Can the Private Sector Ensure the Public Interest? Evidence from Federal Procurement*

Leonardo M. Giuffrida†

Gabriele Rovigatti‡

June 19, 2017

Abstract

We empirically investigate the effect of oversight on contract outcomes in public procurement. In particular, we stress a distinction between *public* and *private* oversight: the former is a set of bureaucratic checks enacted by contracting offices, while the latter is carried out by private insurance companies whose money is at stake through so-called surety bonding. We analyze the universe of U.S. federal contracts in the period 2005-2015 and exploit an exogenous variation in the threshold for both sources of oversight, estimating their causal effects on costs and execution time. We find that: (i) public oversight negatively affects outcomes, in particular for less competent buyers; (ii) private oversight has a positive effect on outcomes by affecting both the ex-ante screening of bidders - altering the pool of winning firms - and the ex-post behavior of contractors.

JEL: D21, D44, D82, H57, L74.

Keywords: oversight, procurement, screening, red tape, moral hazard.

*Special thanks go to Francesco Decarolis, without whom this work would not have seen the light of day. We are grateful to Juan-José Ganuza, Sofia Lundberg, Stéphane Saussier, Giancarlo Spagnolo, and Hidenori Takahashi, who provided crucial insights and expertise. Also, we thank Klenio Barbosa, Decio Coviello, Ruben Enikolopov, Stefano Gagliarducci, John Gatherhood, Alessandro Gavazza, Alberto Iozzi, Vitalijs Jascisens, Calvin Luscombe, Neale Mahoney, Juan Ortner, Jacopo Ponticelli, Johannes Spinnewijn, Francis Vella, two anonymous referees as well as the audiences at the Tor Vergata Lunch Seminars, Barcelona Jamboree 2016 and GRAPE February 2017 meeting for comments that improved the manuscript. Any errors are our own.

†University of Rome Tor Vergata - Department of Economics and Finance. Mail: *leonardo.giuffrida@uniroma2.it*

‡University of Chicago - Booth School of Business. Mail: *grovigatti@uchicago.edu*

I Introduction

Efficient contract procurement is a complex task as sellers have private information on production costs and incentives in exerting effort that are not disclosed through the offers. This informational asymmetry, combined with the intrinsic cost uncertainty at the awarding stage, leaves room for adverse selection and moral hazard during the procurement process. In turn, these issues lead to the renegotiation of contract terms and to an increase in costs and time to completion. To cope with this well-known phenomenon, the previous literature has focused on the role of both awarding procedures in screening bidders and optimal contract design in avoiding misbehavior.[1] On top of that, an efficient procurement regulation should require a balanced level of oversight and even rely on outsourcing to the private sector when this proves to be beneficial (Hart et al. [1997]). Although the optimal level of oversight is well defined in theoretical contributions (Shavell [1984]), however it is fiercely debated in practical applications.[2]

While in the practice of public and private procurement the handling of the contract execution stage is seen as the first-order concern, with a few notable exceptions (Bajari et al. [2014], Lewis and Bajari [2017]) the empirical literature has focused on the contract awarding phase and ignored the operational phase. This paper contributes to the small but growing literature that aims at filling this gap by providing empirical evidence that oversight in public procurement matters. Specifically, we propose a distinction between *public* and *private* oversight, depending on its source. Public oversight includes all formal checks - cost certifications, pricing data transmission, production surveillance - which the contracting authorities enact during the contract awarding phase and execution. It typically involves considerable paperwork for both the buyer and the sellers. At the cost of some *red tape*, it is aimed at alleviating the moral hazard problem (Kaufman [1977], Shleifer and Vishny [1998]). Combinations of public oversight and bureaucrats' ineptitude might be deadly for procurement processes: in February, 2014, the U.S. Court of Appeals for the Federal Circuit sentenced against the federal government because the U.S. Navy officials took

[1] Bajari et al. [2009] and Decarolis [2014] belong to the former group, Bajari and Tadelis [2001] to the latter.

[2] See for example the technical reports GAO [2013] and Garvin et al. [2011].

two years after the original completion date to accept the project as complete and caused million dollars of losses to the contractor.[3] On the other hand, *private oversight* involves third parties - surety companies - issuing bonds (*surety bonds*) to secure the buyer against unpredictable events.[4] If the seller fails to fulfill contractual tasks, contracting authorities make claims to recover losses.[5] A surety is then called on either to complete the public work by themselves (i.e. with their own resources or by subcontracting) or to refund the authority of the bond value. Being liable in case of unsatisfactory contract outcomes, the sureties have strong incentives both to screen bidders (ex ante) and to monitor contractors (ex post). They help mitigate the asymmetry of information between the buyer and the sellers thanks to their experience of the market - i.e. access to private information - and the screening enacted through price discrimination on premia, which directly affects offers placed by potential contractors.[6] Hence, *private oversight* enhances the selection of the best contractors and provides a second tier of monitoring of contractors' progresses.

Identifying the extent to which and the channels through which public and private oversight affect contract outcomes has clear policy implications. Moreover, surety bonds are an increasingly popular tool in procurement governance and in many countries there is an intense debate about their efficacy; we contribute by providing quantitative support.[7] The U.S. constitutes an excellent case study for the outlined framework as both public and private oversight are required depending on the industry and contract value. Furthermore, surety bonding is well-known among all players in the procurement market in the U.S. as it was the first country to introduce it in 1894.[8]

In our work we use a newly available database containing contract-level information on

[3] *Metcalf Construction Company, Inc. v. United States* (http://caselaw.findlaw.com/us-federal-circuit/1656993.html).

[4] Surety companies (more simply *sureties*) usually are subsidiaries of insurance companies.

[5] The rationale of the law was initially to protect the buyer from losses in case of seller's bankruptcy.

[6] Premia paid for surety bonds may vary depending, among the others, on the valuation of bidder quality.

[7] Surety bonds are widely used at government-level procurement not only in the U.S., but also in Japan and Canada. Also, many states in the U.S. have introduced suretyship through the so-called "Little Miller Act". In 1999, the European Commissions Enterprise Section published a report titled "Abnormally low tenders" with detection and rejection rules for abnormally low tenders and started a working group on surety bonds (European Commission Enterprise Section, 1999).

[8] The *Heard Act*, requiring surety bonds on all federally funded projects, was replaced by the *Miller Act* in 1935.

the near-universe of U.S. federal procurement.[9] Focusing on 2005-2015 data, our identification strategy relies on the contemporaneous change (occurring in 2010) of the threshold for (i) the Simplified Acquisition Procedures, exempting *all* federal procurement contracts from public oversight; and (ii) the Miller Act, the law requiring private oversight only in construction projects through surety bonding.[10] We deploy a difference-in-difference-in-differences (DDD) approach to estimate the causal effect of the different sources of oversight on performance outcomes.[11] Specifically, for the whole population of federal procurement awards, we compare the average change in outcomes of contracts that are exempted from public oversight with corresponding changes among those that remain subject to the requirement. To take into account possible differences between construction and other services, due to the additional application of private oversight to the former procurement category, we simultaneously compare changes in outcomes between the two groups.

Our reduced-form analysis yields two main findings. First, exempting contracts from private oversight negatively affects performance in terms of time and cost, leading to a decrease of 9 and 4.2 percentage points, respectively. We then exploit firm-level data to provide evidence that adverse selection plays a key role in driving our estimates. All these findings are in line with Calveras et al. [2004], who develop a model of public procurement with surety bonds, whose premium is proportional to the riskiness of the bidder, and show that their presence improves the selection of winning firms.[12] Second, we find that exempting contracts from public oversight improves both time and cost outcomes, leading to increases in performance of +7.2% and +5.3%, respectively. This is in line with the results of Calvo et al. [2016]; however, we also find that the red tape effect in public oversight is negatively correlated with the contracting authority competence, and we do not find any significant outcome when estimating the treatment effect on the subset of high-competence offices.[13] In

[9]The Federal Procurement Data System - Next Generation (FPDS-NG) is publicly available at https://www.fpds.gov/fpdsng_cms/index.php/en/ and updated on a daily basis. The FPDS database is well documented and was recently used by Liebman and Mahoney [2016], among the others.

[10]This subset of data includes nearly 40 million purchases and totals around $5.6 trillion in government expenditure.

[11]Recently, Bergman et al. [2016] used the same econometric approach in the procurement of elderly care services in Sweden.

[12]The premium is incorporated into the bid and affects the probability of winning the auction. Thus, the higher the risk for the surety, the higher the premium charged and the lower the chance of winning.

[13]The definition of *competence* is controversial and we will not address it in the present paper. In our

the construction sector, where the 2010 reform implied the simultaneous elimination of both public and private oversight, we find that their combined effect on contract performance is ambiguous: we observe a decrease in time performance of 1.8% and an increase in cost performance of .6%. The straightforward implication of our results is that an effective reform should exempt contractors from public oversight and keep the provision of the Miller Act fixed.

This paper contributes to the literature on optimal procurement regulation and to the debate on effectiveness of public vs. private supply of public goods. In turn, the first strand can be divided into two branches depending on the focus of the analysis: i) papers dealing with ex-ante regulations through the analysis of auction formats, contract types, awarding procedures and their effects on participation and performances (recent examples in this literature include Marion [2007], Board [2007], Marion [2009], Krasnokutskaya and Seim [2011], Bajari and Lewis [2014], and Branzoli and Decarolis [2015]); and ii) papers focusing on ex-post tools for enhancement of contract outcomes: oversight (Calvo et al. [2016]) and relational contracting (Coviello et al. [2017], Banerjee and Duflo [2000], Calzolari and Spagnolo [2009]). Our paper combines these approaches in disentangling the role of surety bonding as a regulatory element on the one hand and as a mean to increase monitoring of contractors on the other.

We emphasize the choice between direct provision of public services and outsourcing to private contractors (Hart et al. [1997]). Examples of empirical economic analyses of government efficiency that make use of direct measurements of outcomes, the approach our paper follows, include Di Tella and Schargrodsky [2003], Reinikka and Svensson [2004], Olken [2006, 2007], Bertrand et al. [2004], Fisman and Gatti [2006], Fisman and Miguel [2007], Hyytinen et al. [2009], and Ferraz and Finan [2008, 2011]. In their paper, Bandiera et al. [2009] identify the amount and the sources of public waste in Italian public procurement. They find that inefficiency is by far the most important dimension in explaining public waste, with heterogeneity across different buyers, and that the best performance (both in

exercise we will proxy competence through the closely related concept of *performance persistence*: we will use a weighted distribution of past contractual performance and divide our sample into competent and incompetent offices depending on the median value (Decarolis et al. [2017]).

terms of active and passive waste) is associated with more autonomy/discretion. According to Kelman [1990, 2005], an ultimate cause of passive waste in the U.S. federal government is that an excessive regulatory burden may make procurement cumbersome and increase average prices.

To the best of our knowledge, this paper is the first to empirically assess the role of surety bonding and the associated private oversight in public procurement. Despite not being widely known, surety bonding is a founding pillar of U.S. public construction procurement, which is a crucial economic sector worth approximately $32 billion, and was extensively used during the recent financial crisis as a fiscal policy tool to stimulate the economy (see the *American Recovery and Reinvestment Act*).[14] Both at the federal levek (Miller act) and the state level (Little Miller Acts), there were only slight variations in the regulations before the 2010 reform; therefore, assessing the effectiveness of surety bonds has essentially been impossible. On top of that, the low default rate in the industry (less than 1%) has been interpreted by some as an indication that surety bonds are redundant and represent an unnecessary cost for firms and public buyers, and should therefore be eliminated (Gransberg et al. [2014]). This paper, instead, uses novel variation to identify the causal effect of this instrument and reveals that its quantitative effects on contract performance are large and positive, both in terms of time and costs. Furthermore, providing evidence in favor of the screening role of sureties reverses the causality previously highlighted: surety bonding is what helps keep the default rate low by enhancing the selection of the best contractors.

The remainder of the paper unfolds as follows. In section II we present the concept of suretyship and the related U.S. legislative context. Section III deals with the theoretical background underlying our analysis; section IV outlines the data we employ in our analysis; section V addresses the empirical analysis, outlines the identification strategy and presents results; in section VI we discuss the main drivers of our findings; and section VII concludes.

[14]Year 2013, source: Federal Reserve Bank of St. Louis.

II Context

In this section, we first describe the institution of suretyship, its legislative foundations and the economic rationale underlying its provision in several public procurement regulations. We then shift the focus to the U.S. federal procurement regulation to define and discuss both *public* and *private oversight*. These subsections provide the necessary background to address the theoretical model underlying our analysis, which closes the section.

II.1 Suretyship

Procuring supplies entails strategic considerations on competition, tender design and optimal ex-post rating in order to ensure the maximum benefit for the procurer. When dealing with procurement of services, buyers also face uncertainty related to production cost and business factor dynamics: unexpected negative shocks could hit contractors during the execution of the work, leading to profit erosion and, ultimately, losses.[15] In the worst-case scenario, contractors are forced to declare bankruptcy, leaving the work incomplete and the buyer with no party to make claims against. Avoiding such lose-lose outcomes is a first-order concern for all parties, and situations of this sort are typically handled by renegotiating contract provisions either in terms of time or costs. This leaves room for moral hazard and adverse selection issues, as low-quality firms may take advantage of cost uncertainty at the awarding stage, underbid and then renegotiate once they are awarded the contract - e.g. by pretending to have suffered an unexpected negative cost shock (Guasch et al. [2008]).

Hence, when the contractors' probability of default is high, it makes sense for buyers to take out an insurance to avoid bearing all risks on their own. Suretyship is a specific line of insurance based on the issuance of a *surety bond* involving three parties: the surety

[15] According to the OECD, *services* are "outputs produced to order and which cannot be traded separately from their production". A broader definition provided by the management literature is based on "the five Is": Intangibility, Inventory, Inseparability, Inconsistency and Involvement. Either way, throughout the paper we will distinguish supplies contracts from service contracts according to the underlying timing of production: while goods could - in principle - be stored and sold outright, services are customized and need time to be produced and delivered after the contract award.

guarantees that the contractor will perform the tasks demanded by the buyer.[16] In other words, the suretyship works as a risk-transfer mechanism between the buyer and the surety, but it is demanded by a third party - the contractor - that guarantees the performance of an obligation.

