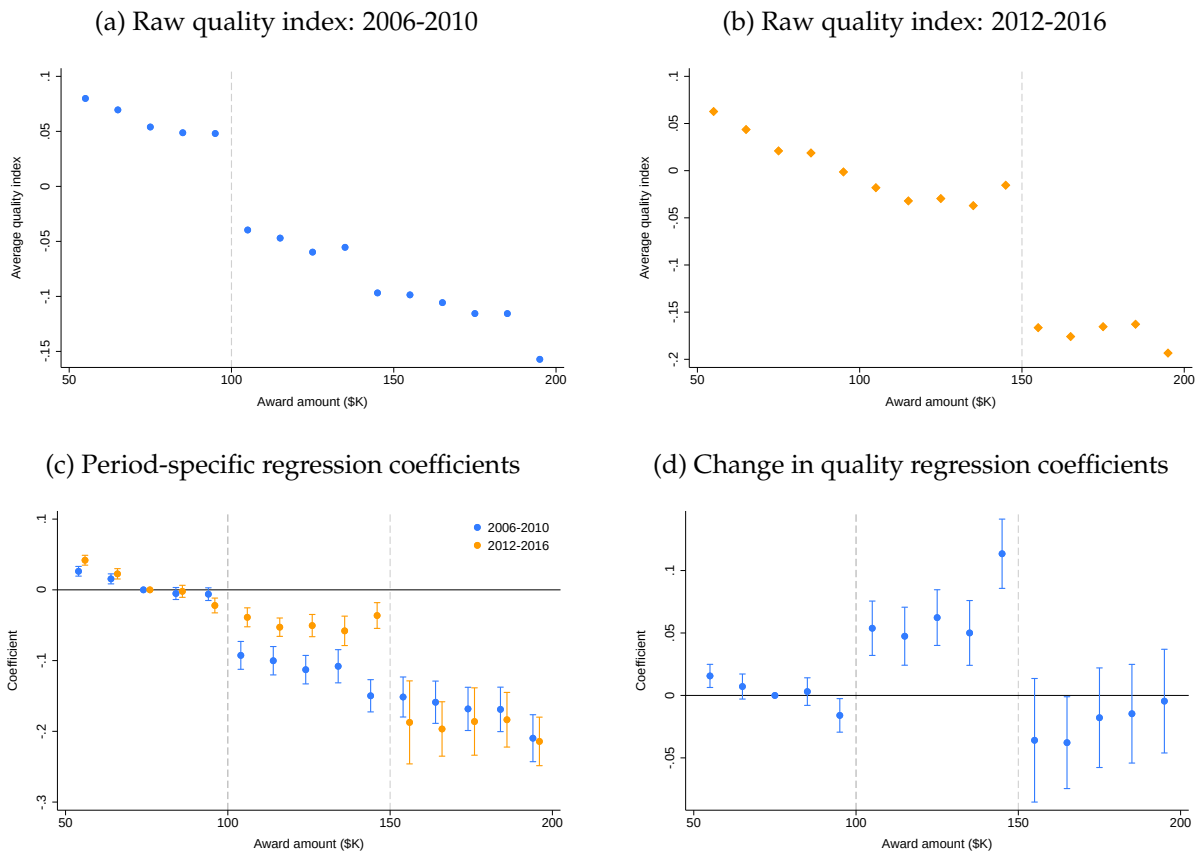


(b) 2015-2016 predicted based on 2013-2014



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 $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 $(\$1000(X-1), \$1000X]$. 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.

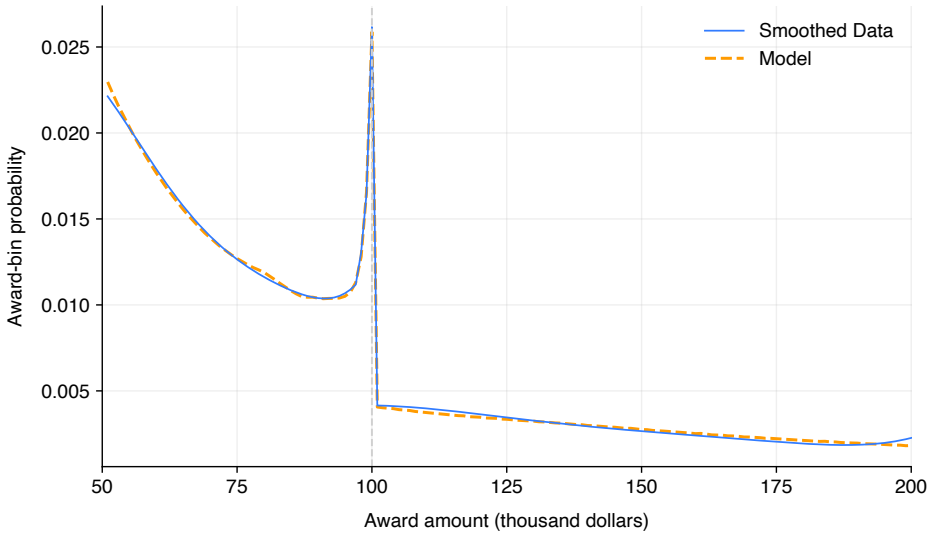
Figure 6: Contract quality



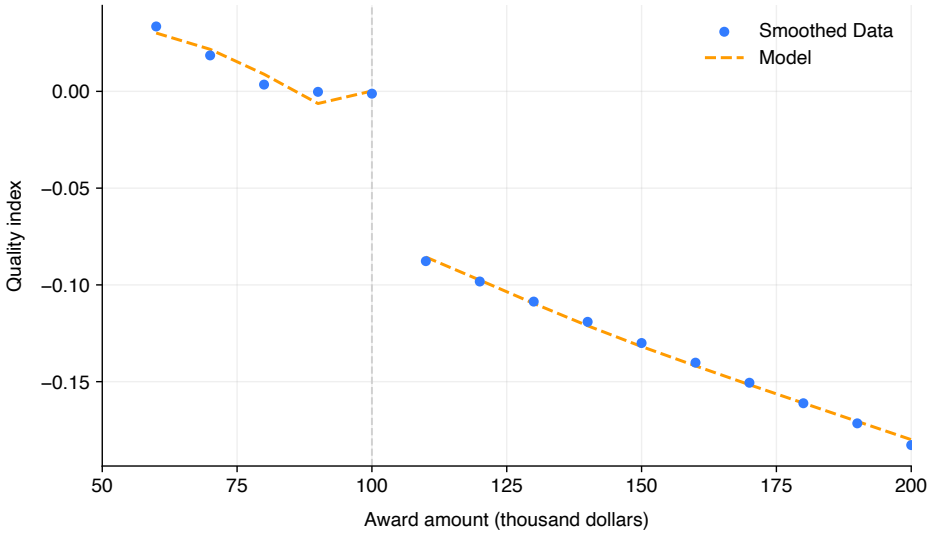
Notes: Panels (a) and (b) respectively show the average contract quality index by award-amount bin in the "pre" (2006-2010) and "post" (2012-2016) periods. Panel (c) shows period-specific regression coefficients from specification (3), using the contract quality index as the dependent variable and no controls. Coefficients can be interpreted as normalized average contract quality by period and bin. Panel (d) presents regression coefficients from specification (4), using the contract quality index as the 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 award-amount bin. 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. Award amounts are discretized into right-inclusive bins of ten-thousand dollars length. The contract quality index averages 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 an award in the last week of the fiscal year. Each proxy is standardized using its mean and standard deviation among pre-period (2006-2010) contracts with award amounts in $(\$50,000, \$300,000]$. The signs are reversed before averaging, so higher index values indicate better execution quality. Standard errors are clustered by awarding office. The vertical dashed lines indicate the simplified acquisition threshold in each period: $100,000 in the pre-period and $150,000 in the post-period.

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

(a) Award distribution



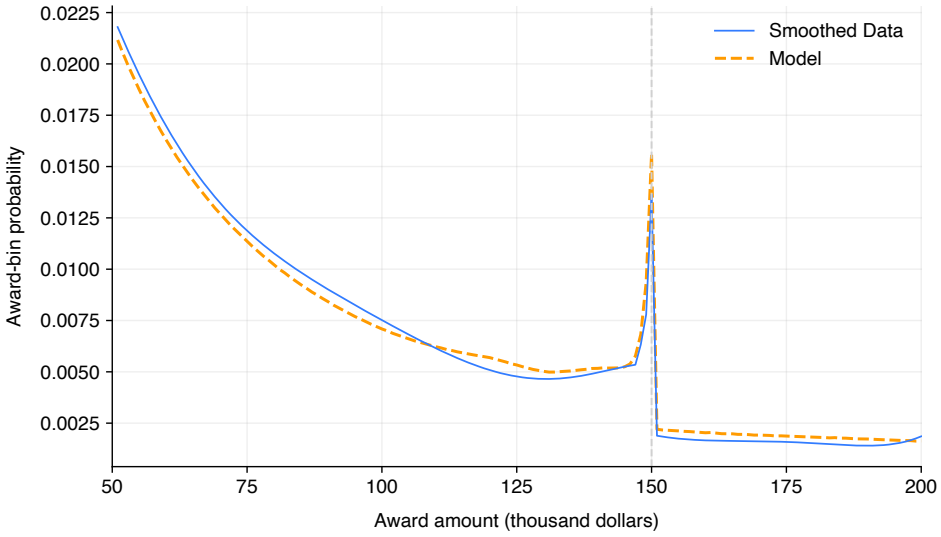
(b) Contract quality



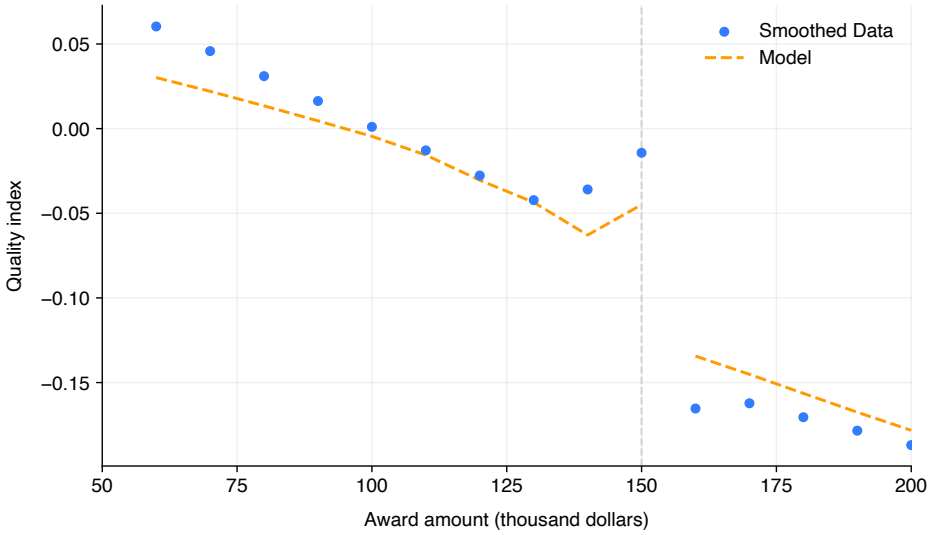
Notes: This figure shows model fit within the estimation sample (2006-2010). The estimation sample consists of non-R&D definitive contracts and purchase orders from the Federal Procurement Data System-Next Generation, with award amounts between $50,000 and $300,000. Both panels compare smoothed data moments with model-simulated moments. Panel (a) uses the normalized frequency in 150 right-inclusive $1,000 award bins between $50,000 and $200,000. Panel (b) uses 15 right-inclusive $10,000 quality bins over the same range. The vertical line indicates the $100,000 simplified acquisition threshold.