Prior to issuing a bond, the potential contractor is subject to a *screening process* by the surety - consisting of an assessment of its entire business operations, financial resources, experience, organization, backlog, profitability and management capability - aimed at extrapolating private information on agent's type. The surety, due to its comprehensive access to firms' information during the prequalification phase and its prior experience of the market, can evaluate the contractor's ability to fulfill the contract provisions.[17] The whole process culminates in the determination of a *premium*, an actuarially based fee that varies depending on the size, type and duration of the project and, notably, on how the characteristics of the contractor that emerged from the screening process match the project complexity. In the U.S., the bond price mostly ranges from 0.5% to 3% of the contract amount and the potential contractor typically incorporates the bond premium amount into the offer. Hence, the screening enacted by the surety makes the premium a prominent *ex-ante* mechanism for discriminating among potential contractors. The premium, by reducing the asymmetry of information and affecting the offered bid amount, takes a relevant role in determining the quality of the winner in a competitive tender and shifts adverse selection and moral hazard issues away from the procurer, whose only piece of information about the sellers at the award stage is the offer placed.

Furthermore, sureties systematically gather and analyze information regarding bonded contractors after the contract award. They have the legal right to access information on work progress, payments and the estimated percentage of completion for bonded projects. Prior to modifying contractual terms, procurers and contractors shall obtain the consent of the surety on the basis of the gathered information on contractor conduct.[18] Hence, in addition to being screened, bonded contractors undergo an effective *ex-post* monitoring process by

[16]The legal definitions for buyer and contractor are *obligee* and *principal*, respectively.

[17]Surety bonds are common across the entire U.S. construction industry. Construction bonds generate two-thirds of total surety premia written and 70% of total revenues.

[18]Federal Acquisition Regulation (FAR) Part 28.

sureties.

A comparison with letters of credit (LOC), widely used in the European procurement market, might be useful to better understand how surety bonds differ from other traditional forms of guarantee in their nature and the underlying incentives provided. A LOC, normally issued by a bank, is a cash guarantee to the buyer who can call on demand and receive a pre-specified amount of money if some breach of contract were to occur. An surety bond protects the buyer from nonperformance and financial exposure, should the contractor default. Hence, while the performance of the contract has no or little relation to the bank's obligation to pay on the LOC, the primary focus of a surety bond is the effective accomplishment of the work. The two instruments also differ with respect to their effect on the contractor's borrowing capacity and the prequalification process. In order to issue an LOC, the bank requires the contractor to pledge specific assets to be paid in case of insolvency. An LOC thus diminishes the contractor's line of credit and appears on financial statements as a contingent liability. The bank examines the quality and liquidity of the asset by checking whether it could back up the debt; if this is the case, no further prequalification is required. Hence, a bank issuing an LOC takes no risk and has no incentive to screen the contractor, whose liquidity is reduced to back up the LOC. Should the applicant be unable to make payment on the purchase, it shall cover the outstanding amount. In contrast, surety bonds are usually issued on an unsecured basis and neither alter firms' assets nor diminish the contractor's borrowing capacity; in other words, the surety bears part of the project risk. In order to ensure the delivery of the contract object in case of contractor's default the surety has to choose between the following: (i) covering production costs by itself and allowing the contractor to finish the works; (ii) selecting a new contractor to conclude the residual tasks; or, only as a last resort, (iii) refunding the bond value to the buyer, leaving the execution incomplete.

These crucial differences imply that an LOC is likely to be unavailable to companies with few assets, which excludes them from participating in the tender and thus reduces competition on dimensions not related to quality. Since sureties, which must have sufficient assets to back up the bonds they issue, are partially responsible for the completion of the works,

they have strong incentives to properly screen potential contractors' financial resources and their ability to execute the job. This point crucially inspired our work. *Ceteris paribus*, a bonded project is more likely to be completed in accordance with the contract provisions as the likelihood of contractor default or any breach of procurement contract clauses is reduced, while the awarding price may be higher due to a premium.

II.2 Legislative Framework and Oversight

According to the Federal Acquisition Regulation (FAR), the guidebook governing the public procurement process in the U.S., auditing offices within each contracting bureau must require sellers to (i) complete expenditure justification forms and submit cost or pricing data certifying that expenses are based on adequate price competition (FAR Part 15); and (ii) submit reports on the project's progress to specific evaluation teams (FAR Part 42). Following Calvo et al. [2016], we will refer to these ex-post formal background checks collectively as *public oversight*.[19] The Simplified Acquisition Procedures, introduced with the Federal Acquisition Streamlining Act of 1994, aim at reducing the administrative burden for the sellers, mainly small businesses, when working for the Government.[20] Under simplified acquisitions, contracting offices do not exert any public oversight over contractors.

In addition to this picture, a series of private checks are introduced by the Miller Act, the law requiring surety bonds in US procurement.[21] The Act applies *only* to contracts awarded for the construction, alteration, or repair of any public building (for the sake of simplicity, we will refer to this subset of contracts as *constructions* henceforth) of the U.S. federal government. The Miller Act imposes that, in order to be allowed to participate in the tender, potential contractors must furnish the federal government with a surety bond pre-approval. Typically, the surety bond amounts to the 100% of the contract price.[22] Throughout this

[19]The number and type of checks are slightly different for each contracting office, but their scope is analogous and we can coherently group them into one set.

[20]Indeed, according to the FAR, the purpose of the Simplified Acquisition Procedures is to (i) reduce administrative costs; (ii) improve opportunities for small and disadvantaged businesses to obtain a fair proportion of government contracts; (iii) promote efficiency and economy in contracting; and (iv) avoid unnecessary burdens for agencies and contractors.

[21]40 U.S.C. sections 3131-3134

[22]Contractors are free to choose their own surety from a list of financial companies which the U.S. De-

paper we will refer to the ex-ante and ex-post checks enacted by the sureties as *private oversight*.

Federal procurers are required to use simplified acquisitions to the maximum extent practicable for purchases of supplies or services whose anticipated dollar value does not exceed the Simplified Acquisition Threshold.[23] The Miller Act applies only above the same monetary cutoff and, for the sake of convenience, we will refer to the *oversight threshold* for the implementation of both regulations. In this sense, a contractor awarded a construction project whose anticipated value lies below the oversight threshold is exempt from both public and private oversight. We exploit a change in the oversight threshold that was enacted in 2010 to inform our identification strategy. The National Defense Authorization Act for Fiscal Year 2005 requires an adjustment every five years of acquisition-related thresholds for inflation. The law applies to simplified acquisitions and the Miller Act through an update of the oversight threshold, which was raised from $100,000 to $150,000 as of October 1, 2010.[24]

Figure (1) provides a stylized timeline of the outlined framework. The horizontal solid line represents the oversight threshold, moving upward in October, 2010, the grid identifies awarded contracts worth $100,000 to $149,999, while the background colors refer to oversight application (dark) or waiver (pale). In the case of construction contracts, the waiver included both public and private oversight, while for all other contracts ("non-constructions" from now on) the exemption was from public oversight only.[25] Over the time span considered, construction contracts valued above $150,000 are always subject to private and public oversight, while non-construction contracts of the same amount are subject to public oversight; construction contracts below $100,000, on the other hand, are always exempted from private and public oversight, while non-construction contracts from public oversight.

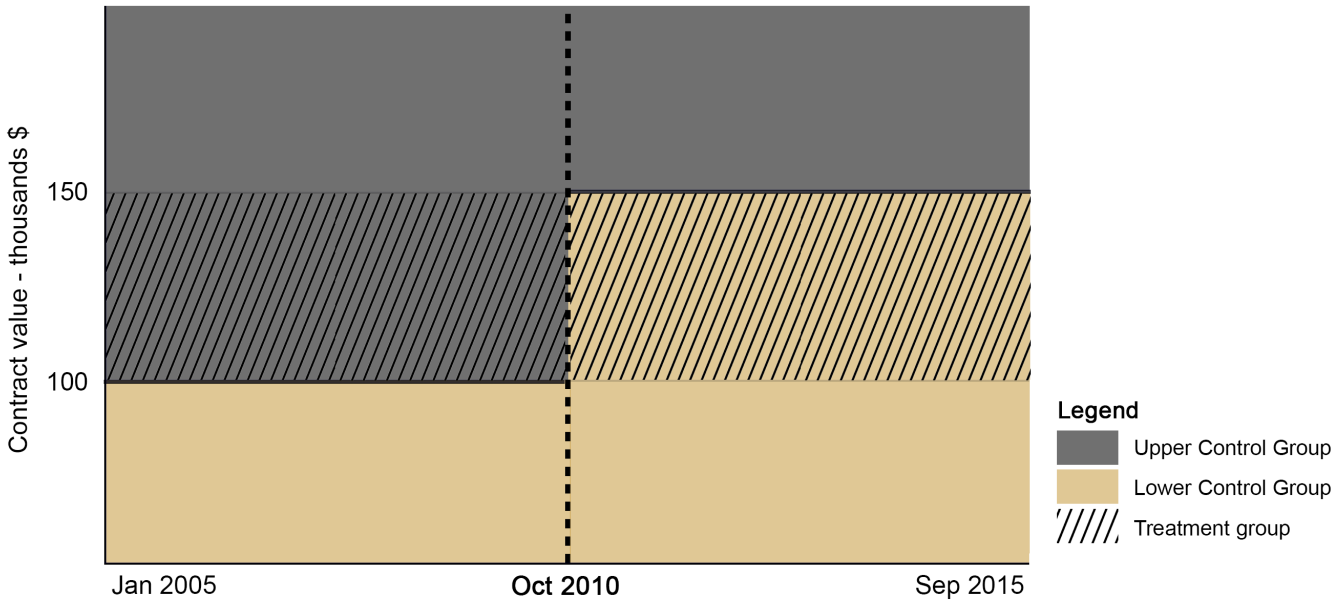
partment of Treasury establishes as qualified to underwrite surety bonds on federal government projects. This certificate of authority also determines the amount of the maximum limits of coverage for each of these. In other words, a surety that wants to issue bonds for federal government construction projects is in turn subject to a financial review that officially sets its bond size limit.

[23]FAR part 13.

[24]The adjustment is rounded - in the case of a dollar threshold that is not less than $100,000, but is less than $1,000,000 - to the nearest $50,000.

[25]The law provides alternate payment protection for contracts that exceed $30,000, so that contracts below the Miller Act threshold are not entirely without payment protection.

Figure 1: Reform Timing



Notes: Contracts exempt (pale) and not exempt (dark) from oversight before and after October, 2010. The grid identifies the treatment group, i.e. those contracts subject to oversight before but not after the reform. Upper control group includes contracts always exposed to oversight (i.e. always dark) while the lower control group consists of contracts never exposed (always pale).

III Theoretical Background

In seeking to restrain vendors' misconduct, public oversight introduces a burden in terms of both time and cost. To produce the required paperwork, contractors must divert resources away from contract-specific tasks, and their autonomy in the execution of works is curtailed by the need for public approval. To sum up, enforcing public oversight may lead to two conflicting phenomena:

- Hypothesis *a.1)* : The introduction of an unnecessary bureaucratic burden for contracting parties causes longer delays and higher costs - *red tape effect*;[26]

[26]See Bozeman [1993] for a review of the theory of red tape and public contracting.

- Hypothesis *a.2)* : project supervision reduces the risk and the extent of opportunism, slack conduct or misbehavior in contract execution - *public moral hazard effect*.[27]

An ex-ante assessment of the effect of public oversight on contractors' performance is not trivial. On the one hand, the two effects are competing;[28] on the other hand, both might be non-linear in the contract amount - monitoring might be a wasteful activity only for small projects, and could lead to savings for larger ones.

On top of public oversight, firms competing for federal construction contracts are required to obtain surety bonds and be subject to private oversight. This entails oversight exerted by private companies, i.e the sureties. The effect of a surety bond on contract outcomes may have two sources:

- Hypothesis *b.1)*: firms subject/not subject to Miller Act provisions are structurally different due to the screening effect induced by sureties - *adverse selection effect*;

- Hypothesis *b.2)*: as for public oversight, being covered and monitored by a surety gives firms more incentives to complete contracts under the terms and conditions agreed - *private moral hazard effect*.

Hypothesis *b.1)* is the one proposed by Calveras et al. [2004] (CGH henceforth), according to which we should observe a different pool of winning firms before and after the reform.[29] Specifically, since sellers are no longer subject to the pre-bidding screening process, we should observe a high turnover rate between firm types. After the reform, low-quality firms are supposed to be more likely to win at the expense of the good types given that their low quality does not reflect on higher premia charged by sureties anymore.[30] Thus we would expect more bad-type contractors to enter the pool of winners, good types to exit and the

[27] See Spiller [2008] for the theory on public contracts and opportunism; see also Decarolis et al. [2016].

[28] Identifying the extent to which the red tape and the moral hazard effects induced by the public oversight interact and affect the contract outcomes goes beyond the scope of the present paper.

[29] The FPDS-NG only reports tender winners. Indeed, according to hypothesis *b.1)*, the pool of potential contractors does not necessarily change with or without screening.

[30] In CGH terms limited liability companies are more willing to bid aggresively and, ultimately, face risks and an unexpected need to revise contract terms.

quality of the average contract outcome to decline accordingly. Hypothesis *b.2)* underlies a different prediction on the pool of winning firms. The assumption that surety companies do not screen potential contractors through a premium discrimination implies that we should not observe any significant change in the composition and structure of awarded firms after the reform. In such a framework, what matters instead is that removing surety bonds reduces the incentives for the *same* firms to exert the effort required to accomplish the contract tasks. To guarantee contract completion, sureties check the status of works and evaluate contractors' performance. In their absence, an issue of moral hazard arises and contractors tend to perform worse.

Hypotheses *b.1)* and *b.2)* are not competing and we expect both to be relevant in the public procurement market. The role designed by the law for surety companies is meant to minimize both effects through an ex-ante and ex-post monitoring of contractors. The overall effect of private oversight on contract outcomes amounts to the sum of selection, monitoring and the interactions of the two, and we expect it to be positive in terms of contract outcomes.

IV Data

IV.1 FPDS Dataset

The data we use are sourced from the Federal Procurement Data System (FPDS), a database to which federal contracting officers in the U.S. submit complete reports on procurement contract actions, as required by the FAR. It contains all contracts, both supply- and service-based, that have been awarded by the U.S. government and exceed an individual transaction value of $2,500, as well as every following activity.[31] The dataset also includes several variables related to the transaction itself, including buyer and seller characteristics in addition to solicitation and contract information, such as the signature, award and insertion dates, the contract object and its category (i.e. service or supply).

[31] Data are gathered by contracting offices in 23 agencies. In tables (A.3) and (A.4) we report the number of contracts per agency/year.

Importantly, we observe the type of solicitation procedures used, which reveals whether a contract is awarded through Simplified Acquisition Procedures or any other procedure (e.g., sealed bid, negotiation). Using this information, we build the binary variable SAP_i, indicating whether contract i has been waived from public and private oversight or not. The SAP variable crucially supports our identification strategy: ideally, we would like to observe the engineers' estimated value (EV_i), which is the piece of information used by the contracting office to assign the public oversight treatment to a contract. However, this is not recorded into FPDS and we are able to overcome the issue only combining information provided by i) SAP, i.e. we identify contracts exempted from public and private oversight, and ii) the ex-post contract value.[32]

The version of FPDS employed dates back to September, 30 2015.

IV.2 Data Management

We split the data into two main groups: contracts and amendment records. The former refer to the first transaction between a procurer and a vendor and correspond to our unit of observation, whose reported characteristics represent the benchmark agreement information. The latter account for all the revisions, modifications or corrections to existing contracts. Each contract is identified through a unique ID which is used to mark all its present and future alterations; therefore, we are able to track the entire contract history and link each contract to its revisions. Amendment records are classified according to the reason for contract modification, which is reported alongside the extra cost and time taken to complete the works. We further group them into *in-scope* or *out-of-scope* revisions, depending on whether the goal of the amendment is consistent with the initial contract terms.[33] We use

[32]Consider two contracts, A and B, whose observed contract value is \$105,000, both awarded before the threshold revision. However, the unobservable engineers' estimated value of A, EV_A, is \$110,000, while $EV_B = \$95,000$. According to the contract value, they are both subject to oversight. However, exploiting the fact that $SAP_A = 0$ and $SAP_B = 1$, we can proceed to the correct identification and avoid any source of bias in the estimates.

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

the *in-scope* amendments to build the outcome measures of our empirical analysis presented below.

Performance Indexes: First, we define: i) *Time Overrun*, representing the days in excess of a project's initial deadline; measured as the difference - in days - between the actual completion date and the estimated one and ii) *Cost Overrun*, standing for the expenses in excess of a project's initial budget; it is the sum - in thousands of dollars - of all renegotiated amounts. *Time Overrun* and *Cost Overrun* are well established measures for contractual performance,[34]; however, there are circumstances in which renegotiating the contract terms leads to optimal outcomes - typically, this is the case for complex, structured projects likely to be subject to unexpected events (negative cost shocks, adverse natural conditions, etc.). Given high-value contracts are the minority in the FPDS data, and according to Spiller [2008], who argues that renegotiations are suboptimal in the public procurement context, we consider the measures built on *in-scope* amendments only to adequately reflect the performance of a contractor.[35]

In order to compare the two overrun measures with the initial expected outcomes - that is, the time/cost of completion specified in the contract terms - we specify two indexes for contract performance:[36] $performance_{ig} = \frac{expected\ outcome_{ig}}{expected\ outcome_{ig} + overrun_{ig}}$, where i refers to the contract and $g = [time, cost]$. By construction, it maps the couple $[expected\ outcome\,; overrun]$ to the interval $[0, 1]$, with an increasing performance approaching 1, i.e. in the case of no overruns. Not surprisingly, the two performance measures are positively correlated (50%).

[34] Among the others, see Lewis and Bajari [2017], Coviello et al. [2017], Decarolis [2014] and Guasch et al. [2008].

[35] Spiller [2008]'s argument unfolds as follows: given the formal, bureaucratic nature of public contracting, any terms renegotiation would add adjustment costs, providing weaker incentives to adapt for both contractors and public authorities. Bajari et al. [2014] provide support to this hypothesis by quantifying in 8 to 14% of the winning bid the adaptation costs in their construction data.

[36] The two overrun measures are positively correlated (48%). This feature of our data differs from that in Decarolis [2014], who finds a nearly zero correlation between the two types of renegotiation and no evidence of a nonlinear relationship. He stresses, however, that designing the contract in such a way that the contractor would be in charge of both the design and the execution of the project would lead to shorter time and greater cost overruns. We are not able to reproduce his results since the FPDS does not contain such information.

Amended vs Non-amended Contracts: These are binary variables indicating whether the contract terms have been amended, i.e. at least one modification follows the initial contract signature in terms of completion time (*Time Amended*) or final cost (*Cost Amended*).

Average Overrun: This is the average amount of time (*Average Time Overrun*) and cost (*Average Cost Overrun*) overruns - i.e. $\frac{\sum_{k=1}^{K} amount\ amended_{i,k}}{number\ amendments_i}$, where i stands for the contract and k the amendment. These variables are defined only for the subset of contracts subject to at least one revision.

The FPDS dataset includes a number of other variables from which we build the controls in our regressions. *SAP* is a binary variable indicating whether the contract has been subject to the Simplified Acquisition Procedures; *Construction* is an indicator variable for construction contracts; *Limited Liability* indicates whether the business structure of the awarded company is limited liability; and *Last Week* highlights whether the transaction occurred in the final week of the fiscal year. In relation to the latter, following the argument in Liebman and Mahoney [2016], many public offices fund their spending out of a fixed budget that expires at the year end. When this deadline approaches, they rush to spend the last funds on possibly lower-quality projects. The average quality of works relating to contracts assigned in the final week is supposed to be lower, accordingly.

IV.3 Sample Selection

We restrict our sample to those contracts awarded through competitive solicitations because the effect of the treatments would otherwise not be observable.[37] For similar reasons, we focus on contracts whose tasks are such that the vendor can influence the outcome metrics through effort. Supply contracts do not allow for renegotiations. Hence, for these contracts our measure of performance does not proxy outcome quality whatsoever and we exclude them from the analysis.[38] The same rationale applies to the service subcategory "Lease or

[37] We consider as competitive a lot for which the extent of competition is labelled "Full and open" and whose participation is not set aside to any specific group of firms. In non-competitive auctions, the participation criteria restrict the competition *ex-ante* to dimensions other than quality (e.g. Athey et al. [2011])

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

Rental of Equipment, Structures, or Facilities".[39] In order to keep a balanced time-window around the SAT update, we rule out observations before January, 1, 2005, and cover the years 2005 to 2015. We eliminate contracts whose expected termination date is beyond the date of data download - September 30, 2015 - to keep only completed projects. We also drop contracts related to certain commercial items that make use of simplified procedures for the acquisition of services for amounts greater than the oversight threshold. This cleaning process yields a sample of 226,161 contracts and 23,870 unique firms .[40]

Two sets of contracts - the two solid colored sections in figure (1) - are potential candidates for use as control groups: the "always exposed" set (*upper control group*) and the "never exposed" set (*lower control group*). According to hypothesis *b.1)*, the presence of the surety company induces selection in awarded firms: hence, we will investigate the extent of these flows among groups as the threshold changes. The reform date and the two treatments cluster the sample into 6 distinct groups: the Service treatment group, counting all contracts - constructions included - valued between \$100,000 and \$149,999 that are subject to public oversight before but not after the reform; Service upper and lower control groups, consisting of all contracts valued more than \$150,000 or less than \$100,000, respectively; and surety bond (SB) treatment, upper and lower control groups, including construction contracts only, subject to private as well public oversight, with the same monetary cutoffs.

In Table (1) we report summary statistics for the Service treatment group and upper control group, both before and after the reform.

[39]Services included in the sample are: Special Studies/Analysis, Not R&D; Architect and Engineering Services; Information Technology and Telecommunications; Purchase of Structures/Facilities; Natural Resources Management; Social; Quality Control, Testing, and Inspection; Maintenance, Repair, and Rebuilding of Equipment; Modification of Equipment; Technical Representative; Operation of Structures/Facilities; Installation of Equipment; Salvage; Medical; Support (Professional/Administrative/Management); Utilities and Housekeeping; Photo/Map/Print/Publication; Education/Training; Transportation/Travel/Relocation.

[40]The firm ID variable is missing in approximately 57% of the contracts in our sample.

Table 1: Summary statistics - Service sample

	Upper Control Group							
	Before				After			
	Mean	SD	Median	N	Mean	SD	Median	N
Time Performance	0.7	0.3	0.74	51,224	0.7	0.3	0.82	52,692
Num Time Amendments	2.2	2.7	1	62,013	1.3	2.0	0	64,434
Prob Time Revision	0.6	0.5	1	62,013	0.5	0.5	0	64,434
Avg Time Overrun	216.2	282.8	130.5	33,223	217.6	293.4	114.7	27,887
Cost Performance	0.7	0.3	0.92	51,854	0.8	0.3	0.99	53,451
Num Cost Amendments	2.6	3.0	1	62,013	1.8	2.4	1	64,434
Prob Cost Revision	0.7	0.5	1	62,013	0.6	0.5	1	64,434
Avg Cost Overrun	274.0	334.4	205.1	31,626	307.7	413.5	176.6	27,457
Contract Value	2,051.9	56,831.2	414.4	62,013	1,068.3	4,139.9	297.5	64,434
# Contractual Days	403.6	395.5	364	62,012	296.7	220.0	347	64,434
Offers received	5.2	18.1	2	62,013	5.5	17.7	2	64,434

	Treatment Group							
	Before				After			
	Mean	SD	Median	N	Mean	SD	Median	N
Time Performance	0.8	0.3	1	13,771	0.8	0.3	1	2,061
Num Time Amendments	1.1	1.9	0	16,834	0.8	1.5	0	2,726
Prob Time Revision	0.4	0.5	0	16,834	0.4	0.5	0	2,726
Avg Time Overrun	251.0	318.1	146.1	6,094	194.2	265.3	92	926
Cost Performance	0.8	0.3	1	14,013	0.9	0.2	1	2,128
Num Cost Amendments	1.3	2.1	0	16,834	1.0	1.6	0	2,726
Prob Cost Revision	0.5	0.5	0	16,834	0.4	0.5	0	2,726
Avg Cost Overrun	101.1	125.8	64.2	5,386	67.6	96.7	34.3	626
Contract Value	122.0	15.2	121.0	16,834	121.6	16.4	120	2,726
# Contractual Days	292.8	336.1	213	16,833	225.5	196.4	196.5	2,726
Offers received	4.6	13.8	2	16,834	3.4	16.5	1	2,726

Notes: the table reports descriptive statistics for both the upper control group (upper panel) and the treatment group (lower panel), before (left side) and after (right side) the threshold revision. *Time* and *Cost Performance* are relative measures of performance - bounded 0 to 1; *Num Time* and *Num Cost Amendments* count the number of amendments per contract, while the relative *Prob* is a binary variable that takes value 1 in case of any amendment occurs; *Avg Time* and *Avg Cost Overrun* account for the average extra time or extra cost and is defined only for contracts which had at least one amendment; *Contract Value* is expressed in US$ thousands; *Offers Received* report the number of offers received per tender.

V Empirical Analysis

In this section, we first explain the econometric strategy used to identify the effect of private oversight and public oversight. Then, we present empirical results and the relative checks for robustness.

V.1 Identification strategy

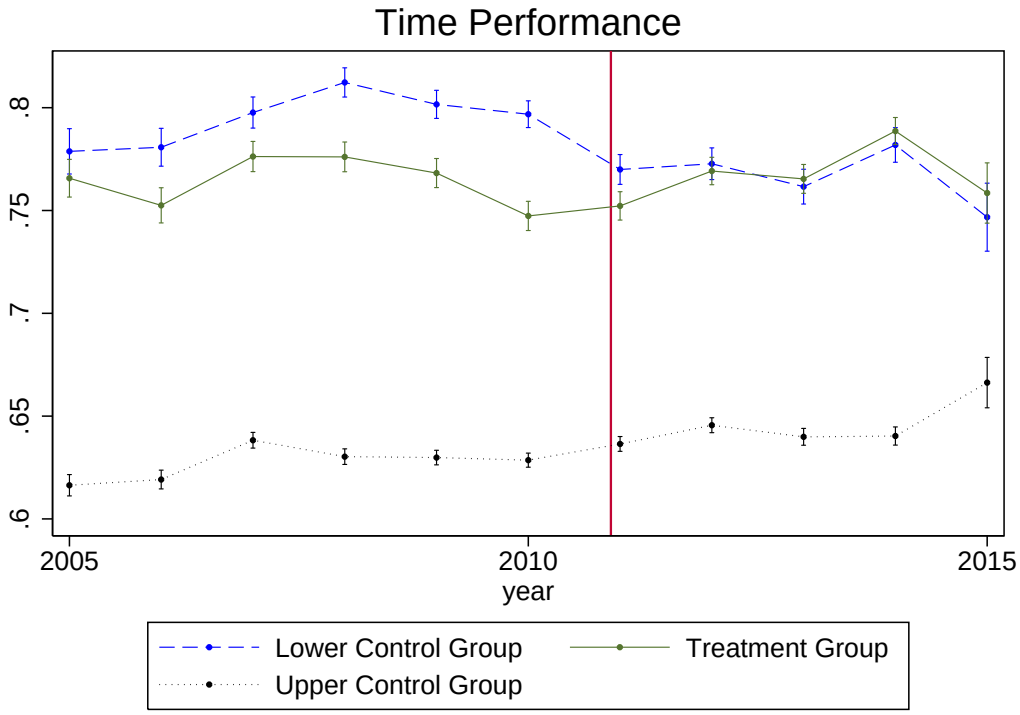
We shall exploit the threshold adjustment in order to identify the effect on performance in terms of cost and time of public oversight and private oversight treatments on all contracts and on constructions contracts only. In principle, we would want to randomly assign the provisions across solicitations and perform a pairwise comparison of the average outcomes of the groups in the two cases. In the absence of a controlled randomized trial, we are forced to turn to non-experimental methods that mimic this under reasonable conditions.