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

(a) Award distribution

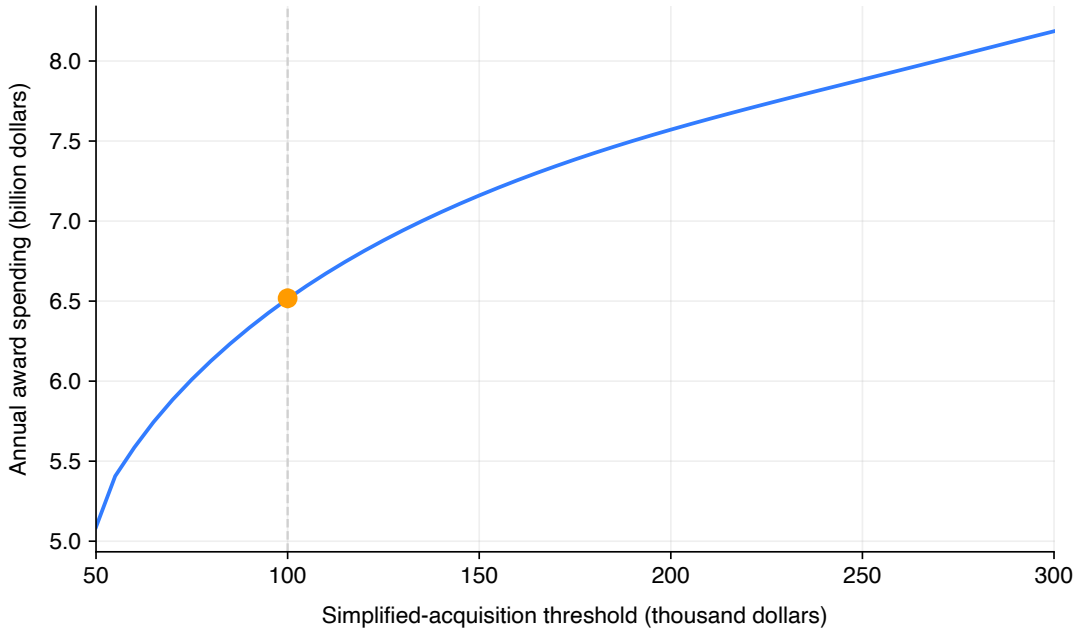


(b) Contract quality



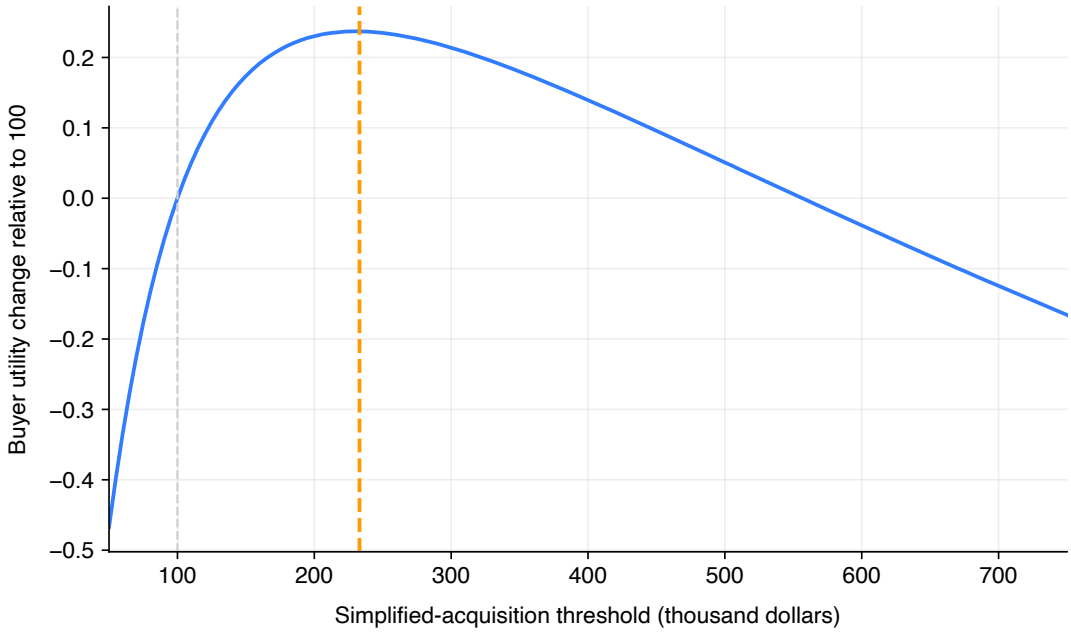
Notes: This figure shows model fit outside the estimation sample (2012-2016). Both panels compare smoothed data moments with model-simulated moments. Panel (a) uses right-inclusive $1,000 award bins and Panel (b) uses right-inclusive $10,000 quality bins. The vertical line indicates the post-period simplified acquisition threshold.

Figure 9: Counterfactual contract spending under different thresholds



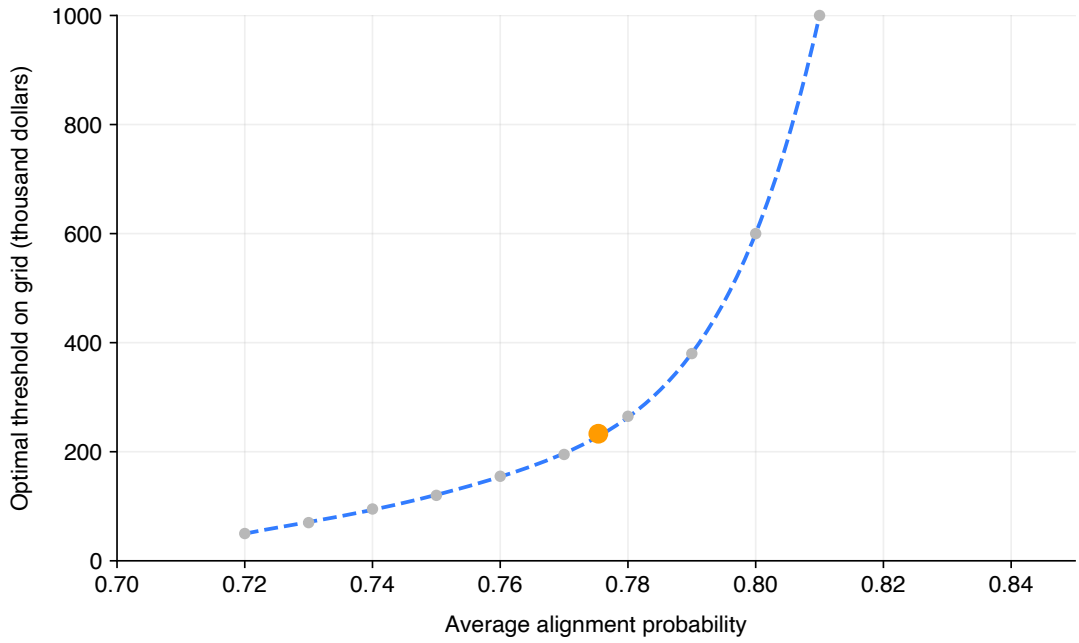
Notes: This figure shows model-implied annual award spending between $50,000 and $300,000 as the simplified acquisition threshold varies. The orange marker is actual average annual pre-period spending at the $100,000 threshold. Dollar amounts are expressed in billions of dollars.

Figure 10: Optimal regulatory threshold



Notes: This figure shows expected buyer utility relative to the $100,000 policy as a function of the simplified acquisition threshold. The dashed light-gray vertical line marks the $100,000 policy value and the dashed orange line marks the maximizing evaluated threshold. Buyer utility is measured in thousands of dollars per procurement opportunity; dollar thresholds are in thousands of dollars.

Figure 11: Optimal threshold as a function of alignment



Notes: This figure shows the buyer-utility-maximizing threshold as a function of the average alignment probability λ. Light-gray dots show point estimates, while the dashed blue line is a fifth-degree polynomial fit. The orange marker shows the estimated λ and its \$233,000 optimum. Dollar thresholds are in thousands of dollars.

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
Veterans 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 awards with expected values, including options, in $(\$5,000, \$5 \text{ million}]$. The right column restricts this sample to $(\$50,000, \$300,000]$.

Table 2: Distribution of contracts by total obligations 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 groups contracts by realized total obligations, summed across all associated contract actions. It includes contracts with total obligations of at least $5,000 and imposes no upper limit. The bracket variable therefore differs from the expected award value used in the threshold analysis. 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 Components	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.)	(0.89)	(6.84)	(3.26)
Excess mass $(\widehat{x})$	27.83	-4.46	-0.62
(s.e.)	(1.21)	(3.38)	(1.28)
Net excess mass $(\widehat{x} - \widehat{m})$	18.24	-1.04	-2.61
(s.e.)	(1.94)	(9.43)	(4.43)
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)	(2)	(3)	(4)	(5)	(6)
	Quality Index	No. of mods	Cost overruns	Delays	Terminated	Last Week
$\in (\$75K, \$100K] \times$ Post	-0.0008	0.0237**	0.1246	-0.2129	-0.0363	-0.1308
	(0.0036)	(0.0113)	(0.4018)	(1.2937)	(0.0790)	(0.1916)
$\in (\$100K, \$150K] \times$ Post	0.0332***	-0.0883***	-3.6086***	-5.4169	-0.1602*	-1.0045***
	(0.0103)	(0.0277)	(1.3457)	(3.6404)	(0.0942)	(0.2631)
$\in (\$150K, \$300K] \times$ Post	-0.0040	0.0306	-2.4679	6.1000	0.0297	0.4874
	(0.0171)	(0.0539)	(2.7865)	(5.3518)	(0.1530)	(0.3260)
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

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 fixed effects for awarding office, 4-digit PSC code, fiscal year, and $1,000-wide award bin. Cost overruns are measured in thousands of dollars and delays in days. Termination and last-week indicators are multiplied by 100, so their coefficients are in percentage points. 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 clustered by awarding agency are shown in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 6: Model Estimates

Parameter	Estimate	Std. error
Valuation location μ_v	-1.3324	(0.0022)
Valuation dispersion σ_v	3.0041	(0.0006)
Discretionary performance magnitude d	0.2531	(0.0007)
Government bargaining weight ϕ	0.4693	(0.0006)
Average alignment probability λ	0.7754	(0.0279)
Baseline bunching friction η	0.7319	(0.0202)
Compliance-cost latent location μ_κ	16.5771	(0.1124)
Compliance-cost latent dispersion σ_κ	10.8272	(0.1136)
Mean threshold buffer s_0	1.1622	(0.0849)
Commercial-size slope b_C	0.3561	(0.7846)
Extra misaligned bunching friction ρ_{BM}	0.9958	(0.1126)
Log cost-ratio shape ℓ_a	0.4697	(0.0024)
Value-dependent alignment slope b_A	0.0774	(0.1425)
Quality level correction u_q	-0.0227	(0.0042)
Quality size correction v_q	0.0080	(0.0169)
Government direct compliance-cost share γ	0.0000	fixed
Mean compliance cost $E[\kappa]$	16.8726	(0.1091)
Cost-ratio shape a	1.5995	(0.0039)
Mean relative cost $E[c/\bar{v}]$	0.8077	(0.0003)

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 (11). 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. Estimation details are provided 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

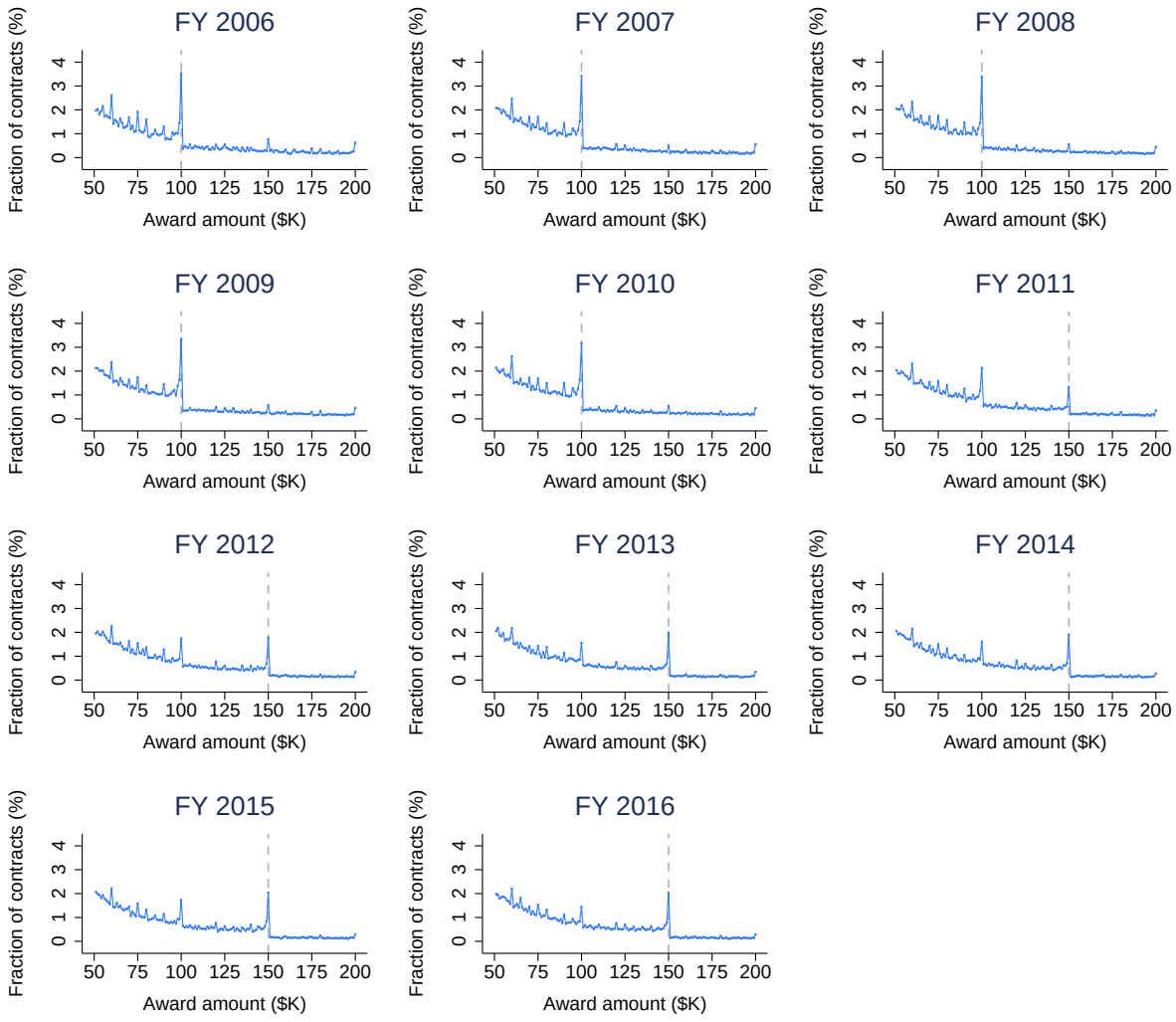
E. Validating Quality Measures Using IT Dashboard