Construction contracts above the oversight threshold are exposed to both public and private oversight, while non-construction contracts are subject only to public oversight. For both construction and all contracts as a whole, the grid in figure (1) identifies the treatment group: in the former case, the treatment results in exemption from both types of oversight; in the latter case it amounts to an exemption from public oversight only. Solid color sections identify upper (dark) and lower (pale) control groups. We start by considering all contracts and present a plain difference-in-differences (DD) strategy. We then focus on the construction/non-construction distinction and discuss how to nest two difference-in-differences analyses through the difference-in-difference-in-differences (DDD, or triple difference) approach.

The simplest framework for a DD estimation requires a set of individuals observed over two periods. A subset of observations - the treatment group - is exposed to a treatment in the second period; the other subset - the control group - is never exposed. Measuring the difference in the average outcome between the groups, while keeping everything else constant, yields the average treatment effect on the treated. The underlying assumption,

which crucially informs the DD identification, states that the difference in expected outcome between the groups is constant across periods, conditional on observables; in other words, one assumes the trends of the variable of interest in the two groups would have been parallel had the treatment not occurred. In our setting, the contract is the unit of observation, the treatment is the waiver of oversight, the periods are determined according to the reform date, and the groups are defined as above. To verify whether the parallel trend assumption is reasonable in our data, we plot the yearly average time series of time performance and cost performance for the Service treatment and both control groups in Figures (2) and (3).[41] The trends appear to be parallel throughout the pre-treatment period.[42]

Figure 2: Time Performance: Yearly Averages - Services



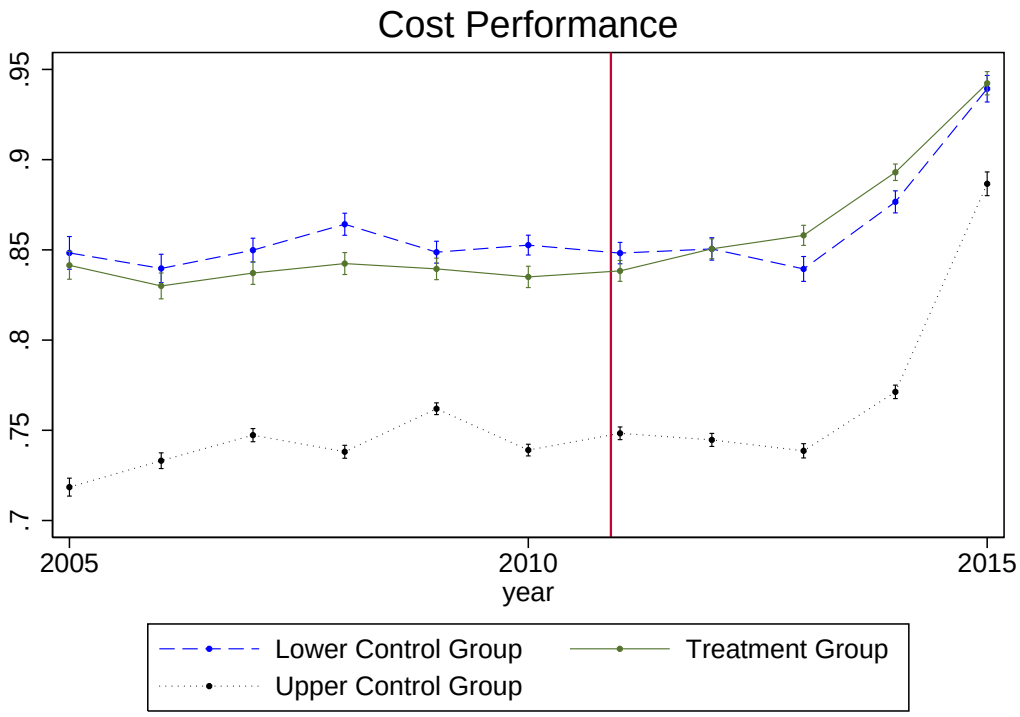
Notes: Trends in yearly averages of *Time Performance* for treatment, treatment, upper and lower control groups. The vertical line corresponds to October, 2010.

If the parallel trends assumption holds, it is then possible to identify the average treat-

[41] Cost performance shows a sharp increase in both the treatment and the control groups. This is possibly due to the presence of uncompleted contract in our sample when approaching the date of download. We show that the results are robust to the narrowing of time window.

[42] Figures (A.6) and (A.7) in appendix show the same graphs for construction contracts.

Figure 3: Cost Performance: Yearly Averages - Services



Notes: Trends in yearly averages of *Cost Performance* for treatment, treatment, upper and lower control groups. The vertical line corresponds to October, 2010.

ment effect on the treated by running a linear regression:

$$Y_{it} = \beta_1 D1_{it} + \beta_2 D2_{it} + \theta_a (D1_{it} * D2_{it}) + \epsilon_{it} \tag{1}$$

where $D1_{it}$ and $D2_{it}$ are binary indicators for group (treatment/control) and period (before/after), respectively. The term $(D1_{it} * D2_{it})$ identifies the treatment and its parameter $\hat{\theta}_a$ amounts to the average treatment effect on the treated. In our setting, in which one treatment is nested onto the other, however, $\hat{\theta}_a$ is biased and the very definition of treatment is ambiguous, as it encompasses effect of the waiver of both public and private oversight. The latter is relevant to treated construction contracts only, but its effect is estimated jointly on the whole sample and cannot be disentangled via a plain DD.

In order to deal with two nested treatments, we rely on an augmented version of the

DD. The triple differences approach nests two DD models like (1) in a single equation and, controlling for the relative differences between treatment and control groups, consistently estimates the average treatment effects.[43] Specifically, starting from equation (1) we define $D3$ as an indicator variable for the subset of individuals subject to the second treatment and augment the model with another tier of differences:

$$Y_{it} = \alpha + \beta_1 D1_{it} + \beta_2 D2_{it} + \theta_a(D1_{it} * D2_{it}) + \beta_3 D3_{it} + \\ + \beta_4(D1_{it} * D3_{it}) + \beta_5(D2_{it} * D3_{it}) + \theta_b(D1_{it} * D2_{it} * D3_{it}) + \epsilon_{it} \quad (2)$$

In equation (2) the triple interaction term $(D1_{it} * D2_{it} * D3_{it})$ indicates the individuals subject to both treatments. In our framework, the coefficients of interest $\theta_i, i \in [a, b]$ capture the effect of the waiver of both types of oversight. As in the case of the plain DD, these are identified as the difference between the observed effects of treatment on the treated and the counter-factual outcome in the absence of treatment, which is assumed to be parallel to that of the control group.

Intensive margin We treat our data as a pooled cross-section and use *upper control group* in the baseline and main robustness specifications.[44] In the core analysis of the paper, we examine the treatment effects on the intensive margin; more specifically, we estimate a DDD on cost and time performance metrics. Indicating the contract outcome variable by Y_{ijt}, we specify the following linear equation:

$$Y_{ijt} = \alpha + \beta_1 Waiver_{it} + \beta_2 Post_{it} + \theta_{public}\left(Waiver_{it} * Post_{it}\right) + \beta_3 W_{it} + \beta_4\left(W_{it} * Waiver_{it}\right) \\ + \beta_5\left(W_{it} * Post_{it}\right) + \theta_{private}\left(Waiver_{it} * Post_{it} * W_{it}\right) + \gamma X_{it} + \zeta_j + \delta_t + \varepsilon_{ijt} \quad (3)$$

[43] See Berck and Villas-Boas [2016], among others, for further details.

[44] The population of contracts in the construction industry never subject to public oversight but always subject to Miller Act provisions and whose estimated value was above $150,000.

where i refers to the contract, j is the contracting office and t indicates the year. $Waiver_{it}$ is the binary variable marking whether the contract value lies between 100,000 and 150,000$ and captures differences between the treatment and control groups prior to the policy change; $Post_{it}$ is a dummy variable for contracts awarded after the reform and captures aggregate factors that would cause changes in Y_{ijt} even in the absence of a policy change and the interaction term $Waiver_{it} * Post_{it}$ captures the effect of exempting contracts from public oversight. W_i is a binary indicator for construction works and the triple interaction term $Waiver_{it} * Post_{it} * W_{it}$ indicates the construction contracts subject to private oversight.[45] Finally, X_{it} are contractor-specific characteristics at the time of the award and ζ_j and δ_t are contracting office and year fixed effects, respectively. The coefficients of interest are θ_{public}, representing the average treatment effect of the exemption from public oversight, and $\theta_{private}$, capturing the effect of the exemption from private oversight.

In order to fully characterize the treatment effect on the treated, we will analyze both the *intensive margin* - the total and average amount - and the *extensive margin* - the probability - of contract amendment. This approach is crucial to unveil the channels through which contractual performance is affected by the reform.

Extensive margin The triple difference analysis identifies the treatment effects on the intensive margin of outcome measures. In order to fully describe the causal effects of the treatments on the performance, we need to investigate whether treated firms are more likely to renegotiate. More specifically, we are interested in assessing the treatment effects on the probability of amend the contract. When not being monitored, firms have more discretionary power during job planning and execution. On the other hand, this leaves room for opportunistic incentives in contract revisions and they may find it more convenient to bargain with the public administration more often at lower amounts. We expect this effect to be even stronger in the construction industry: the decision to renegotiate with the sponsor must be arranged with the surety whose money is at stake, and represents a last resort for contractors. Any minor issue in terms of costs or time could be managed by the surety itself. Hence, in the absence of private oversight, contract revisions become a viable option

[45] In terms of equation (2), $Waiver_{it}$ corresponds to $D1_{it}$, $Post_{it}$ to $D2_{it}$ and W_{it} to $D3$, respectively

to overcome unexpected shocks.

We will test these conjectures running a DDD logistic regression of *Time Amended* and *Cost Amended* on treatments and controls. The above premises underlie a second set of conjectures regarding the intensive margin of amendments. If sureties handle minor issues and help contractors to overcome them without contract revisions, we would expect the average renegotiation to be higher in their presence, since otherwise the sponsor itself has to take care of minor issues. Hence, we proceed with a DDD analysis of the average overrun - *Average Time Overrun* and *Average Cost Overrun* - only for those contracts subject to at least one amendment.

Identification issues and data features The chief concern in our empirical framework is that we do not explicitly observe the engineers' estimated value (EV_i). Since we rely on a combination of i) *SAP*, identifying contracts with public oversight, and ii) their ex-post contract value, we cannot identify the lower bound of treated contracts: this exposes our treatment group sample to the risk of spurious contamination. When testing for robustness of our results, we show that the contract award amount is a good proxy of the engineers' estimate and that the misclassification of contracts to the treatment group is residual.

A very nice feature of our data is that we can run the model on two equally valid sets of control groups: switching from one to the other, as long as the parallel trends assumption holds, should not alter the DDD estimates. In fact, as shown in section V.3, our results are robust to the choice of either group. Finally, it is crucial to remark that contractors decide whether to participate in the tenders and the choice to be subject to the treatment is endogenous. On top of that, the surety company exerts an ex-ante selection on potential contractors, affecting the pool of winners on the quality dimension (see section VI for further details). For all these reasons, a regression discontinuity design (RDD) approach is not a viable option. In order to test for endogenous sorting or discontinuities in the forcing variable, we performed the McCrary [2008] density test for post-law data (see Figure (A.3)). The sharp discontinuity of the running variable at the threshold, highlighted by the graph and confirmed by the highly significant test results, rules out any possibility of running a

usual RDD with our data; see Appendix for further details.

V.2 Results

Triple difference Table (2) reports the DDD regression of contract outcomes - time performance in panel (a) and cost performance in panel (b) - on the treatment variables as defined in equation (3). Column (1) reports results of a plain triple difference model based on equation (1). Specifications (2) to (6) include controls plus an increasing number of fixed effects (bureau, year, state, and contract category). To deal with a collection of minor problems about normality, heteroscedasticity or observations that exhibit large residuals, leverage or influence, standard errors are estimated using the Eicker-Huber-White estimator.[46]

The public oversight treatment ($\hat{\theta}_{public}$) is extremely robust to any choice of controls and fixed effects. In particular, adding bureau fixed effects - column (3) - seems to have significant effects on the magnitude of estimates. Similarly, the effect of a private oversight waiver ($\hat{\theta}_{private}$) is boosted and becomes statistically significant once accounting for bureau specific features. This is not surprising: as shown in Section VI, persistency of the bureau's performance matters in terms of contract outcomes. On the other hand, controlling for year (column (4)), state (column (5)) or object fixed effects does not alter results substantially.

Our baseline estimates, in column (6), show that waiving public oversight positively affects contract performance, but the absence of private oversight offsets such gains. Specifically, we find that the waiver of public oversight effect is positive both in terms of time performance (+7.2%) and cost performance (+5.3%). On the other hand, removing private oversight worsens both measures of performance: it leads to a 9% decrease in terms of time performance and to a 4.2% decrease in terms of cost performance. The composite effect is ambiguous and depends on the dimension considered: adding $\hat{\theta}_{public}$ to $\hat{\theta}_{private}$ allows us to evaluate the composite effect of public oversight and private oversight on construction contract outcomes. The upward shift of the oversight threshold produces an overall decrease

[46]Standard errors estimates are robust to various choices of clusterization level. In table A.5 we report the estimated parameters of the baseline model for both *Time Performance* and *Cost Performance* with standard errors clustered at different levels.

in time performance (-1.8%) and a slight increase in cost performance (+0.6%).[47]

Extensive and Intensive Margin In Table (3) we present the estimated treatment effect on the extensive margin for time outcomes - panel (a) - and cost outcomes - panel (b) - and confront them with the intensive margin of overrun, where we report the treatment effect on the average overrun. Column (1) reports the results of a linear probability model on an indicator function for amendment while column (2) reports the estimates of a DDD - equation (2) - on the average overrun. The model employed for each regression includes all fixed effects and controls of column (6) in Table (2). The probability of a time amendment falls in the Service treatment group (-5.6%), while no significant effect is found for the SB treatment group. Results are similar for cost margin, with an estimated decline in the Service group (-3.2%) and no effect in the SB group. The intensive margin analysis yields a similar picture: public oversight causes lower average overruns in terms of both time and cost (-44.5 days and -$47,351, respectively), while private oversight leads to higher average time overruns (+49.3 days) and lower cost overruns (-$69,692). The latter is in contrast with our baseline finding on the SB treatment effect on cost performance, although the effect is possibly counterbalanced by the non-significance of the effect at the extensive margin. In words, this implies that firms in the Service treatment group are less likely to revise contract terms and, when they do, the overrun is less on both time and cost dimensions; instead, waiving private oversight does not affect the likelihood of revision and has an opposite effect at the intensive margin for cost and time dimensions.