F. Model Details

G. Model Estimation Details

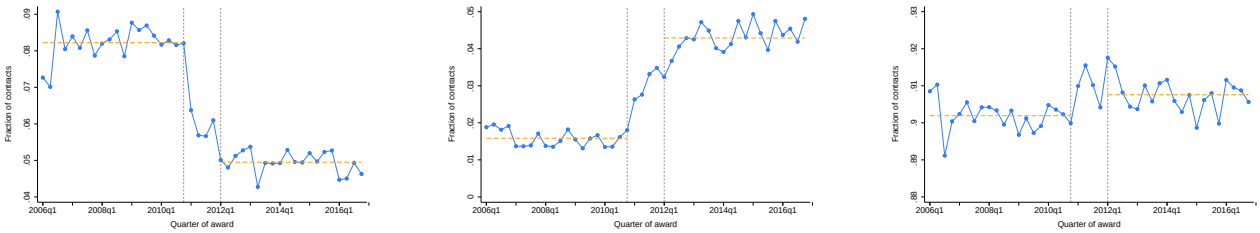
A. Additional Figures

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 the more extensive requirements of nonsimplified acquisition.

Figure A2: Dynamics of Bunching at the Quarterly Level

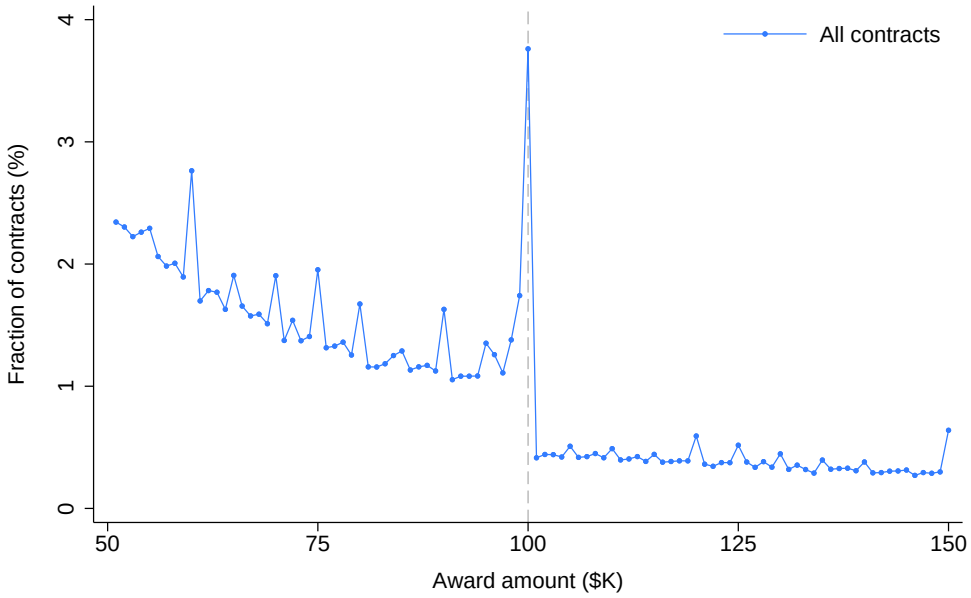


(a) Just Below $100,000 (b) Just Below $150,000 (c) Rest

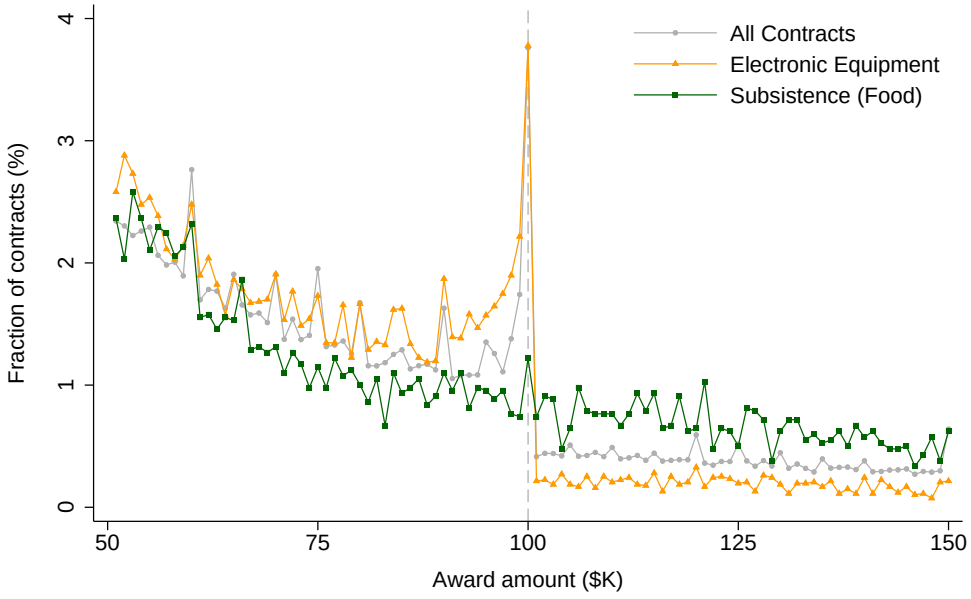
Notes: This figure shows the evolution of the fraction of contracts in different parts of the award amount distribution. The sample is non-R&D definitive contracts and purchase orders from the Federal Procurement Data System-Next Generation, with an award amount between $50,000 and $200,000. Panel (a) shows the fraction of contracts between $95,000 and $100,000. Panel (b) shows the fraction of contracts between $145,000 and $150,000. Panel (c) shows the fraction of contracts that do not fall in either of these two categories. Vertical dashed lines indicate the end of the "pre" period (2010q4) and the beginning of the "post" period (2012q1), leaving the four quarters of fiscal year 2011 in between. Horizontal dash lines show the mean fraction of contracts within each period.

Figure A3: Distribution of contract awards by product categories, 2006-2010

(a) All contracts



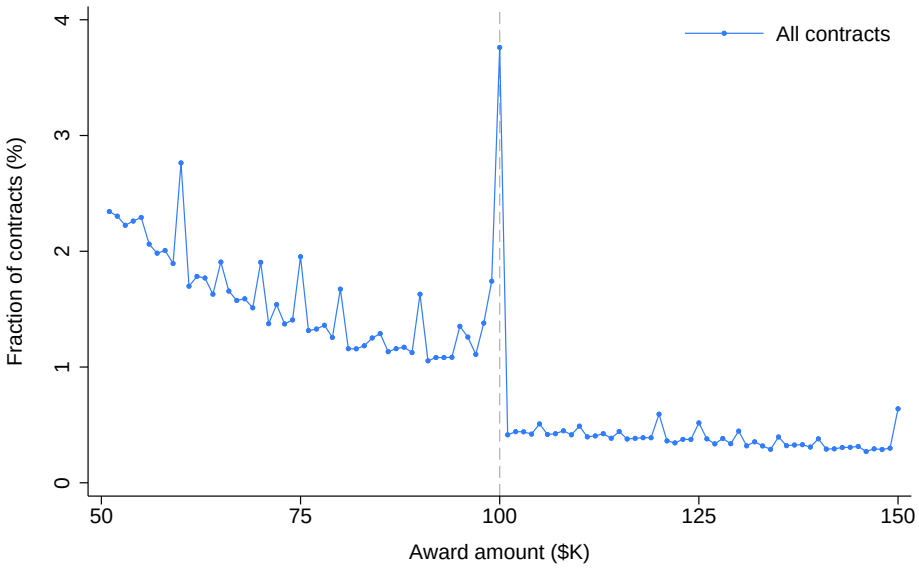
(b) Selected product categories



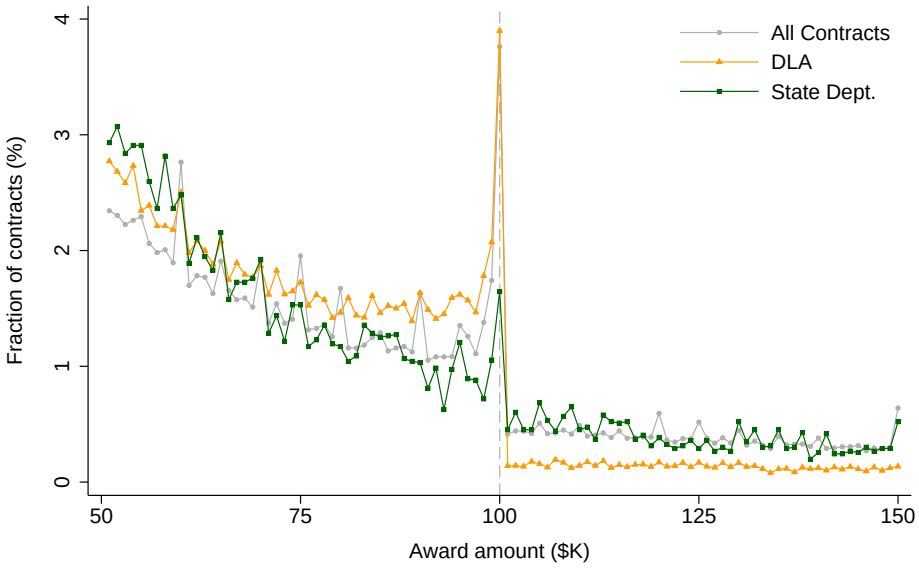
Notes: This figure shows contract awards frequency distributions in 2006-2010, by selected product categories. The sample is non-R&D definitive contracts and purchase orders from the Federal Procurement Data System-Next Generation, with an award amount between $50,000 and $150,000. Panel (a) features 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 the more extensive requirements of nonsimplified acquisition.

Figure A4: Distribution of contract awards by awarding agency, 2006-2010

(a) All contracts



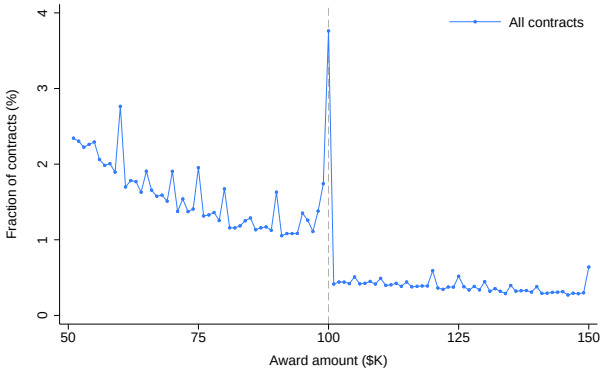
(b) Selected agencies



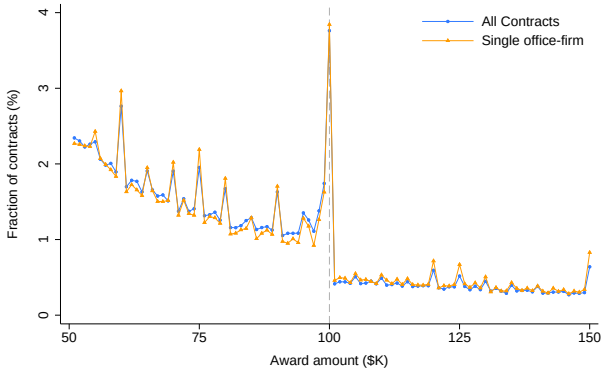
Notes: This figure shows contract awards frequency distributions in 2006-2010, by selected awarding agencies. The sample is non-R&D definitive contracts and purchase orders from the Federal Procurement Data System-Next Generation, with an award amount between $50,000 and $150,000. Panel (a) features 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 the more extensive requirements of nonsimplified acquisition.