V.3 Robustness Checks

We test the robustness of our baseline DDD findings on three dimensions. We first check in Table (4) whether the results are robust in different subsets of the sample; then, we run the estimation on absolute renegotiation values, i.e. *Cost Overrun* and *Time Overrun*, in Table (5); finally we repeat the baseline exercise with the lower control group.

[47] We calculate the percentage change for constructions only, since both types of oversight apply to these contracts.

Table 2: Triple Difference - Contractual Performances

Panel (a): Time Performance

	(1)	(2)	(3)	(4)	(5)	(6)
$\hat{\theta}_{public}$	0.026***	0.027***	0.050***	0.050***	0.049***	0.048***
	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)
$\hat{\theta}_{private}$	-0.001	-0.002	-0.067***	-0.067***	-0.070***	-0.071***
	(0.025)	(0.025)	(0.025)	(0.025)	(0.026)	(0.026)
N	99,578	99,578	99,578	99,578	99,578	99,578
R^2	0.011	0.014	0.088	0.090	0.093	0.106
$Avg_{services}$	0.763	0.763	0.763	0.763	0.763	0.763
Avg_{works}	0.762	0.762	0.762	0.762	0.762	0.762

Panel (b): Cost Performance

	(1)	(2)	(3)	(4)	(5)	(6)
$\hat{\theta}_{public}$	0.019***	0.017***	0.039***	0.039***	0.039***	0.043***
	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)
$\hat{\theta}_{private}$	0.001	-0.002	-0.028*	-0.030**	-0.036**	-0.039***
	(0.013)	(0.013)	(0.015)	(0.014)	(0.015)	(0.014)
N	100,861	100,861	100,861	100,861	100,861	100,861
R^2	0.030	0.035	0.139	0.142	0.148	0.176
$Avg_{services}$	0.837	0.837	0.837	0.837	0.837	0.837
Avg_{works}	0.920	0.920	0.920	0.920	0.920	0.920
Controls		✓	✓	✓	✓	✓
Bureau Fixed Effects			✓	✓	✓	✓
Year Fixed Effects				✓	✓	✓
State Fixed Effects					✓	✓
Object Fixed Effects						✓

Notes: results of the DDD regression of *Time Performance* - panel (a) - and *Cost Performance* - panel (b) - on public oversight and private oversight treatment indicators. Column (1) reports the results of a plain DDD regression, specification (2) adds controls for firm size, auction type, whether the firm is a limited liability company and whether the contract was signed during the last week of the fiscal year. Columns (3)-(6) include an increasing number of fixed effects (bureau, year, state, and contract object). To deal with a collection of minor problems about normality, heteroscedasticity or observations that exhibit large residuals, leverage or influence, standard errors are estimated using the Eicker-Huber-White estimators. In each panel, $Avg_{services}$ and Avg_{works} account for the average outcome in the Services and the Public Works treatment group, respectively.

Table 3: Triple Difference - Contract Outcomes

Panel (a): Time Margin

	(1) Extensive	(2) Intensive
$\hat{\theta}_{public}$	-0.050***	-40.050***
	(0.009)	(6.990)
$\hat{\theta}_{private}$	0.046	49.174**
	(0.040)	(19.920)
N	99,578	57,010

Panel (b): Cost Margin

	(1)	(2)
$\hat{\theta}_{public}$	-0.029***	-41.124***
	(0.009)	(7.259)
$\hat{\theta}_{private}$	0.021	-69.277***
	(0.040)	(17.053)
N	100,861	54,273

Notes: *Time* - panel (a) - and *Cost* - panel (b) - extensive and intensive margin analysis. For each panel, column (1) reports the results of a linear probability model of *Time/Cost Amended* on public oversight and private oversight treatment indicators plus controls for firm size, auction type, whether the firm is a limited liability company, whether the contract was signed during the last week of the fiscal year, bureau, year, state, and contract object fixed effects. Column (2) reports the estimates of a DDD with the same set of controls on the *Average Time/Cost Overrun*.

Our first set of robustness checks begins in Table (4). We are concerned with possible sources of contamination due to the unobserved engineers' value: we do not observe the *ex-ante* valuation of the project or whether it lay within the $100,000-150,000 band. In order to test whether the misallocation of contracts to the treatment group drives our results, in Table (4) we rule out observations in a (-$10,000, + $10,000) window around $100,000 - *Contamin*, column (2) - and in a (-$15,000, + $15,000) window around $150,000 - *Sanitary*, column (3). Results are robust and, if anything, there is a positive variation in the magnitude of the estimates. In column (4) we run the baseline model on a narrower time window (2008-2013), while in column (5) we reduce the control group to contracts whose values range from

$150,000 to $500,000.

Table 4: Triple Difference - Performance Robustness Checks

Panel (a): Time Performance

	(1) Base	(2) Contamin	(3) Sanitary	(4) 2008-2013	(5) Range	(6) Public Check	(7) Input
$\hat{\theta}_{public}$	0.048***	0.050***	0.051***	0.046***	0.051***	0.047***	0.048***
	(0.006)	(0.006)	(0.007)	(0.007)	(0.007)	(0.008)	(0.006)
$\hat{\theta}_{private}$	-0.071***	-0.065**	-0.071**	-0.060**	-0.066**	-0.070***	-0.069***
	(0.026)	(0.026)	(0.029)	(0.030)	(0.028)	(0.026)	(0.026)
N	99,578	93,529	83,174	68,858	66,787	99,578	99,578
R^2	0.106	0.106	0.105	0.116	0.106	0.106	0.107
$Avg_{services}$	0.763	0.760	0.751	0.751	0.779	0.763	0.763
Avg_{works}	0.762	0.763	0.746	0.752	0.776	0.762	0.762

Panel (b): Cost Performance

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
$\hat{\theta}_{public}$	0.043***	0.044***	0.045***	0.050***	0.042***	0.041***	0.043***
	(0.005)	(0.005)	(0.006)	(0.006)	(0.005)	(0.006)	(0.005)
$\hat{\theta}_{private}$	-0.039***	-0.039***	-0.040**	-0.040**	-0.051***	-0.037**	-0.038***
	(0.014)	(0.015)	(0.017)	(0.016)	(0.016)	(0.015)	(0.014)
N	100,861	94,698	84,177	69,746	67,801	100,861	100,861
R^2	0.176	0.178	0.180	0.183	0.166	0.176	0.177
$Avg_{services}$	0.837	0.831	0.818	0.822	0.850	0.837	0.837
Avg_{works}	0.920	0.915	0.905	0.915	0.924	0.920	0.920

Notes: results of various DDD models of *Time Performance* - panel (a) - and *Cost Performance* - panel (b) - on public and private oversight treatment indicators plus controls for firm size, auction type, whether the firm is a limited liability company, whether the contract was signed during the last week of the fiscal year, bureau, year, state, and contract object fixed effects. *Base* reports the baseline model results; in *Contamin* we account for the possible contamination at the $100,000 threshold dropping all contracts whose face value lies between $90,000 and $110,000; in a similar fashion *Sanitary* model deals with the contamination at both thresholds ($100,000 and $150,000) by dropping two 10% sanitary bands around; *2008-2013* is self-explanatory and *Range* reports results of the same model applied to a sample of contracts trimmed at a face value of $500,000. *Public Check* reports results of a modified model with a specific treatment dummy for public on-site monitoring while in the *Input* model we added controls for public works input prices throughout the period. In each panel, $Avg_{services}$ and Avg_{works} account for the average outcome in the Service and the Public Works treatment group, respectively.

Public officers are specifically demanded to carry out an ex-post monitoring of work status in all services contracts except constructions. In the construction industry, instead, this is a task for the surety. We want to check whether the presence of an additional service-only specific treatment drives our results - mainly with respect to the *public oversight* estimates. Hence, we modify the triple difference model in equation (3) to:

$$Y_{ijt} = \alpha + \beta_1 Waiver_{it} + \beta_2 Post_{it} + \theta_{public}\left(Waiver_{it} * Post_{it}\right) + \beta_3 W_{it} + \beta_4\left(W_{it} * Waiver_{it}\right) + \beta_5\left(W_{it} * Post_{it}\right) + \theta_{private}\left(W_{it} * Waiver_{it} * Post_{it}\right) + \theta_{pc} T_{it}^{pc} + \gamma X_{it} + \delta_t + \zeta_j + \varepsilon_{ijt} \quad (4)$$

where T_{it}^{pc} is a binary treatment variable active for services-only contracts above the threshold. Results of the estimation are reported in column (6): parameter $\hat{\theta}_{pc}$ (unreported) is not statistically significant.

The second set of checks are presented in Table (5), in which the same DDD model is implemented on different dependent variables: panel (a) reports results on *Time Overrun* (in days), while panel (b) reports *Cost Overrun* (in dollars).

Finally, there are two contract groups potentially suitable for use as a control. Although in the baseline regressions we use contracts always subject to oversight (upper control group), we also have data on all those contracts that are never subject to oversight (lower control group). Tables (6)-(7) present the triple difference regressions with lower control group in terms of absolute renegotiation values and performance measures, respectively. Our estimates prove to be robust to the change in control group both in terms of sign and magnitude. In the appendix we report summary statistics for both the lower control group and the treatment group (bottom panel).

Table 5: Triple Difference - Overruns

Panel (a): Time Overrun

	(1)	(2)	(3)	(4)	(5)	(6)
$\hat{\theta}_{public}$	-42.648***	-38.831***	-64.723***	-66.014***	-65.759***	-61.272***
	(6.895)	(7.021)	(7.349)	(7.311)	(7.339)	(7.376)
$\hat{\theta}_{private}$	43.919**	47.287**	123.920***	125.505***	130.784***	124.239***
	(21.674)	(21.641)	(22.028)	(21.973)	(22.319)	(22.244)
N	99,578	99,578	99,578	99,578	99,578	99,578
R^2	0.023	0.025	0.104	0.108	0.111	0.120
$Avg_{services}$	185.279	185.279	185.279	185.279	185.279	185.279
Avg_{works}	118.269	118.269	118.269	118.269	118.269	118.269

Panel (b): Cost Overrun

	(1)	(2)	(3)	(4)	(5)	(6)
$\hat{\theta}_{public}$	15.680	41.824***	-9.115	-9.914	-9.457	-8.565
	(12.308)	(12.569)	(12.652)	(12.602)	(12.572)	(12.512)
$\hat{\theta}_{private}$	-251.752***	-235.036***	-201.133***	-193.493***	-181.137***	-203.886***
	(32.167)	(31.261)	(34.411)	(34.056)	(34.779)	(34.810)
N	100,861	100,861	100,861	100,861	100,861	100,861
R^2	0.043	0.052	0.147	0.149	0.155	0.173
$Avg_{services}$	252.298	252.298	252.298	252.298	252.298	252.298
Avg_{works}	71.257	71.257	71.257	71.257	71.257	71.257
Controls		✓	✓	✓	✓	✓
Bureau Fixed Effects			✓	✓	✓	✓
Year Fixed Effects				✓	✓	✓
State Fixed Effects					✓	✓
Object Fixed Effects						✓

Notes: results of the DDD regression of *Time Performance* - panel (a) - and *Cost Performance* - panel (b) - on public oversight and private oversight treatment indicators. Column (1) reports the results of a plain DDD regression, specification (2) adds controls for firm size, auction type, whether the firm is a limited liability company and whether the contract was signed during the last week of the fiscal year. Columns (3)-(6) include an increasing number of fixed effects (bureau, year, state, and contract object). To deal with a collection of minor problems about normality, heteroscedasticity or observations that exhibit large residuals, leverage or influence, standard errors are estimated using the Eicker-Huber-White estimators. In each panel, $Avg_{services}$ and Avg_{works} account for the average outcome in the Services and the Public Works treatment group, respectively.

VI Discussion

Surety Bonding: Moral Hazard vs. Adverse Selection With the triple difference approach we are able to identify the net effect of surety bonding on contract outcomes.

Table 6: Triple Difference - Lower Control Group (Overruns)

Panel (a): Time Overrun

	(1) Base	(2) Contamin	(3) Sanitary	(4) 2008-2013	(5) Range	(6) Public Check	(7) Input
$\hat{\theta}_{public}$	-70.933***	-66.514***	-59.891***	-72.384***	-87.333***	-51.647***	-70.851***
	(5.365)	(5.884)	(6.143)	(6.204)	(9.491)	(5.819)	(5.363)
$\hat{\theta}_{private}$	64.588***	52.850***	47.403***	53.575***	98.512***	45.278***	61.939***
	(15.010)	(15.946)	(16.798)	(16.978)	(29.932)	(15.181)	(15.096)
N	75,682	61,014	55,361	52,339	26,811	75,682	75,682
R^2	0.087	0.088	0.087	0.089	0.096	0.088	0.088
$Avg_{services}$	137.355	136.092	137.906	141.596	139.084	137.355	137.355
Avg_{works}	73.718	70.541	72.607	71.410	87.024	73.718	73.718

Panel (b): Cost Overrun

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
$\hat{\theta}_{public}$	-70.766***	-61.104***	-52.264***	-74.904***	-83.925***	-56.029***	-70.663***
	(3.949)	(4.119)	(4.168)	(4.755)	(7.783)	(3.940)	(3.945)
$\hat{\theta}_{private}$	51.514***	45.428***	41.333***	47.715***	55.917***	36.737***	51.169***
	(10.916)	(12.286)	(12.108)	(13.189)	(16.883)	(10.892)	(11.039)
N	77,229	62,302	56,551	53,383	27,316	77,229	77,229
R^2	0.080	0.076	0.075	0.083	0.087	0.080	0.080
$Avg_{services}$	48.409	45.658	44.845	54.206	59.337	48.409	48.409
Avg_{works}	16.780	17.295	17.680	18.460	15.078	16.780	16.780

Notes: results of various DDD models of *Time Performance* - panel (a) - and *Cost Performance* - panel (b) - on public and private oversight treatment indicators plus controls for firm size, auction type, whether the firm is a limited liability company, whether the contract was signed during the last week of the fiscal year, bureau, year, state, and contract object fixed effects. *Base* reports the baseline model results; in *Contamin* we account for the possible contamination at the \$100,000 threshold dropping all contracts whose face value lies between \$90,000 and \$110,000; in a similar fashion *Sanitary* model deals with the contamination at both thresholds (\$100,000 and \$150,000) by dropping two 10% sanitary bands around; *2008-2013* is self-explanatory and *Range* reports results of the same model applied to a sample of contracts trimmed at a face value of \$110,000. *Public Check* reports results of a modified model with a specific treatment dummy for public on-site monitoring while in the *Input* model we added controls for public works input prices throughout the period. In each panel, $Avg_{services}$ and Avg_{works} account for the average outcome in the Service and the Public Works treatment group, respectively.

However, the estimates alone do not help distinguish whether this is due to adverse selection, to moral hazard, or to both. In order to test hypotheses *b.1)* and *b.2)*, we partition firms in the SB treatment group into three groups: *Stayers* - those firms that win at least one contract both before and after October, 2010 within the treatment band; *Exiters* that win at least one contract within the band before, but are not awarded a contract worth less than

Table 7: Triple Difference - Lower Control Group (Performance Measures)

Panel (a): Time Performance

	(1) Base	(2) Contamin	(3) Sanitary	(4) 2008-2013	(5) Range	(6) Public Check	(7) Input
$\hat{\theta}_{public}$	0.054***	0.047***	0.041***	0.058***	0.071***	0.042***	0.054***
	(0.005)	(0.005)	(0.006)	(0.006)	(0.008)	(0.005)	(0.005)
$\hat{\theta}_{private}$	-0.013	0.001	0.005	-0.007	-0.030	-0.000	-0.011
	(0.018)	(0.019)	(0.020)	(0.020)	(0.031)	(0.018)	(0.018)
N	75,682	61,014	55,361	52,339	26,811	75,682	75,682
R^2	0.111	0.112	0.113	0.127	0.117	0.111	0.111
$Avg_{services}$	0.802	0.802	0.800	0.803	0.802	0.802	0.802
Avg_{works}	0.828	0.826	0.824	0.830	0.834	0.828	0.828

Panel (b): Cost Performance

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
$\hat{\theta}_{public}$	0.069***	0.062***	0.058***	0.075***	0.084***	0.056***	0.069***
	(0.004)	(0.004)	(0.004)	(0.004)	(0.007)	(0.004)	(0.004)
$\hat{\theta}_{private}$	-0.056***	-0.052***	-0.049***	-0.058***	-0.063***	-0.043***	-0.055***
	(0.010)	(0.010)	(0.011)	(0.011)	(0.017)	(0.010)	(0.010)
N	77,229	62,302	56,551	53,383	27,316	77,229	77,229
R^2	0.127	0.127	0.129	0.142	0.133	0.127	0.127
$Avg_{services}$	0.883	0.883	0.881	0.878	0.881	0.883	0.883
Avg_{works}	0.949	0.948	0.948	0.948	0.945	0.949	0.949

Notes: results of various DDD models of *Time Performance* - panel (a) - and *Cost Performance* - panel (b) - on public and private oversight treatment indicators plus controls for firm size, auction type, whether the firm is a limited liability company, whether the contract was signed during the last week of the fiscal year, bureau, year, state, and contract object fixed effects. *Base* reports the baseline model results; in *Contamin* we account for the possible contamination at the \$100,000 threshold dropping all contracts whose face value lies between \$90,000 and \$110,000; in a similar fashion *Sanitary* model deals with the contamination at both thresholds (\$100,000 and \$150,000) by dropping two 10% sanitary bands around; *2008-2013* is self-explanatory and *Range* reports results of the same model applied to a sample of contracts trimmed at a face value of \$500,000. *Public Check* reports results of a modified model with a specific treatment dummy for public on-site monitoring while in the *Input* model we added controls for public works input prices throughout the period. In each panel, $Avg_{services}$ and Avg_{works} account for the average outcome in the Service and the Public Works treatment group, respectively.

\$150,000 afterwards; and *Entrants*, which are never awarded a contract worth more than \$100,000 before the threshold revision, but win at least one contract in the treatment band afterwards. Within the latter two clusters, we can further identify two subgroups: *Segment Entrants* (*Segment Exiters*) are those firms that enter (exit) the treatment band after being awarded at least one contract below (above) \$100,000 during the pre-treatment period.

According to CGH (hypothesis *b.1*), the pool of winning firms must change with the treatment. This would be reflected in a moderate share of *Stayers* and a high number of both *Exiters* and *Entrants*. On top of that, the share of limited liability firms should be higher in the pool of entrants, given that their aggressive bidding strategy is not counterbalanced by higher surety premia.[48] Conversely, hypothesis *b.2* does not entail any structural adjustment in the set of winning firms: according to this hypothesis sureties do not exert any ex-ante screening and affect the outcomes only through the ex-post supervision of work progress. In this scenario, the same firms are awarded contracts both before and after the reform but their performance is negatively affected by the absence of the surety, whose monitoring role is crucial to avoid misbehavior. Hence, we would expect to observe a relatively high number of *Stayers* and a moderate or null turnover; moreover, *Entrants* and *Exiters* should be similar in size and observable characteristics.

In Table (8), panel *(a)* reports summary statistics of each group within the SB treatment group. The treated sample spans 2008-2013 with the treatment threshold fixed at October, 2010. In order to enhance comparability we repeat the exercise (panels *b* and *c*) on two placebo subsamples spanning 5 years but not containing the threshold revision date - i.e. all contracts before and all contracts after October, 2010. We place two placebo thresholds in 2008 and 2012, respectively. *Stayers* constitute 6.8% of the sample in the treatment group - *Panel (a)* - while *Entrants* and *Exiters* account for 31.2% and 62%, respectively. The *Stayers* figure is similar and remains low in all subsamples, this being an indicator of a high turnover level in the market, but the proportion of exiters to entrants is reversed in both panels *(b)* and *(c)* with respect to the treatment group.[49] This indicates that an unusually high number of firms exit the federal construction market following the 2010 reform - three times as many as those in the Placebo 1 subsample and almost 7 times the number of *Exiters* in the Placebo 2 subsample - but are not replaced by a comparable number of firms.[50]

[48]Limited liability firms are able to bid more aggressively because their financial responsibility in case of loss is bounded. For more details on this, see CGH.

[49]See, for example, the Bureau of Labor Statistics on turnover levels at http://www.bls.gov/iag/tgs/iag23.htm.

[50]Weighting the comparison for the ratio of total number of firms in the samples yields 2.45 and 4.57 times more, respectively.

An interesting piece of evidence in favor of the presence of adverse selection is provided by the *Segment Entrants* and the *Segment Exiters* figures. The share of exiters is remarkably regular across the three panels, lingering below the 5 percentage points.[51] On the other hand, the volume of *Segment Entrants* varies substantially, starting at similar levels before the reform and increasing dramatically during treatment period (15%), only to decrease again (9%) afterwards. This is consistent with hypothesis *b.1)*: a negligible share of firms step up from below the $100,000 threshold to the treatment group when sureties screen bidders ex ante, while the transition is relatively easy in their absence. We document an analogous pattern in the share of entrant limited liability companies, which peaks during the treatment period and decreases afterwards. Following CGH, we interpret this as a signal of adverse selection.

Public Monitoring: Competence and the Red Tape Effect In public procurement, the quality of the buyer is a crucial feature to account for. Competence in providing goods and services helps in selecting the best contractors, properly designing contracts and avoiding misconduct. Moreover, competent offices are able to minimize the red tape effect of bureaucracy through rapid and simple monitoring activities. Thus, we would expect the public oversight treatment effect to be particularly intense in contracting offices with low levels of competence and, conversely, contracts managed by expert officials not to be affected. Decarolis et al. [2017] already showed the importance of assessing the contracting office's competence in dealing with public procurement performance using the same data. In order to do this, they propose a measure of office quality defined as the persistency of contract performance within the same purchasing organization and for the same category of procured good or service over time. More specifically, for each contract awarded, they measure the weighted average outcome (in terms of past cost performance and past time performance, separately) of the contract awarded by the same bureau in the past. They differentiate by category in order not to homogenize competence in different sectors within the same federal office. The same rationale applies to our framework, and we build the same measure of

[51] We obtain similar results for segment Exiters with different placebo subsamples, where we define placebo treatments according to different levels of contract value thresholds. Statistics are available upon request.

Table 8: Summary statistics: Stayers, Entrants and Exiters in the Construction Market

Panel (a): Treated - 2010 Threshold

	Stayers			Entrants			Exiters		
	Mean	Median	N	Mean	Median	N	Mean	Median	N
No. of Employees	5,868.09	41.4	37	3,119.84	18	165	158.17	12	336
Annual Revenue	1,443.38	12.1	37	293.29	3.40	165	62.09	2.17	336
Limited Company	0.16	0	37	0.21	0	165	0.10	0	336
Past Time Perf e	0.79	0.80	37	0.80	0.80	161	0.82	0.83	334
Past Cost Perf e	0.90	0.92	37	0.87	0.89	161	0.89	0.90	334
Segment Entrants				0.15	0	165			
Segment Exiters							0.04	0	336

Panel (b): Placebo 1 - 2008 Threshold

	Stayers			Entrants			Exiters		
	Mean	Median	N	Mean	Median	N	Mean	Median	N
No. of Employees	7,164.09	11.9	26	225.48	10.5	307	1,321.22	25	113
Annual Revenue	1,365.78	5.71	26	74.77	2.15	307	102.08	3	113
Limited Company	0.08	0	26	0.12	0	307	0.03	0	113
Past Time Perf e	0.78	0.75	26	0.82	0.84	306	0.81	0.81	109
Past Cost Perf e	0.91	0.93	26	0.89	0.90	306	0.91	0.93	109
Segment Entrants				0.05	0	307			
Segment Exiters							0.04	0	113

Panel (c): Placebo 2 - 2012 Threshold

	Stayers			Entrants			Exiters		
	Mean	Median	N	Mean	Median	N	Mean	Median	N
No. of Employees	2,256.63	34.1	19	3,828.86	19	144	455.11	30	51
Annual Revenue	684.95	11.5	19	484.14	4.35	144	79.47	4.50	51
Limited Company	0.26	0	19	0.19	0	144	0.18	0	51
Past Time Perf e	0.77	0.75	19	0.80	0.80	142	0.80	0.81	48
Past Cost Perf e	0.92	0.93	19	0.86	0.86	142	0.90	0.93	48
Segment Entrants				0.10	0	144			
Segment Exiters							0.04	0	51

Notes: descriptive statistics relative to the group of *Stayers*, *Entrants* and *Exiters*. These have been defined according to the actual threshold revision in 2010 - panel (a) - or with two placebo reform dates in 2008 - panel (b) - and 2012 - panel (c).*No. of Employees* is the average number of employees over the time span considered; *Annual Revenue* reports the last 3 years' average revenue - in US$ thousands ; *Limited Company* indicates a limited liability company; *Past Time* and *Past Cost Performance* are defined as in Decarolis et al. [2017]. *Segment Entrants* indicates the share of Entrant firms which only appeared below the threshold prior to October, 2010, while *Segment Exiters* reports the number of Exiter firms which appear only above the threshold before the reform.

past performance for each contract i as $past\,performance_{ik} = \frac{\sum_{j_k<i_k} w_{j_k} * performance_{j_k}}{\sum_{j_k} w_{j_k}}$, where k stands for the contracting office, $performance_{j_k}$ is the performance value achieved in the j^{th} contract awarded by k and w_{j_k} are the Bartlett weights for smoothing past observations. We repeat the exercise for both cost and time performance and further partition our sample into *High* (H)- and *Low* (L)-type contracting office competence, depending on the median value. In Table (9) we report the results of our baseline DDD estimation on High and Low contracts subsamples. We find that the detrimental effect of red tape is large and significant only for contracts awarded by less competent contracting offices. Hence, public oversight worsens contractors' performance if and only if it is carried out by a low-quality bureau. The scores for offices with a track record of excellence in project management do not alter after the treatment.

VII Conclusions

Contracting authorities rely on procurement both to fulfill their standard activities and as a mean to indirectly promote technological progress and the development of strategic sectors. As public procurement is a major source of public spending, contractual, legal and institutional design play a crucial role in all contemporary economies. Oversight is among the most keenly debated tools for policy-makers and practitioners due to its heterogeneous nature: it aims at reducing moral hazard and asymmetry of information, but it is frequently a source of red tape, waste and inefficiency. This paper contributes to the ongoing discussions and to the literature on optimal procurement practices by providing field evidence of the effect of private oversight and public oversight on contractual performance. Our estimates show that the two effects are competing and that their magnitude is comparable. This is especially relevant from a policy perspective as it delineates a feasible strategy for future reforms, i.e. moving thresholds independently and keeping the benefits of both private and public oversight. Our complementary exercises allow us also to disentangle *private moral hazard* from *adverse selection*, and *public moral hazard* from *red tape*, respectively.