Figure A5: Distribution of contract awards in selected samples, 2006-2010

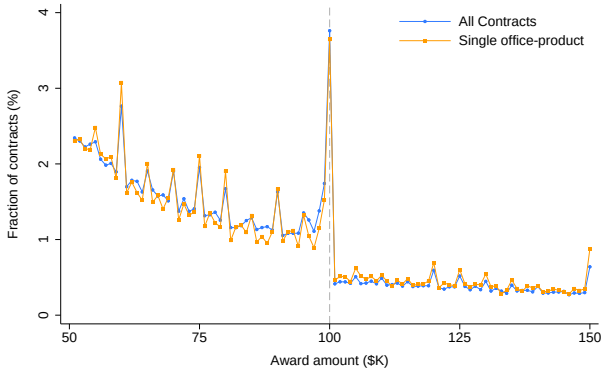
(a) All contracts in FY2006-2010



(b) Versus single-office-firm contracts



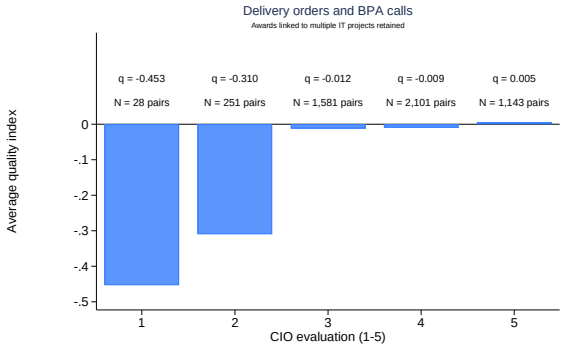
(c) Versus single-office-product contracts



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 the more extensive requirements of nonsimplified acquisition. 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 A6: Validating quality measures

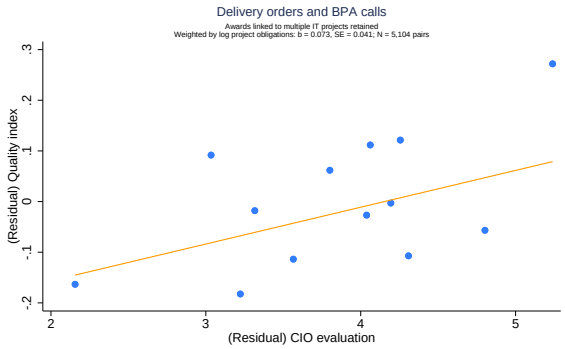
(a) Mean quality index by CIO evaluation



(b) Controlling for agency and year



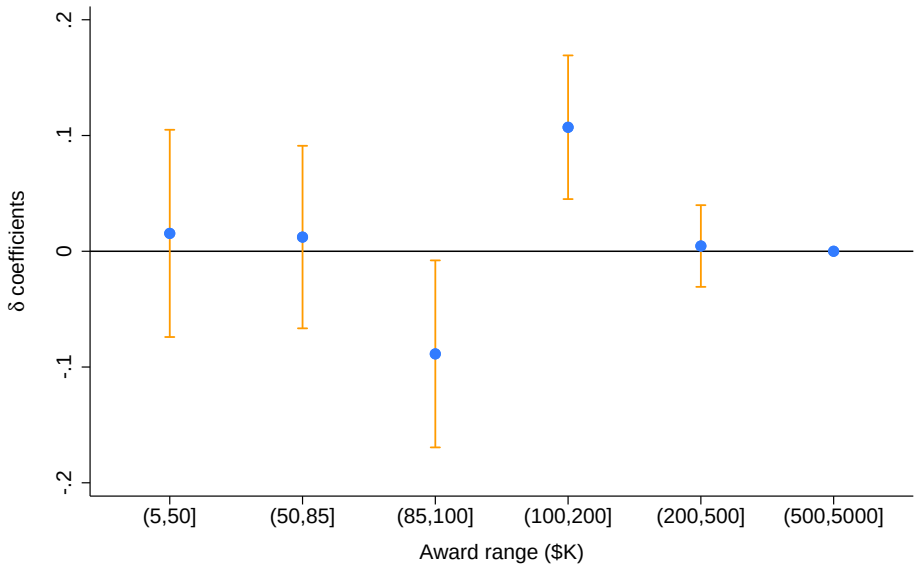
(c) Weighting by the size of the project



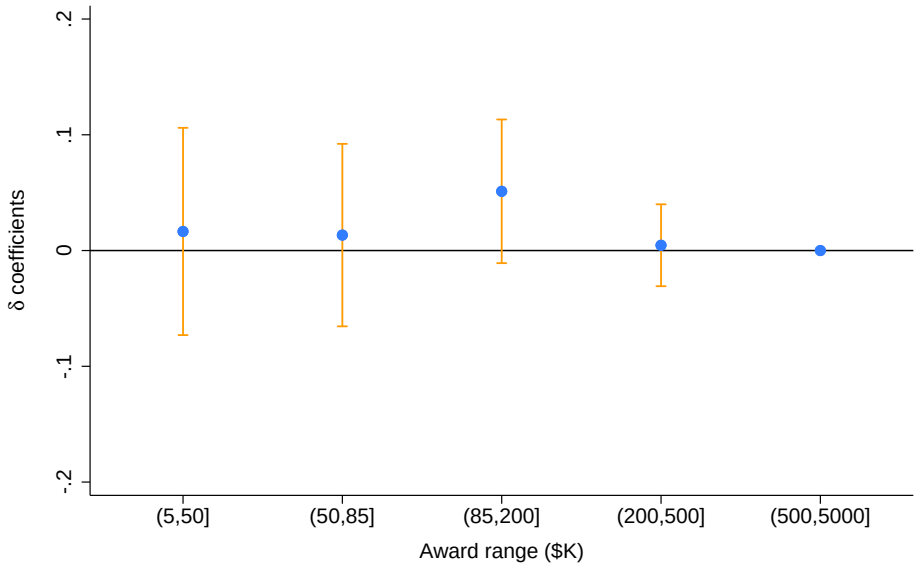
Notes: This figure shows the relationship between the execution-based quality proxies used in the main analysis and Chief Information Officer (CIO) evaluations from the IT Dashboard. The sample contains 5,104 project–award pairs obtained by matching IT Dashboard annual files from 2013 through 2018 to 4,657 distinct delivery orders and BPA calls, covering 506 IT projects. An award linked to multiple projects appears once for each project–award link. Panel (a) presents the unconditional mean award-level quality index by CIO evaluation. Panel (b) presents a binned scatter plot controlling for project-agency and Dashboard file-year fixed effects. Panel (c) repeats this exercise weighting observations by the natural logarithm of total obligations for awards linked to the project. Standard errors associated with the fitted slopes are clustered at the IT-project level.

Figure A7: Change in number of contracts in high vs. low bunching offices, by award size

(a) Baseline



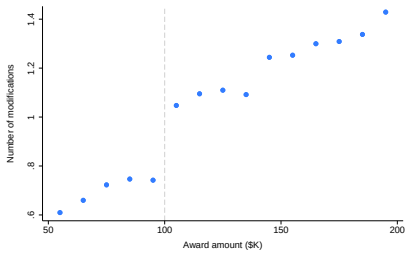
(b) Combining award ranges around the simplified acquisition threshold



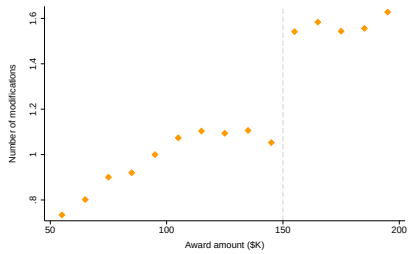
Notes: This figure presents coefficient estimates and their 95% confidence intervals for two triple-difference (DDD) specifications. The sample consists of non-R&D definitive contracts and purchase orders from the Federal Procurement Data System-Next Generation in fiscal years 2006 through 2016, with award amounts between $5,000 and $5,000,000. The data are aggregated by contracting office, fiscal year, and $1,000-wide award-amount bin. The dependent variable is the log number of contracts in each office-year-bin; the regression includes only observations with a positive count. The regressors include fixed effects for office, fiscal year, and award-amount bin, as well as the lower-order interactions described in Appendix C. The plotted estimates are the coefficients on the triple interaction between a high-bunching indicator, a post indicator, and indicators for award-amount ranges. The high-bunching indicator equals one for offices with above-median pre-2011 bunching, measured by the share of total awards with amounts between $99,000 and $100,000. The post indicator equals zero in 2010 or earlier and one in 2012 or later; fiscal year 2011 is excluded. Standard errors are clustered by contracting office.

Figure A8: Raw quality sub-measures around the threshold

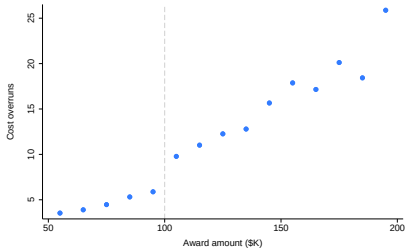
(a) Number of modifications: 2006-2010



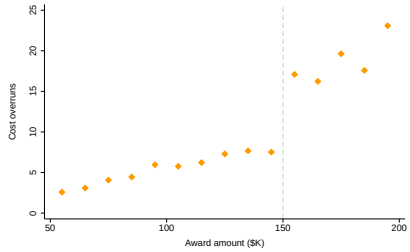
(b) Number of modifications: 2012-2016



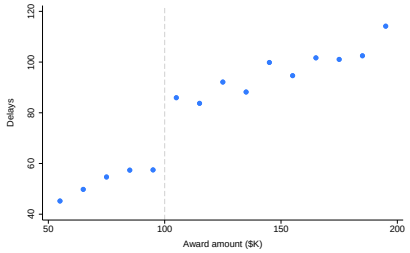
(c) Cost overruns: 2006-2010



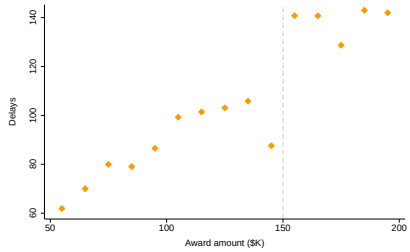
(d) Cost overruns: 2012-2016



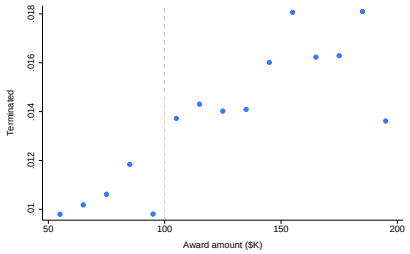
(e) Delays: 2006-2010



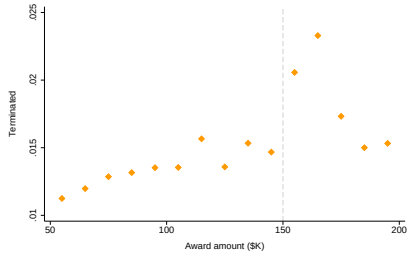
(f) Delays: 2012-2016



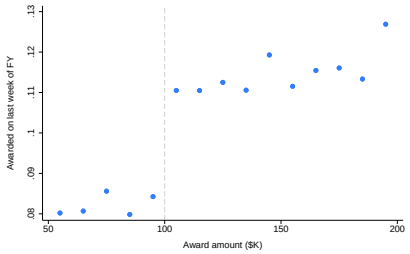
(g) Terminated: 2006-2010



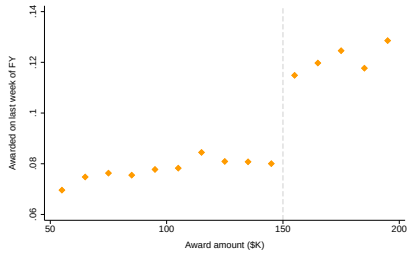
(h) Terminated: 2012-2016



(i) Last week: 2006-2010

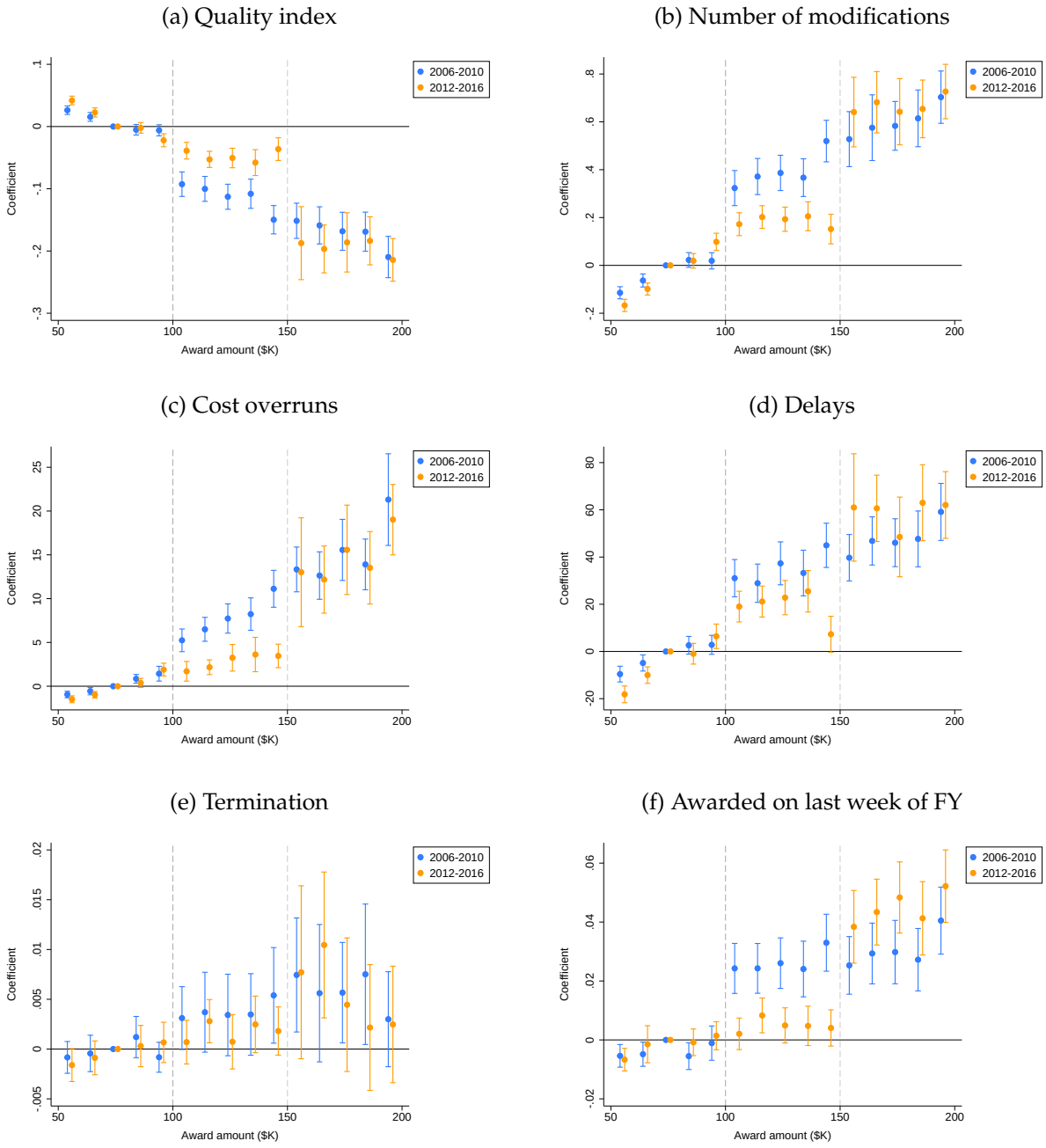


(j) Last week: 2012-2016



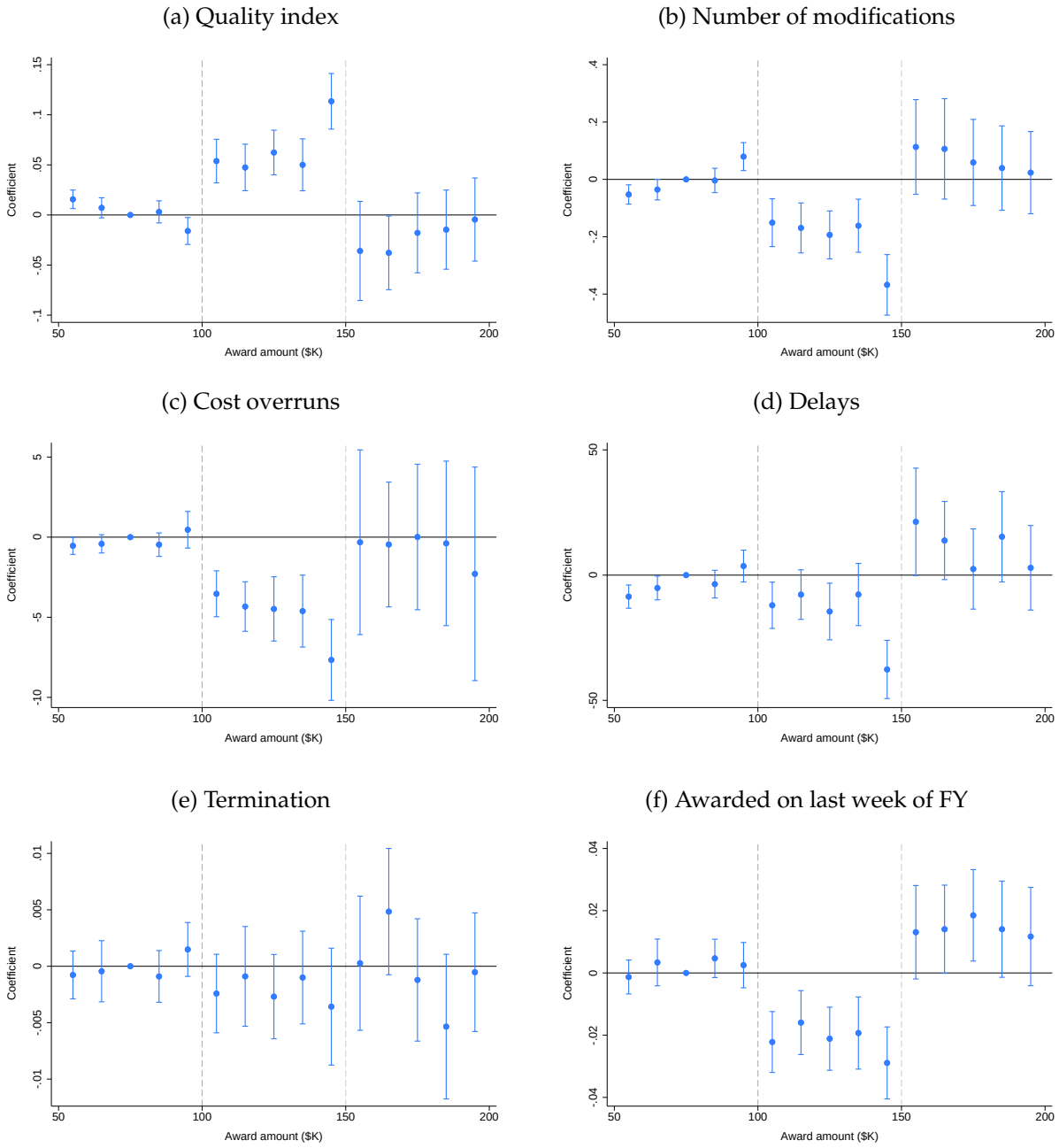
Notes: Points show mean outcomes in right-inclusive $10,000 award bins, plotted at bin midpoints. The sample consists of non-R&D definitive contracts and purchase orders from the Federal Procurement Data System-Next Generation with award amounts in $(\$50,000, \$200,000]$, in 2006-2010 (left) and 2012-2016 (right). Outcomes are within-scope modifications (counts), within-scope cost overruns (thousands of dollars), within-scope delays (days), and the shares terminated or awarded in the last week of the fiscal year. These are raw means, without standardization or sign reversal. Dashed lines mark the $100,000 and $150,000 simplified acquisition thresholds, respectively.

Figure A9: Contract quality by each performance proxy



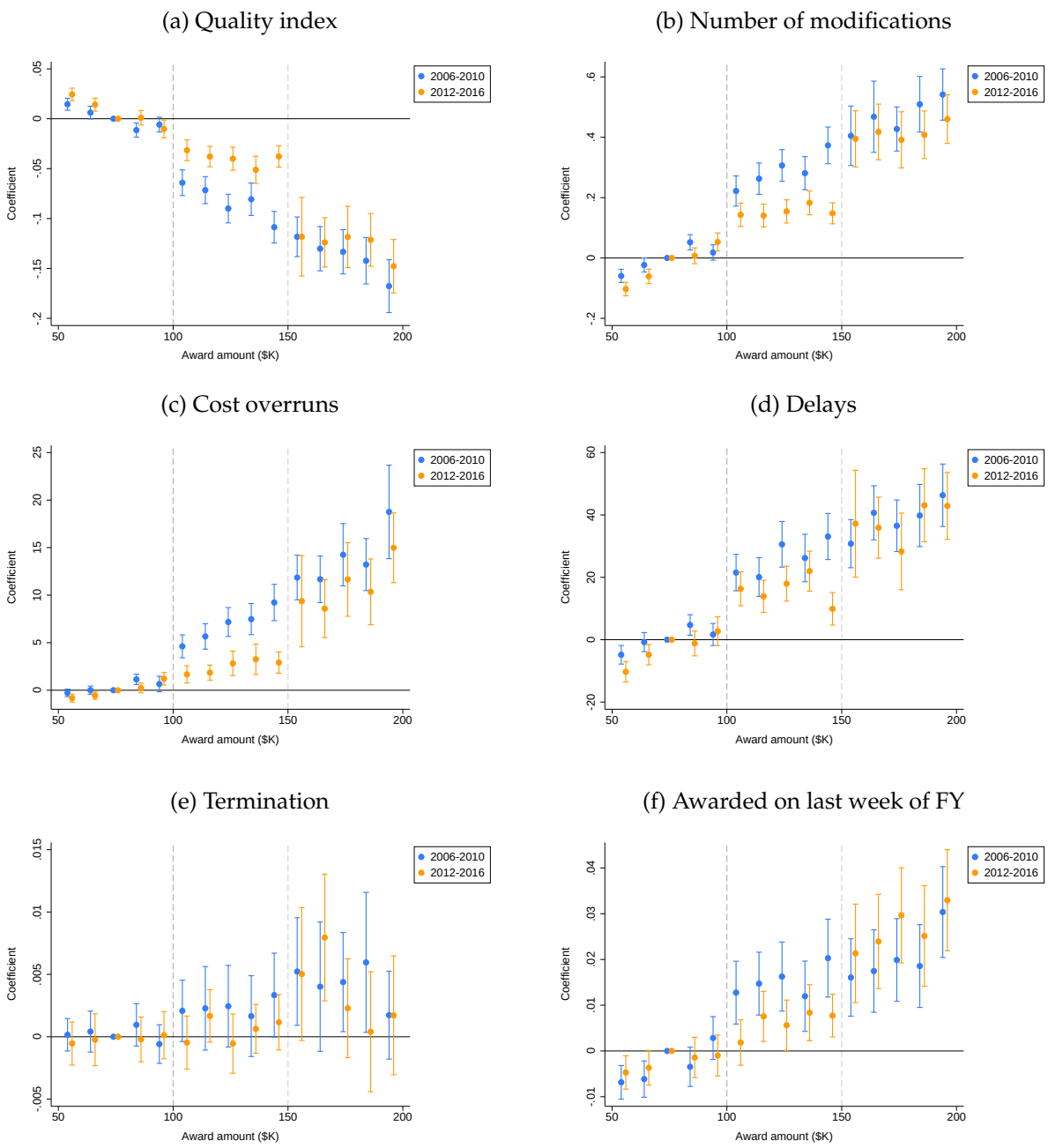
Notes: This figure shows period-specific regression coefficients from specification (3), using different measures of contract performance as dependent variables and no controls. Coefficients can be interpreted as normalized average outcomes by period and bin. The sample consists of non-R&D definitive contracts and purchase orders from the Federal Procurement Data System-Next Generation, with award amounts between $50,000 and $300,000. Panel (a) shows the quality index, panel (b) the number of within-scope modifications, panel (c) cost overruns, panel (d) delays, panel (e) the fraction of terminated contracts, and panel (f) 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. Standard errors are clustered by awarding office. The vertical dashed lines indicate the simplified acquisition threshold in each period: $100,000 in the pre-period (2006-2010) and $150,000 in the post-period (2012-2016).