The positive estimate of the effect of public oversight on performance shows that the red

Table 9: Triple Difference - High- and Low-type contracting offices

Panel (a): Performance Measures

	Time Performance		Cost Performance	
	(1)	(2)	(3)	(4)
	H	L	H	L
$\hat{\theta}_{public}$	0.018	0.084***	-0.001	0.079**
	(0.036)	(0.024)	(0.023)	(0.030)
N	50,612	48,966	49,027	51,834
R^2	0.096	0.056	0.126	0.102
$Avg_{services}$	0.772	0.754	0.852	0.820
Avg_{works}	0.764	0.757	0.916	0.928

Panel (b): Renegotiation Measures

	Time Overrun		Cost Overrun	
	H	L	H	L
$\hat{\theta}_{public}$	-26.989	-107.952***	47.000	-54.045
	(46.308)	(37.238)	(78.588)	(85.374)
N	50,612	48,966	49,027	51,834
R^2	0.095	0.110	0.136	0.118
$Avg_{services}$	188.459	181.852	152.530	231.897
Avg_{works}	1.000	149.418	57.439	44.715

Notes: DDD regressions of *Time Performance* and *Cost Performance* - panel (a) - and *Time Overrun* and *Cost Overrun* - panel (b) - on public and private oversight treatment indicators plus controls for firm size, auction type, whether the firm is a limited liability company, whether the contract was signed during the last week of the fiscal year, bureau, year, state, and contract object fixed effects. Odd columns report results for contracts managed by high-competence contracting offices, even columns refer to low-competence. Standard errors are in parentheses. In each panel, $Avg_{services}$ and Avg_{works} account for the average outcome in the Services and the Public Works treatment group, respectively.

tape effect outweighs the moral hazard issue for low-value contracts. Hence, enforcing public oversight leads to extra time overrun and cost overrun. On the other hand, the interpretation of the negative estimated parameter of private oversight is not straightforward. The result is consistent with both the facts that sureties carry out an ex-ante selection of bidders and that firms renegotiate less because they exert higher effort under the supervision of sureties. We further investigate the dynamics of treatment effects through an extensive/intensive margin analysis. This reveals that public oversight leads contractors to renegotiate fewer times - mostly over cost overrun - and for lower amounts, while exempting firms from issuing surety bonds does not affect their probability of renegotiating. We also present descriptive evidence on the relationship between competence of the contracting office and treatment effects. Red tape is associated with incompetent officers, while contracts awarded by high-competence bureaus do not alter their performance depending on the level of public oversight. A second descriptive exercise clarifies the role of the surety companies in both the ex-ante selection of contractors and the ex-post monitoring of progress. To summarize, public and private oversight have different impacts in terms of sign but comparable impacts in terms of magnitude on contract outcomes. Moreover, they affect performance through distinct, non-overlapping channels. As a result, we find that the net effects of the upward oversight threshold revision are discordant and document an improvement in cost performance but a decline in time performance for construction contracts.

Most interestingly from a policy perspective, we show that private and public oversight yield competing effects. Therefore, their thresholds could be more efficiently exploited if moved independently: public oversight could be enforced on high-value contracts only - reducing the risk and the extent of red tape - without affecting the cutoff for surety bond requirements on low-value contracts. However, private oversight, whose benefits show that the private sector can ensure the public interest, comes at a cost. Surety bonds premia are included in the bids and passed on to the sponsor; therefore, the contract price rises accordingly. Future research - with more detailed data on bid distribution, premia amounts, engineers' ex-ante evaluations, and the direct cost of public oversight - should exploit a similar framework to run a more comprehensive welfare analysis.

References

Athey, S., Levin, J. and Seira, E. [2011]. Comparing open and sealed-bid auctions: Evidence from timber auctions., *The Quarterly Journal of Economics* **126**: 207–257.

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

Bajari, P. and Lewis, G. [2014]. Moral hazard, incentive contracts, and risk: Evidence from procurement, *The Review of Economic Studies* **forthcoming**.

Bajari, P., McMillan, R. S. and Tadelis, S. [2009]. Auctions versus negotiations in procurement: An empirical analysis, *Journal of Law, Economics and Organization* **25**(2): 372–399.

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

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

Banerjee, A. V. and Duflo, E. [2000]. Reputation effects and the limits of contracting: a study of the Indian Software Industry, *The Quarterly Journal of Economics* **115**(3): 989–1017.

Berck, P. and Villas-Boas, S. [2016]. A note on the triple difference in economic models, *Applied Economics Letters* .

Bergman, A. M., Johansson, P., Lundberg, S. and Spagnolo, G. [2016]. Privatization and quality: Evidence from elderly care in sweden, *Journal of Health Economics* **49**: 109–119.

Bertrand, M., Duflo, E. and Mullainathan, S. [2004]. How much should we trust differences-in-differences estimates?, *The Quarterly Journal of Economics* **119**(1): 249–275.

Board, S. [2007]. Bidding into the red: A model of post-auction bankruptcy, *Journal of Finance* **62**(6): 2695–2723.

Bozeman, B. [1993]. A Theory Of Government "Red Tape", *Journal of Public Administration Research and Theory* **3**(3): 273–303.

Branzoli, N. and Decarolis, F. [2015]. Entry and subcontracting in public procurement auctions, *Management Science* **61**(12): 2945 – 2962.

Calveras, A., Ganuza, J. and Hauk, E. [2004]. Wild bids. gambling for resurrection in procurement contracts, *Journal of Regulatory Economics* **26**(1): 41–68.

Calvo, E., Cui, R. and Serpa, J. [2016]. Oversight and efficiency in public projects: A regression discontinuity analysis, *Working Paper* .

Calzolari, G. and Spagnolo, G. [2009]. Relational contracts and competitive screening, *CEPR Discussion Papers* .

Coviello, D., Guglielmo, A. and Spagnolo, G. [2017]. The effect of discretion on procurement performance, *Management Science* **63**(5).

Coviello, D. and Mariniello, M. [2014]. Publicity requirements in public procurement: Evidence from a regression discontinuity design, *Journal of Public Economics* **Forthcoming**.

De Silva, D. G., Dunne, T., Kosmopoulou, G. and Lamarche, C. [2017]. Contract Modification and Bidding in Highway Procurement Auctions, *mimeo* .

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., Giuffrida, L., Iossa, E., Mollisi, V. and Spagnolo, G. [2017]. Buyer quality and procurement outcomes: Explorative evidence from the u.s., *Working Paper* .

Decarolis, F., Pacini, R. and Spagnolo, G. [2016]. Past performance and procurement outcomes, *Working Paper* .

Di Tella, R. and Schargrodsky, E. [2003]. The role of wages and auditing during a crackdown on corruption in the city of buenos aires, *Journal of Law & Economics* **46**(1): 269–292.

Ferraz, C. and Finan, F. [2008]. Exposing corrupt politicians: the effects of brazil's publicly released audits on electoral outcomes, *The Quarterly Journal of Economics* **123**(2): 703–745.

Ferraz, C. and Finan, F. [2011]. Motivating politicians: The impacts of monetary incentives on quality and performance, *Working Paper* .

Fisman, R. and Gatti, R. [2006]. Bargaining for bribes: The role of institutions, *CEPR Discussion Papers* .

Fisman, R. and Miguel, E. [2007]. Corruption, norms, and legal enforcement: Evidence from diplomatic parking tickets, *Journal of Political Economy* **115**(6).

GAO, U. [2013]. Agencies need to strengthen oversight of multibillion dollar investments in operations and maintenance, *Technical Report GAO-14-66*, United States Government Accountability Office.

Garvin, M., Molenaar, K., Navarro, D. and Proctor, G. [2011]. Key performance indicators in public-private partnerships, *Technical Report FHWA-PL-10-029*, US Department of Transportation.

Gransberg, D. D., Kraft, P. and Park, H. [2014]. Performance bond: Cost, benefit, and paradox for the public highway 3 agencies, *Transportation Research Record* **2408**: 3–9.

Guasch, J. L., Laffont, J. and Straub, S. [2008]. Renegotiation of concession contracts in latin america: Evidence from the water and transport sectors, *International Journal of Industrial Organization* **22**(7): 1267–1294.

Hart, O., Shleifer, A. and Vishny, R. W. [1997]. The proper scope of government: Theory and an application to prisons, *Quarterly Journal of Economics* **112**(4): 1127–1161.

Hyytinen, A., Lundberg, S. and Toivanen, O. [2009]. Politics and procurement: Evidence from cleaning contracts, *HEER Working Paper* .

Kaufman, H. [1977]. *Red Tape: Its origins, uses, and abuses*, Brookings Institution Press.

Kelman, S. [1990]. *Procurement and Public Management: The Fear of Discretion and the Quality of Government Performance*, University Press of America Lanham.

Kelman, S. [2005]. *Unleashing Change: A Study of Organizational Renewal in Government*, Brookings Institution Press.

Krasnokutskaya, E. and Seim, K. [2011]. Bid preference programs and participation in highway procurement auctions, *American Economic Review* **101**(6): 2653–2686.

Lewis, G. and Bajari, P. [2017]. Procurement Contracting with Time Incentives: Theory and Evidence, *The Quarterly Journal of Economics* **forthcoming**.

Liebman, J. B. and Mahoney, N. [2016]. Do expiring budgets lead to wasteful year-end spending? evidence from federal procurement, *Working Paper* .

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

Marion, J. [2009]. How costly is affirmative action? government contracting and california's proposition 209, *The Review of Economics and Statistics* **91**(3): 503–522.

McCrary, J. [2008]. Manipulation of the running variable in the regression discontinuity design: A density test, *Journal of Econometrics* .

Olken, B. A. [2006]. Corruption and the costs of redistribution: Micro evidence from indonesia, *Journal of Public Economics* .

Olken, B. A. [2007]. Monitoring corruption: Evidence from a field experiment in indonesia, *Journal of Political Economy* **115**(2).

Palguta, J. and Pertold, F. [forthcoming]. Manipulation of procurement contracts: Evidence from the introduction of discretionary thresholds, *American Economic Journal: Economic Policy* .

Reinikka, R. and Svensson, J. [2004]. Local capture: Evidence from a central government transfer program in uganda, *The Quarterly Journal of Economics* .

Shavell, S. [1984]. A model of the optimal use of liability and safety regulation, *RAND Journal of Economics* **15**: 271–280.

Shleifer, A. and Vishny, R. W. [1998]. *The Grabbing Hand: Government Pathologies and Their Cures*, Harvard University Press.

Spiller, P. T. [2008]. An institutional theory of public contracts: regulatory implications, *NBER Working Paper* (14152): 1–33.

A Appendix

A.1 Data features

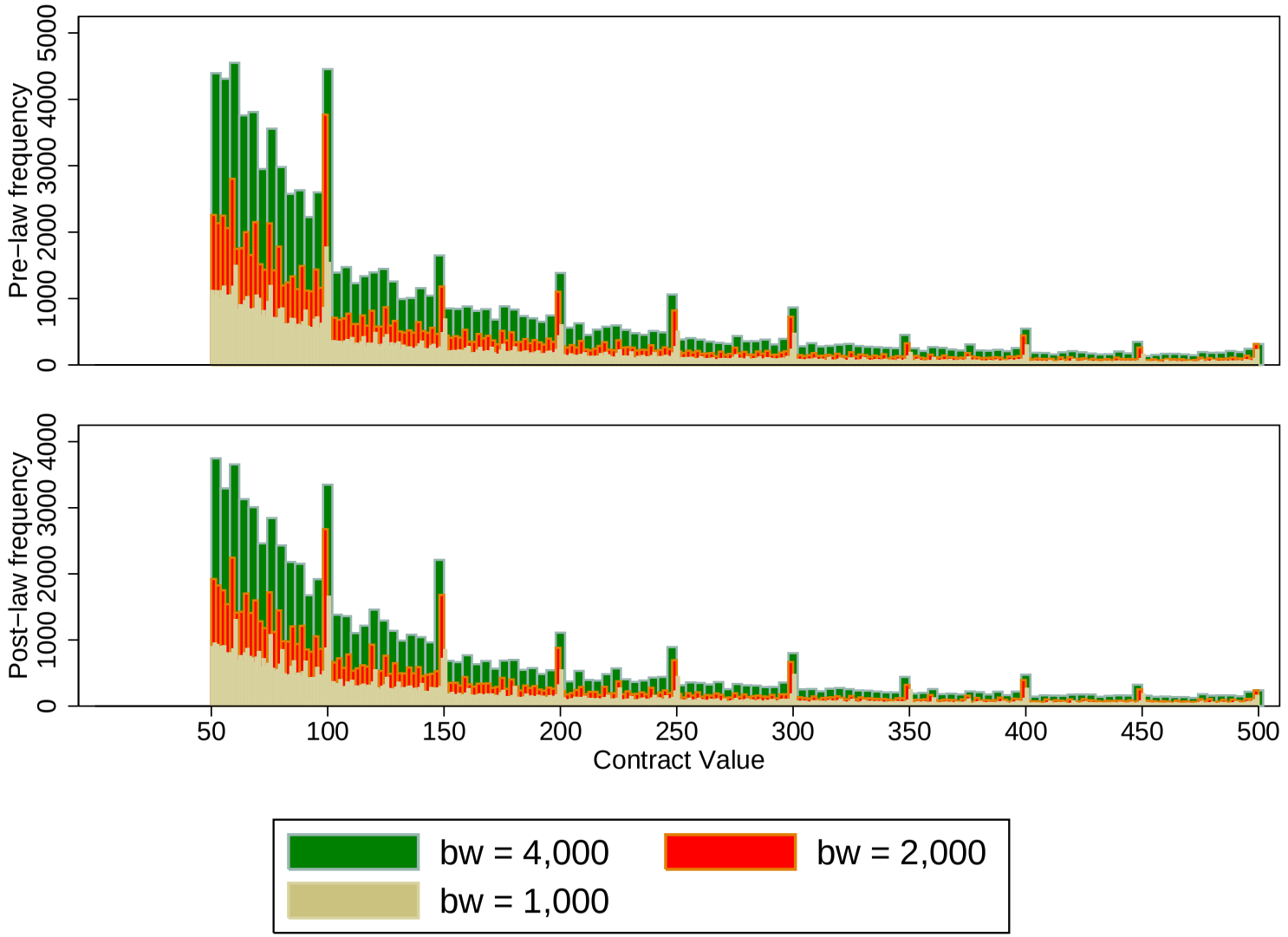
The previous literature used legislative thresholds in RDD in order to analyze whether contract-allocation procedures with different levels of discretion affect procurement outcomes (see Coviello and Mariniello [2014], Coviello et al. [2017]). As shown in Figure (A.1), most of the "round" values - i.e. multiples of $50,000 - show significant jumps in frequency both in pre- and post-law update: this is a clear indication of sorting before different procedural thresholds. Some of these cutoffs, such as the one at $200,000, seem to be mainly psychological, since FAR provides no legislative provision attached to these contracts values. As a result, due to this likely endogenous sorting of contracts both prior to and after October, 2010, the performance of firms above and below the threshold may have been different even before the reform. Had we ignored this and performed an RDD analysis, we could have retrieved severely upward biased estimates due to a pre-existent discontinuity.[52]

In order to test for endogenous sorting or discontinuities in the forcing variable, we performed the McCrary [2008] density test for post-law data for both construction (Figure A.2) and services (Figure A.3). The sharp discontinuity of the running variable at the $150,000 threshold, highlighted by the graph and confirmed by the significant test results, rules out any possibility of running a usual RDD with our data.