Figure A10: Change in contract quality by each performance proxy



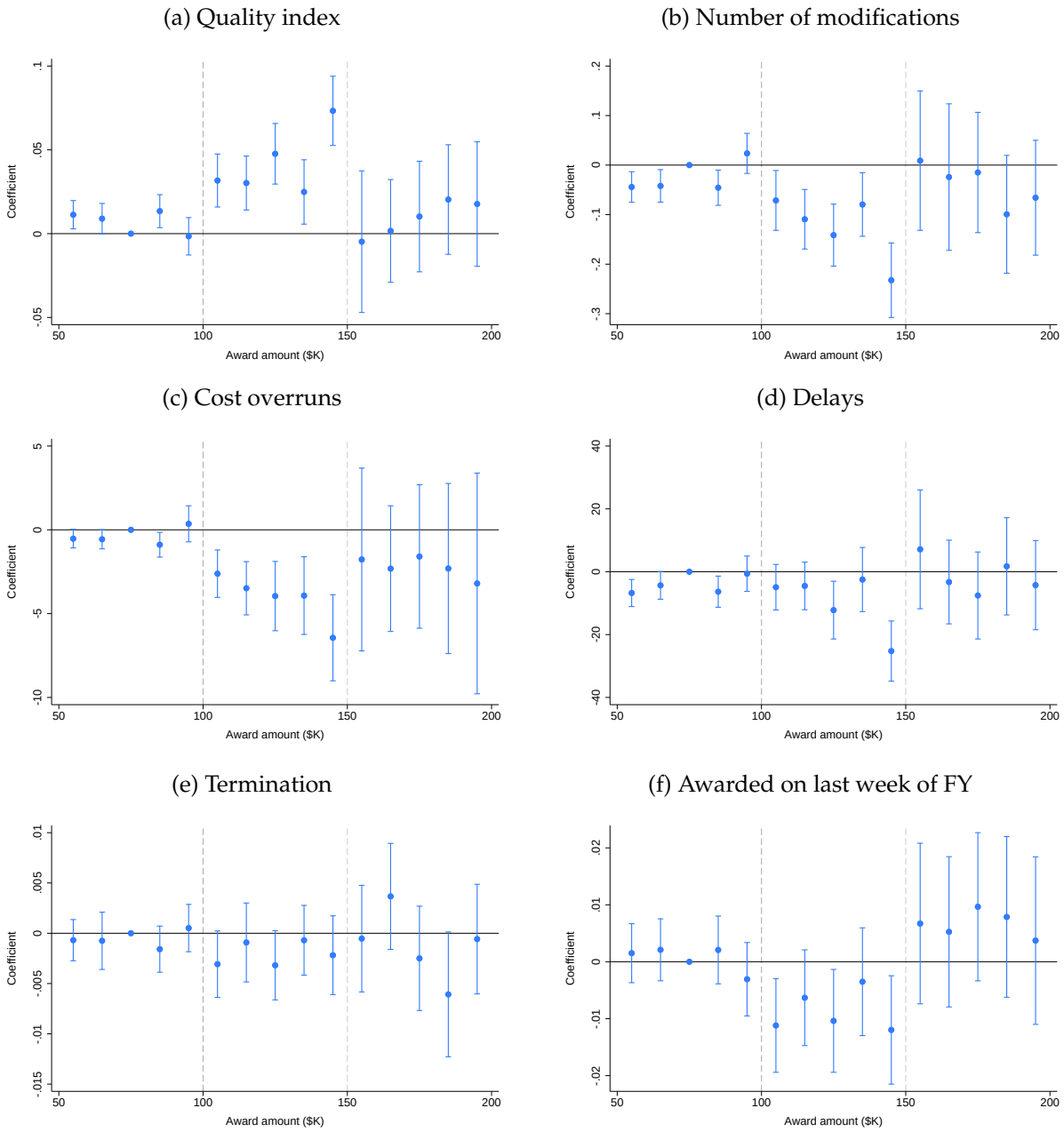
Notes: This figure shows regression coefficients from specification (4), using different measures of contract performance as dependent variables and no controls. Coefficients can be interpreted as changes in average performance between the pre and post periods by award-amount bin. The sample consists 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. Panel (a) shows the quality index, panel (b) the number of within-scope modifications, panel (c) cost overruns, panel (d) delays, panel (e) the fraction of terminated contracts, and panel (f) 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. Standard errors are clustered by awarding office. The vertical dashed lines indicate the simplified acquisition threshold in each period: $100,000 in the pre-period (2006-2010) and $150,000 in the post-period (2012-2016).

Figure A11: Contract quality by performance proxy, with controls



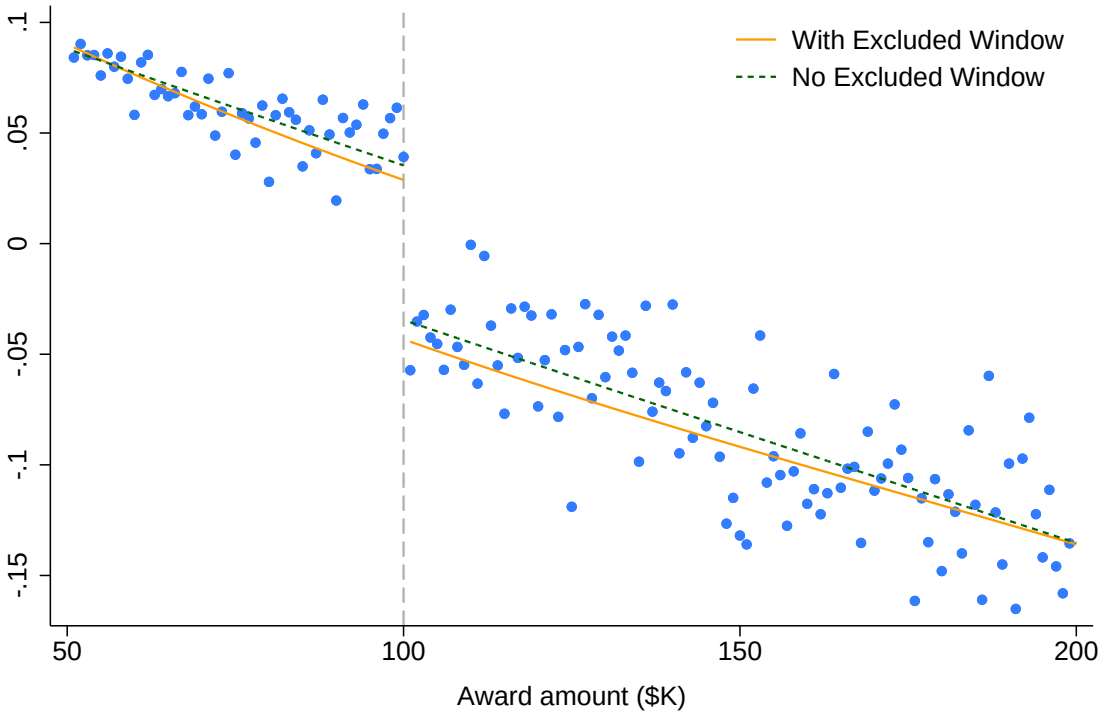
Notes: This figure replicates Figure A9 with controls for product code, awarding office, and fiscal year. Coefficients can be interpreted as conditional normalized average outcomes by period and bin. The sample consists of non-R&D definitive contracts and purchase orders from the Federal Procurement Data System-Next Generation, with award amounts between $50,000 and $300,000. Panel (a) shows the quality index, panel (b) the number of within-scope modifications, panel (c) cost overruns, panel (d) delays, panel (e) the fraction of terminated contracts, and panel (f) 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. Standard errors are clustered by awarding office. The vertical dashed lines indicate the simplified acquisition threshold in each period: $100,000 in the pre-period (2006-2010) and $150,000 in the post-period (2012-2016).

Figure A12: Change in contract quality by each performance proxy, with controls



Notes: This figure replicates Figure A10 with controls for product code, awarding office, fiscal year, and award-amount bin. Coefficients can be interpreted as conditional changes in average performance between the pre and post periods by award-amount bin. 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. Panel (a) shows the quality index, panel (b) the number of within-scope modifications, panel (c) cost overruns, panel (d) delays, panel (e) the fraction of terminated contracts, and panel (f) 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. Standard errors are clustered by awarding office. The vertical dashed lines indicate the simplified acquisition threshold in each period: $100,000 in the pre-period (2006-2010) and $150,000 in the post-period (2012-2016).

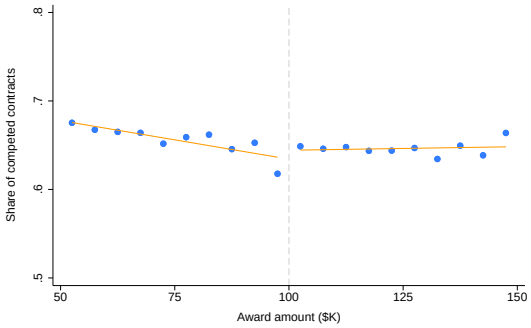
Figure A13: RD graph for quality index, with and without excluded window



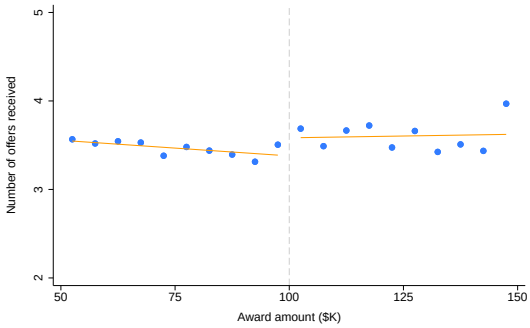
Notes: This figure shows average contract quality across the award-amount distribution. The blue dots represent the average contract quality index for each $1,000 award-amount bin in the "pre" (2006-2010) period. The green dashed line corresponds to fitted values from a regression of the quality index on a third-degree polynomial of award amounts and an indicator for bins above the simplified acquisition threshold ($100,000). The orange solid line corresponds to fitted values from the same regression, but excluding bins between $85,000 and $138,000. The specification is given by Equations (5) and (6). The excluded window is determined following Kleven and Waseem (2013). 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. Award amounts are discretized into right-inclusive bins of one-thousand dollars length. The contract quality index averages 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 an award in the last week of the fiscal year. Each proxy is standardized using its mean and standard deviation among pre-period (2006-2010) contracts with award amounts in $(\$50,000, \$300,000]$. The signs are reversed before averaging, so higher index values indicate better execution quality.

Figure A14: Competition, contractor characteristics, and quality around the threshold

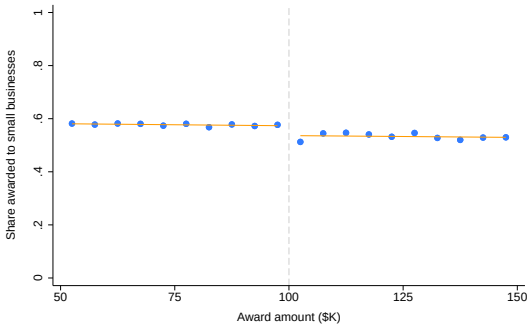
(a) Share of competed awards



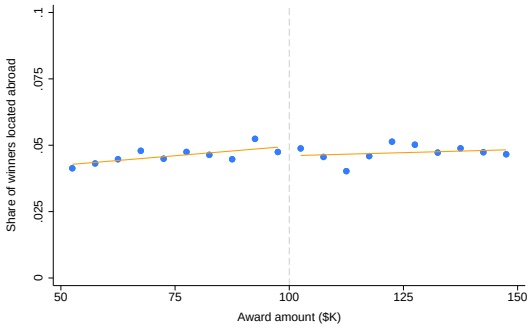
(b) Average number of offers



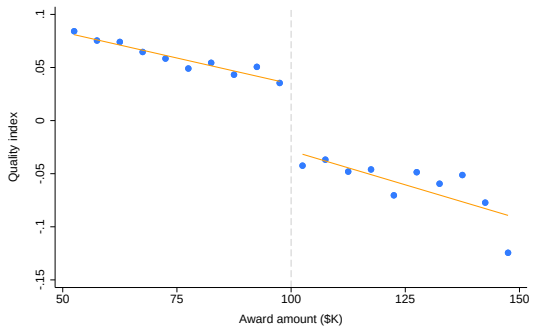
(c) Small-business winners



(d) Winners located abroad

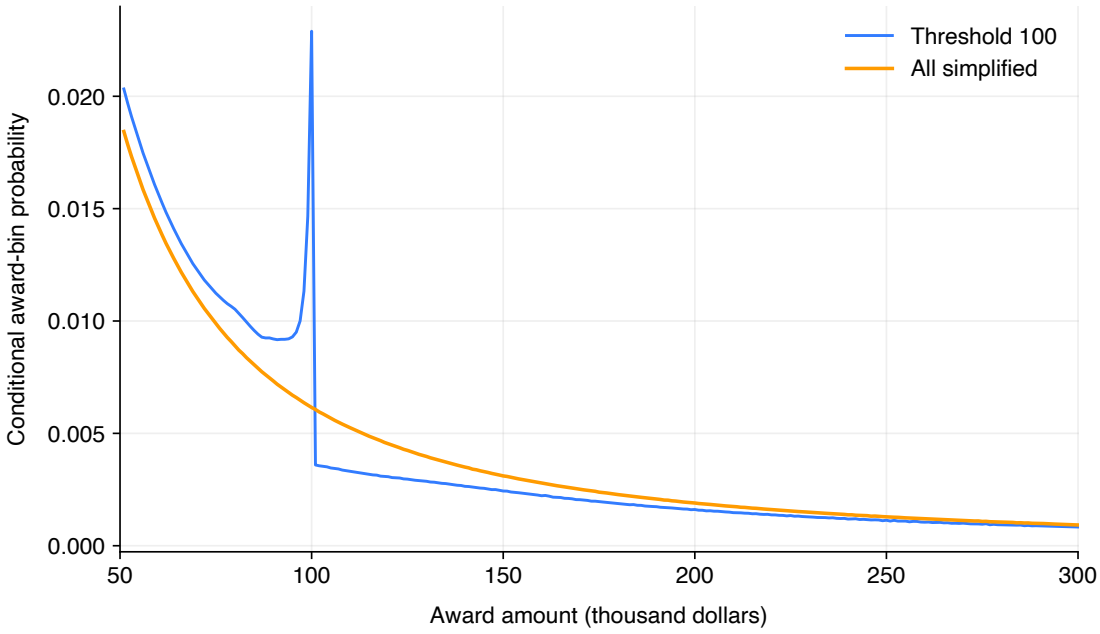


(e) Contract quality index



Notes: This figure shows binned scatter plots for two competition measures, two contractor characteristics, and the contract quality index around the simplified acquisition threshold. Blue dots represent average values in right-inclusive $5,000 award-amount bins in the pre-period (2006–2010); orange lines are separate linear fits on either side of $100,000. The sample consists of non-R&D definitive contracts and purchase orders from the Federal Procurement Data System-Next Generation with award amounts in $(\$50,000, \$300,000]$; the graphs display amounts up to $150,000. Bin means are adjusted for round-number effects. The number of offers is measured among competed awards and winsorized at the 99th percentile. Small-business status is the recorded award-time classification. Foreign location is based on the contractor's recorded country, not ownership. The contract quality index averages 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 an award in the last week of the fiscal year. Each proxy is standardized using its mean and standard deviation among pre-period contracts with award amounts in $(\$50,000, \$300,000]$. Signs are reversed before averaging, so higher index values indicate better execution quality.

Figure A15: Contract distribution with simplified acquisition for all purchases



Notes: This figure compares the model-simulated award distribution at the $100,000 baseline with a counterfactual in which every purchase is eligible for simplified acquisition. Each distribution is conditional on an award between $50,000 and $300,000. Award amounts are in thousands of dollars.

B. Additional Tables

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

		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 $(\hat{x} - \hat{m})$ as a function of key parameters of the bunching estimation procedure. The estimation procedure is described in Section 3.2. Each number is an estimate of $(\hat{x} - \hat{m})$ for a given polynomial degree p in Equation (1) and lower and upper bounds of the excluded region $(\underline{R}, \overline{R}]$. Both bounds are measured in thousand dollars. The baseline specification uses $p = 3$, $\underline{R} = 50$, and $\overline{R} = 300$. The baseline estimate (= 18.24) appears at the center of panel (b) and coincides with the estimate in Table 4.

Table B2: Summary statistics of IT Dashboard–federal awards sample

(a) Number of projects, awards, and total obligations by agency

Department or Agency	Number of IT Projects	Number of Awards	Total Obligated ($M)
Agriculture	45	277	1,879.0
Commerce	24	587	2,090.1
Defense	52	454	6,149.3
Health and Human Services	85	393	13,688.4
Interior	35	138	440.7
State	10	114	3,607.0
Treasury	30	74	2,271.5
General Services Administration	22	62	1,265.4
Homeland Security	89	1,240	16,576.2
Veterans Affairs	25	797	8,191.0
All Others	135	1,005	7,005.9
Total	552	5,141	63,164.5

(b) Summary statistics of quality measures

Quality measure	Mean	s.d.	p10	p90
Award-level (Source: USAspending; N=4,663)				
Number of modifications	6.86	10.81	0.00	18.00
Cost overruns ($M)	1.22	5.09	0.00	1.89
Delays (days)	335.04	498.43	0.00	1126.00
Termination (0-1)	0.01	0.09	0.00	0.00
Last week of FY (0-1)	0.11	0.32	0.00	1.00
Quality index	0.03	0.95	-1.09	0.71
Project-level (Source: IT Dashboard; N=547)				
CIO evaluation (1-5)	3.86	0.91	3.00	5.00

Notes: This table presents summary statistics for federal IT projects from the IT Dashboard annual files for 2013 through 2018 that can be matched to delivery orders and BPA calls in the federal award data. Panel (a) counts each IT project and each award once and reports total obligated dollars by department or agency. Panel (b) summarizes the five execution-based quality measures and the standardized quality index among distinct awards with complete outcomes, and CIO evaluations among distinct matched projects for which a rating is observed. The quality index is normalized over the broader matched sample containing all four actual-award types before the delivery-order/BPA-call restriction is imposed.

Table B3: The effect of manipulation on the use of simplified procedures

Period	2006–2010	2012–2016
Threshold	\$100K	\$150K
Simplified Procedures	0.060	0.100
	(0.004)	(0.006)
Average outcome	0.590	0.631
No. of obs.	250	250

This table presents intent-to-treat (ITT) estimates of the effect of manipulation on the use of simplified procedures for the "pre" (2006-2010) and "post" (2012-2016) periods. 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, aggregated to one-thousand dollar right-inclusive bins. An observation is a contract award amount bin. The methodology follows Diamond and Persson (2016) and Coviello et al. (2026), and has three steps. First, a bunching estimation procedure is implemented on the frequency distribution of contract awards, obtaining counterfactual counts in the absence of manipulation for each bin. Second, we estimate Equation (5) using an indicator for simplified acquisition procedures as the outcome, and obtain counterfactual outcomes in the absence of manipulation for each bin according to Equation (6). The third step is to compute ITT estimates as the difference between average observed and counterfactual outcomes, according to Equation (7). The simplified-procedure indicator is measured on a 0–1 scale. Bootstrapped standard errors are shown in parentheses.

Table B4: The effect of manipulation on contract quality, ITT and LATE estimates

Panel A: 2006–2010 (Threshold = $100K)

	Quality Index	Modifications	Cost Overruns	Delays	Terminations	Last Week
ITT	0.027	-0.075	-2.354	-2.332	-0.003	-0.008
	(0.006)	(0.026)	(0.830)	(2.329)	(0.001)	(0.002)
LATE	0.448	-1.241	-38.982	-38.611	-0.050	-0.138
	(0.095)	(0.424)	(13.900)	(38.723)	(0.016)	(0.041)
Average outcome	-0.097	1.251	18.436	97.710	0.015	0.110
No. of obs.	250	250	250	250	250	250

Panel B: 2012–2016 (Threshold = $150K)

	Quality Index	Modifications	Cost Overruns	Delays	Terminations	Last Week
ITT	0.031	-0.120	-2.003	-17.998	0.003	-0.008
	(0.008)	(0.026)	(1.015)	(3.895)	(0.002)	(0.004)
LATE	0.310	-1.202	-19.987	-179.640	0.034	-0.082
	(0.075)	(0.269)	(10.286)	(39.585)	(0.016)	(0.036)
Average outcome	-0.126	1.427	15.638	125.983	0.018	0.103
No. of obs.	250	250	250	250	250	250

Notes: This table presents intent-to-treat (ITT) and local average treatment effect (LATE) estimates of the effect of manipulation on contract quality. Panel A presents estimates for the "pre" (2006-2010) period and Panel B presents results for the "post" (2012-2016) period. 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, aggregated to one-thousand dollar right-inclusive bins. An observation is a contract award amount bin. The methodology follows Diamond and Persson (2016) and Coviello et al. (2026), and has three steps. First, a bunching estimation procedure is implemented on the frequency distribution of contract awards, obtaining counterfactual counts in the absence of manipulation for each bin. Second, we estimate Equation (5) using contract quality measures as outcomes, and obtain counterfactual outcomes in the absence of manipulation for each bin according to Equation (6). The third step is to compute ITT and LATE estimates. ITT estimates are the difference between average observed and counterfactual outcomes, according to Equation (7). LATE estimates are equal to ITT estimates, divided by the ITT estimates on simplified acquisition procedures, presented in Appendix Table B3. Cost overruns are measured in thousands of dollars, delays in days, and termination and last-week indicators on a 0–1 scale. Quality-index effects are in index points. Bootstrapped standard errors are shown in parentheses.

Table B5: Quality changes at the threshold, with and without excluded window

	Quality Index	Modifications	Cost Overruns	Delays	Terminations	Last Week
RD	-0.070	0.243	2.760	21.481	0.002	0.027
	(0.006)	(0.023)	(0.587)	(2.847)	(0.001)	(0.003)
donut-RD	-0.072	0.218	3.905	12.367	0.005	0.032
	(0.016)	(0.069)	(1.996)	(7.183)	(0.003)	(0.007)
Average outcome	-0.097	1.251	18.436	97.710	0.015	0.110
No. of obs.	250	250	250	250	250	250

Notes: This table presents estimates of the change in contract quality at the simplified acquisition threshold, with and without an excluded window around the threshold. The sample consists of non-R&D definitive contracts and purchase orders from the Federal Procurement Data System-Next Generation in 2006-2010, with award amounts between $50,000 and $300,000, aggregated to one-thousand dollar right-inclusive bins. An observation is a contract award amount bin. Each coefficient comes from a separate regression, where the dependent variable is a contract quality measure indicated in the column header. Coefficient estimates correspond to an indicator for bins above the $100,000 threshold. In both cases, the regression includes a common third-degree polynomial in the award amount and an indicator for bins above the threshold. Coefficients in the first line ("RD") are obtained by including all bins in the sample, while coefficients in the second line ("donut-RD") are obtained by excluding bins between $85,000 and $138,000. This excluded window is determined in the first step of the bunching estimation using the methodology by Diamond and Persson (2016) and Coviello et al. (2026). Cost overruns are measured in thousands of dollars, delays in days, and termination and last-week indicators on a 0–1 scale. Quality-index effects are in index points. Robust standard errors are shown in parentheses.

Table B6: CIO evaluations and award execution quality: sample robustness

	Delivery orders and BPA calls		All award types
	Baseline	Exclude shared awards	Exclude shared awards
Panel A: Unweighted			
CIO evaluation	0.069	0.065	0.021
	(0.040)	(0.037)	(0.036)
Project–award pairs	5,104	4,297	5,369
IT projects	506	459	578
Distinct awards	4,657	4,297	5,369
Panel B: Weighted by log project obligations			
CIO evaluation	0.073	0.067	0.021
	(0.041)	(0.038)	(0.038)
Project–award pairs	5,104	4,297	5,368
IT projects	506	459	577
Distinct awards	4,657	4,297	5,368

Notes: This table reports regressions of the award-level quality index on the CIO evaluation. An observation is a project–award pair. The baseline sample contains delivery orders and BPA calls and retains awards linked to multiple IT projects. The second column excludes awards linked to multiple IT projects while retaining delivery orders and BPA calls. The third column applies the same exclusion and adds definitive contracts and purchase orders. All specifications include project-agency and Dashboard file-year fixed effects. Standard errors clustered at the IT-project level appear in parentheses. Panel B weights project–award pairs by the natural logarithm of project obligations. Baseline project obligations are computed over all four matched award types. For the two restricted samples, project obligations are recomputed after awards linked to multiple IT projects are removed.

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 change, across contracting offices, and across contract award sizes.

Formally, I estimate the following specification:

$$\log n_{kbt} = \alpha_k + \alpha_b + \alpha_t + \rho\, HighBunch_k \times Post_t + \sum_{q \in Q} \beta_q \cdot \mathbf{1}[b \in q] \times HighBunch_k + \sum_{q \in Q} \gamma_q \cdot \mathbf{1}[b \in q] \times Post_t + \sum_{q \in Q} \delta_q \cdot \mathbf{1}[b \in q] \times HighBunch_k \times Post_t + \epsilon_{kbt} \quad (12)$$

where n_{kbt} is the number of awards made by contracting office k in award-amount bin b and fiscal year $t \in \{2006, ..., 2010, 2012, ..., 2016\}$. Bins are right-inclusive and have \$1,000 length, so bin $b = 6$ includes awards in $(\$5,000, \$6,000]$. The regression includes only office–bin–year observations with positive awards. $Post_t$ equals 1 for years 2012 and later. For each office with more than 500 awards in the pre period, I compute the share with values in $(\$99,000, \$100,000]$ and set $HighBunch_k = 1$ for offices above the median of this measure. Finally, Q partitions the award-amount space. The coefficients of interest, δ_q, are on the triple interactions between award-size range, high-bunching offices, and the post period. Standard errors are clustered by awarding office. The estimation sample contains 592,589 observations.

Appendix Figure A7 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 approximate proportional changes in the number of contracts for offices with high pre-period bunching relative to those with low pre-period bunching, for award amounts in range q and relative to awards between \$500,000 and \$5,000,000. The results indicate that any differential change between low- and high-bunching offices occurred only within a window around the old and new simplified acquisition thresholds. When these ranges are combined into a single indicator (panel (b)), high-bunching offices exhibit a net increase in transactions relative to low-bunching offices. This increase was not offset elsewhere in the distribution: the estimates show no differential change below \$85,000 or above \$200,000. This evidence is inconsistent 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 $(\hat{x} - \hat{m})$. 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 2012-2016, and (ii) the estimated counterfactual distribution that would have been observed in the absence of a change in the simplified threshold. This difference is normalized by the average frequency over a 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 $(\hat{x} - \hat{m})$ 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 $(\hat{x} - \hat{m}) = 18.24$, with a standard error of 1.94. Therefore, a 95% confidence interval based on asymptotic normality corresponds to $[14.44, 22.04]$.

Appendix Table B1 presents 75 different estimates of $(\hat{x} - \hat{m})$, 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). Fixing the polynomial degree at $p = 3$, estimates range from 10.68 to 26.23, with 14 of the 25 estimates lying within the 95% confidence interval of the baseline estimate, and several others lying marginally outside of it.

Panel (a) presents results using a second-degree polynomial fit ($p = 2$). The estimate using the baseline excluded region decreases to 16.84, and 19 of the 25 estimates lie within the 95% confidence interval of the baseline estimate. The coefficient is relatively stable for changes of ± 10 thousand dollars to the lower bound and ± 50 thousand dollars to the upper bound of the excluded region.

Finally, the estimate is *not* stable when I consider a fourth-degree polynomial ($p = 4$), as presented in panel (c). The baseline point estimate nearly doubles, and alternative excluded regions yield coefficients as large as 72.93. This result is not surprising. The interpolation method fits a smooth function over relative frequency changes across the award distribution (see Figure 3) to capture changes that would occur even in the absence of a regulatory change. Although the polynomial allows some curvature, these changes are expected to be relatively uniform across the distribution. Higher-degree polynomials can therefore overfit observations outside the excluded region and reduce the accuracy of predictions within it. This differs from standard bunching methods that interpolate distribution *levels*, which typically feature more curvature and therefore use higher-order polynomials (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 project rating is based on three components: cost, schedule, and performance. In the same spirit as my measures of cost overruns and delays, the first two components compute deviations between planned and actual project cost and timeliness. 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 use 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 use a series of objective performance measures. These measures seek to capture the quality of the service being provided, which may be unrelated to cost and schedule performance. While the metrics vary by 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", and "percent of customer issues that are addressed and fully resolved within 24-hours." Furthermore, as Liebman and Mahoney (2017) note, CIOs have strong incentives to rate projects accurately.

I next ask whether and to what extent these project-quality measures correlate with the execution-based quality proxies used in the main analysis.

E.2. CIO evaluations versus award quality proxies

Using Dashboard files from 2013 through 2018, I match 552 major IT projects to 5,141 distinct delivery orders and BPA calls in the federal award data. Appendix Table B2 presents summary statistics for this matched sample. Panel (a) counts projects and awards once and reports total obligations by agency. Panel (b) summarizes the execution-based quality measures for 4,663 distinct awards with complete outcomes and CIO evaluations for 547 distinct matched projects with available ratings. The quality proxies are measured at the award level, whereas the CIO evaluation is measured at the project level.

I match Dashboard records to USAspending using direct or parent award identifiers, with agency and vendor information used to validate or resolve matches. An IT project can involve several awards, which need not contribute equally to the performance assessed by the CIO. Because delivery orders and BPA calls under indefinite-delivery vehicles (IDVs) meet specific requirements within an ongoing umbrella arrangement, their execution may be more closely related to the IT project's implementation and operation, relative to standalone purchases. For this reason, I use these orders and calls in the baseline analysis, and check robustness by first excluding awards linked to multiple projects and then adding standalone definitive contracts and purchase orders.

Figure A6 reports results for 5,104 project–award pairs, consisting of 4,657 distinct awards linked to 506 projects. 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 awards with a higher value of the quality proxy index based on contract execution measures. Panel (b) controls for project-agency and year fixed effects. The estimated slope is 0.069, with a standard error of 0.040 clustered at the IT-project level. Panel (c) weights observations by the natural logarithm of total obligations for awards linked to the project. The corresponding estimate is 0.073, with a clustered standard error of 0.041.

Taken together, these results suggest that, for this sample of matched IT projects, CIO evaluations based on actual quality metrics are positively correlated with the quality proxies computed from award execution variables.

Appendix Table B6 shows that excluding awards linked to multiple projects leaves the estimates nearly unchanged. Adding definitive contracts and purchase orders to this restricted sample yields smaller positive coefficients. This pattern is consistent with attenuation of the association if standalone awards are less closely related to the project's core activities.

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: $\tilde{U}(p) = v_A + (v_R - v_A) \cdot R - p - \gamma\kappa \cdot R$. The payoff for the firm is: $\Pi(p) = p - c - (1-\gamma)\,\kappa \cdot R$. Here R denotes the chosen procedure, which may be regulated even if the resulting award is below the threshold. Conditional on this choice, the Nash bargaining (NB) objective ($\zeta_R(p)$) is:

$$\zeta_R(p) = (v_A + (v_R - v_A) \cdot R - p - \gamma\kappa \cdot R)^{\phi} \cdot (p - c - (1-\gamma)\,\kappa \cdot R)^{(1-\phi)}$$

The four cases (A, B, C and D) outlined in Section 4 represent, respectively, an unconstrained simplified award; a bunching award; a regulated award; and no transaction. The buyer chooses between feasible bunching and regulation by comparing its net payoffs, anticipating the conditional bargaining prices. I proceed to describe each case 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 (8). Taking first order conditions, it is easy to check that the equilibrium award is:

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

Payoffs obtained in this case are:

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

Finally, the NB objective evaluated at this solution is:

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

Case B: A second candidate award is to bunch (subject to the buffer s): $\bar{P} - s$. This outcome is available when the unconstrained simplified award exceeds the threshold, bunching is not prevented by a friction, and both parties obtain nonnegative payoffs. The award is:

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

Payoffs are:

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

The NB objective is:

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

Case C: The third candidate solution uses regulated procurement. The parties divide the surplus after taking the κ costs into account, provided that $v_R - c - \kappa \geq 0$. The award maximizes the conditional objective in Equation (9) and need not exceed the threshold. 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:

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

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 not to transact. This gives a payoff of zero to the transacting parties. The award, payoffs, and value of the NB objective are given by:

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

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

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

Equilibrium: If the unconstrained simplified award does not exceed $\bar{P}$, Case A applies. Otherwise, an aligned buyer chooses Case B when bunching is available and $v_A - \bar{P} + s \geq \phi(v_R - c - \kappa)$. If bunching is unavailable or yields a lower buyer payoff, Case C applies whenever regulation is feasible. If neither alternative is feasible, Case D applies. This comparison uses the buyer's net payoff, rather than a Nash product maximized across procedures.

F.2. Misaligned bureaucrat

The selected specification uses a regulated fallback rather than a separate private reservation payoff for regulated procurement. If the unconstrained discretionary award v_A does not exceed $\bar{P}$, the misaligned agent awards $p_M^* = v_A$. If it exceeds the threshold, the agent can bunch at $\bar{P} - s$ when bunching is feasible, the transaction is not frictional, and the firm's payoff is at least as high as

under regulated procurement. The model sets the effective friction probability for a noncommercial misaligned transaction to

$$\eta_{BM} = \eta + (1 - \eta)\rho_{BM}.$$

If the simplified alternatives are unavailable, the parties use regulated procurement whenever $v_R - c - \kappa \geq 0$. The regulated award is

$$p_R = \phi c + (1 - \phi)v_R + (\phi - \gamma)\kappa.$$

Otherwise the transaction is not completed.

This ordering implements the institution's asymmetric threshold. Simplified procurement cannot be used above the threshold for noncommercial purchases, but a regulated procedure can generate an award below the threshold. It also ensures that an otherwise feasible simplified transaction remains available when regulation is infeasible.

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 denote the number of contracts in bin $b \in \{51, ..., 200\}$, and let R be a vector of round-number multiples for which there may be a special bias. I estimate the following regressions:

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

$$n_b = \sum_{p=0}^{P} \delta_p \cdot b^p + \sum_{r \in R} \lambda_r \cdot \mathbf{1}\left[\frac{b}{r} \in \mathbb{N}\right] + \nu_b, \qquad \text{for } b = 101, ..., 200.$$

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:

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

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 150 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

I smooth quality with a contract-level piecewise-linear regression over awards between \$50,000 and \$300,000. The regression includes a linear award-value term, a change in level and slope above the threshold, separate indicators for the three ten-thousand-dollar windows around the threshold, and fiscal-year fixed effects. I remove the fiscal-year component and average the fitted award-value component in 15 right-closed bins $(50, 60], (60, 70], \ldots, (190, 200]$. These bin means are the quality targets. The same construction is applied at the post-period threshold of 150 for out-of-sample evaluation.

G.3. Moment weighting

I use a fixed, non-optimal weighting matrix that emphasizes the award distribution near the threshold and the associated changes in quality. The weights were chosen to balance fit across these features, and are held fixed throughout optimization.

Let $r_{Nb,k}$, $r_{Qb,k}$, and $r_{Sj,k}$ denote model-minus-data residuals for award frequencies, mean quality, and simplified-procedure shares on simulation panel k. The loss is

$$\begin{aligned} J_k = &\sum_b w_{Nb} r_{Nb,k}^2 + \sum_b w_{Qb} r_{Qb,k}^2 + w_S \sum_j r_{Sj,k}^2 \\ &+ \omega_1 (r_{Q100,k} - r_{Q80,k})^2 + \omega_2 (r_{Q110,k} - r_{Q100,k})^2, \end{aligned}$$

where the sums run over the respective bins described above, and the weights incorporate normalizations used to compare moments measured on different scales.

Award-frequency weights emphasize bins near the threshold while retaining positive weight farther away. At equal distances from the threshold, bins below it receive approximately 3.3 times the weight of bins above it. Quality moments receive 60 percent additional weight in the bins ending at \$100,000 and \$110,000, and approximately 28 percent additional weight in the neighboring bins ending at \$90,000 and \$120,000. The seven procedure-share moments receive a common weight.

The final two terms place additional weight on the change in quality approaching the bunching region and the change across the threshold. They reinforce particular contrasts between existing quality moments rather than introduce independent information. The replication code records the exact numerical weights.

G.4. Optimization procedure

The estimated criterion is $(J_1 + J_2)/2$. Each panel contains 200,000 draws from an independently scrambled Sobol sequence, after skipping its first 4,096 entries; common draws are held fixed throughout optimization. Twenty independent broad initial populations each contain 100 points in the thirteen-dimensional structural search space. For each population I use a genetic algorithm with at most 160 generations and 16,000 objective calls, followed by normalized-coordinate pattern search with up to 8,000 calls. The local search starts with mesh size 0.002 and uses mesh tolerance 10^{-6} and step tolerance 10^{-8}; the objective tolerance is 10^{-8}. A separate stationarity audit checks feasible positive and negative coordinate changes of 0.0001, 0.001, and 0.005 times each parameter's search range, with the quality corrections profiled again at each probe. An estimate passes the stationarity test if no probe improves the native criterion by more than 10^{-6}. The full set is ranked both by the original criterion and by an independent two-million-draw evaluation with the native parameters, quality corrections, and alignment intercept fixed. I select the top-ranked solution after the genetic algorithm and local search, provided it passes the stationarity audit and retains its rank in the independent two-million-draw evaluation.

G.5. Standard errors

The SMM estimator is asymptotically normal under standard regularity conditions. These include (i) a compact parameter space Θ with the true parameter θ_0 in its interior; (ii) a positive semidefinite weighting matrix W; (iii) correctly specified and locally identifying moment conditions, with a continuously differentiable population moment vector and nonsingular $G'WG$ at θ_0; and (iv) a local linear approximation and a joint central limit theorem for the empirical and simulated moment discrepancies as the sample and simulation sizes increase (Pakes and Pollard, 1989; McFadden, 1989; Newey and McFadden, 1994).

For inference, I treat the thirteen structural parameters and the two quality corrections as fifteen jointly estimated coordinates (θ). Let G stack the two panel Jacobians of the moment vector (N, Q, S), and let $W = \mathrm{diag}(W_n/2, W_n/2)$ reproduce the criterion above. The covariance matrix is

$$\widehat{V}(\hat{\theta}) = (G'WG)^{-1}G'W\Sigma_g WG(G'WG)^{-1}.$$

Numerical derivatives with respect to structural parameters use common draws and a step of 0.001 times the parameter's search range, holding (u_q, v_q) fixed and recalibrating the alignment intercept.

I construct Σ_g to account for both sampling and simulation error. I draw 999 bootstrap samples from the 6,123 contracting offices, sampling offices with replacement and keeping each office's contracts together. In each sample, I recompute the empirical moments and re-estimate the intercept and slope of the reference quality regression. I use the delta method to account for uncertainty in these coefficients and their covariance with the empirical moments. I then add the covariance of the stacked simulated moments across 32 independent pairs of 200,000-draw panels, holding the estimated parameters and empirical inputs fixed.