The endogenous sorting and its increment after the update of the SAT threshold are clear-cut evidence of the facts that (i) winning firms' incentives to sort themselves below $150,000 became stronger, (ii) the effect of the confounding policy discontinuity on the potential outcome is not constant over time, and (iii) this effect was the same in the pre- and post-treatment period (as Figure (A.4), displaying the pre- and post-reform contract value density around the $150,000 threshold, displays). This is confirmed by Figure (A.5), showing that pre-reform contract value density around the $100,000 threshold is higher than post-reform.

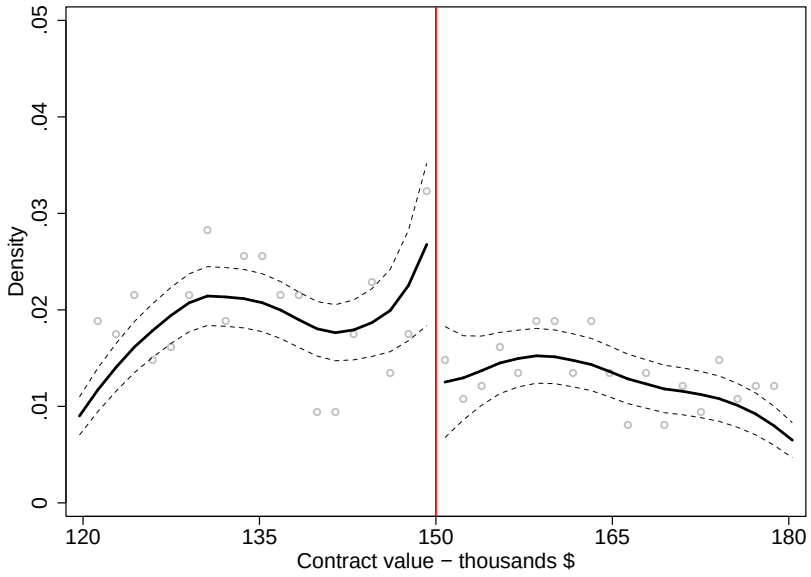
[52] An instance of bunching in procurements just below legislative thresholds is presented in Palguta and Pertold [forthcoming]

Figure A.1: Contract frequency: various binsizes



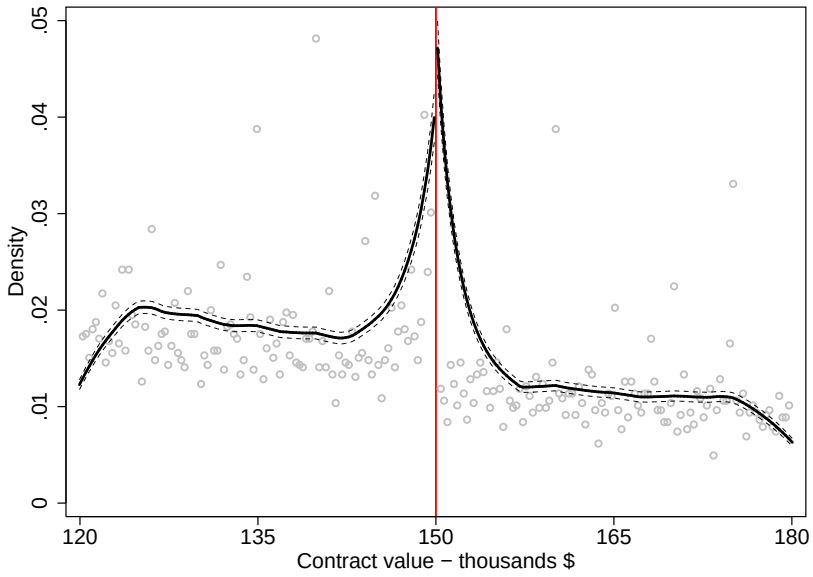
Note: Contract frequencies pre- and post-reform, absolute values. The histogram is plotted for different bin widths: \$4,000 (green), \$2,000 (red) and \$1,000 (yellow).

Figure A.2: McCrary Density Test - Construction Contracts



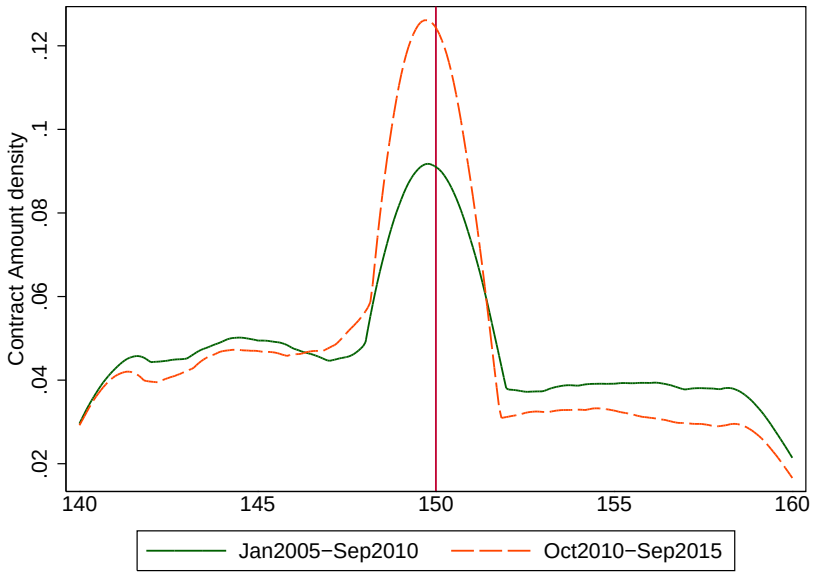
Notes: Mccrary density in a +/- $30,0000 window around the $150,000 threshold. The dots represent the density of projects in different intervals of project budget, the solid line represent a kernel estimate of the density, and the two dashed lines are 95% confidence intervals. Construction contracts only.

Figure A.3: McCrary Density Test - Full Sample



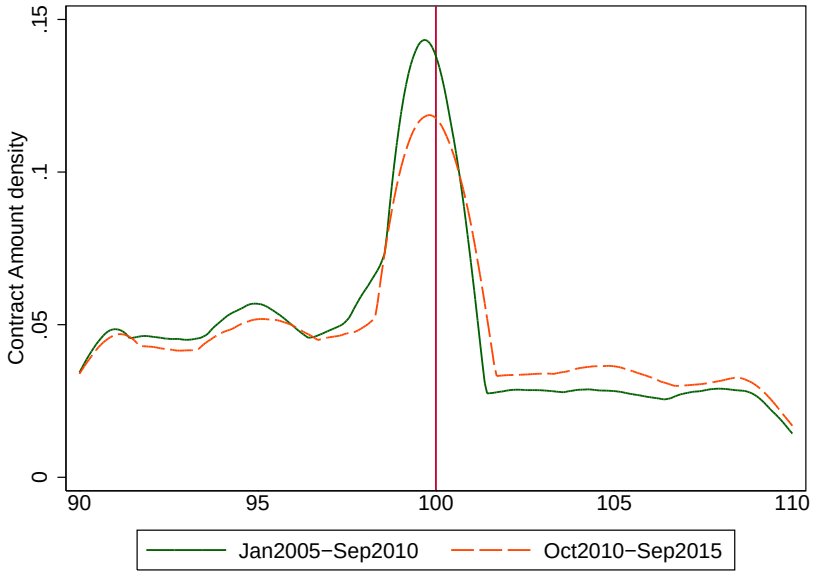
Notes: McCrary density in a +/- $30,000 window around the $150,000 threshold. The dots represent the density of projects in different intervals of project budget, the solid line represent a kernel estimate of the density, and the two dashed lines are 95% confidence intervals.

Figure A.4: Pre- and Post-reform contract value density



Notes: Contract amount density in a +/- $20,000 window around the $150,000 threshold before (solid) and after (dotted) - the threshold revision.

Figure A.5: Pre- and Post-reform contract value density - 100,000



Notes: Contract amount density in a +/- $20,000 window around the $100,000 threshold before (solid) and after (dotted) the threshold revision.

A.2 Summary statistics

Table A.1 reports selected summary statistics for the lower control group of services - along with those for the treatment group, already presented in 1 - before and after the October, 2010 reform. In Table A.2 we report summary statistics for the constructions sample and distinguish, again, the above treatment/control before/after groups scheme. Tables A.3 and A.4 display the number of contracts by department and year before and after the reform when considering the upper and the lower control group, respectively, for the analysis.

Table A.1: Summary statistics - Services Treatment Group with Lower Control Group

	Lower Control Group							
	Before				After			
	Mean	SD	Median	N	Mean	SD	Median	N
Time Performance	0.7	0.3	0.74	51,224	0.7	0.3	0.82	52,692
Num Time Amendments	2.2	2.7	1	62,013	1.3	2.0	0	64,434
Prob Time Revision	0.6	0.5	1	62,013	0.5	0.5	0	64,434
Avg Time Overrun	216.2	282.8	130.5	33,223	217.6	293.4	114.7	27,887
Cost Performance	0.7	0.3	0.92	51,854	0.8	0.3	0.99	53,451
Num Cost Amendments	2.6	3.0	1	62,013	1.8	2.4	1	64,434
Prob Cost Revision	0.7	0.5	1	62,013	0.6	0.5	1	64,434
Avg Cost Overrun	274.0	334.4	205.1	31,626	307.7	413.5	176.6	27,457
Contract Value	2,051.9	56,831.2	414.4	62,013	1,068.3	4,139.9	297.5	64,434
# Contractual Days	403.6	395.5	364	62,012	296.7	220.0	347	64,434
Offers received	5.2	18.1	2	62,013	5.5	17.7	2	64,434

	Treatment Group							
	Before				After			
	Mean	SD	Median	N	Mean	SD	Median	N
Time Performance	0.8	0.3	1	13,771	0.8	0.3	1	2,061
Num Time Amendments	1.1	1.9	0	16,834	0.8	1.5	0	2,726
Prob Time Revision	0.4	0.5	0	16,834	0.4	0.5	0	2,726
Avg Time Overrun	251.0	318.1	146.1	6,094	194.2	265.3	92	926
Cost Performance	0.8	0.3	1	14,013	0.9	0.2	1	2,128
Num Cost Amendments	1.3	2.1	0	16,834	1.0	1.6	0	2,726
Prob Cost Revision	0.5	0.5	0	16,834	0.4	0.5	0	2,726
Avg Cost Overrun	101.1	125.8	64.2	5,386	67.6	96.7	34.3	626
Contract Value	122.0	15.2	121.0	16,834	121.6	16.4	120	2,726
# Contractual Days	292.8	336.1	213	16,833	225.5	196.4	196.5	2,726
Offers received	4.6	13.8	2	16,834	3.4	16.5	1	2,726

Notes: the table reports descriptive statistics for both the Lower Control Group (upper panel) and the treatment group (lower panel), before (left side) and after (right side) the threshold revision. *Time* and *Cost Performance* are relative measures of performance - bounded 0 to 1; *Num Time* and *Num Cost Amendments* count the number of amendments per contract, while the relative *Prob* is a binary variable that takes value 1 in case of any amendment occurs; *Avg Time* and *Avg Cost Overrun* account for the average extra time or extra cost and is defined only for contracts which had at least one amendment; *Contract Value* is expressed in US$ thousands; *Offers Received* report the number of offers received per tender.

Table A.2: Summary statistics - Public Works Sample with Upper Control Group

	Upper Control Group							
	Before				After			
	Mean	SD	Median	N	Mean	SD	Median	N
Time Performance	0.7	0.3	0.67	5,787	0.7	0.3	0.76	2,788
Num Time Amendments	3.1	3.0	2	6,546	2.4	2.8	1	3,249
Prob Time Revision	0.7	0.4	1	6,546	0.7	0.5	1	3,249
Avg Time Overrun	121.5	169.9	71.5	4,538	106.0	150.8	58.9	1,990
Cost Performance	0.9	0.2	0.94	5,816	0.9	0.2	0.96	2,836
Num Cost Amendments	3.4	3.3	2	6,546	2.6	3.0	1	3,249
Prob Cost Revision	0.8	0.4	1	6,546	0.7	0.5	1	3,249
Avg Cost Overrun	123.4	186.6	48.0	4,473	190.1	357.1	66.8	1,878
Contract Value	7,985.3	173,000.2	636.1	6,546	4,383.3	11,851.6	498.8	3,249
# Contractual Days	326.3	281.9	238.5	6,546	265.4	217.2	196	3,249
Offers received	4.3	4.0	3	6,546	4.2	4.4	3	3,249

	Treatment Group							
	Before				After			
	Mean	SD	Median	N	Mean	SD	Median	N
Time Performance	0.7	0.3	0.95	947	0.8	0.3	1	90
Num Time Amendments	1.1	1.6	0	1,110	0.6	1.0	0	102
Prob Time Revision	0.5	0.5	0	1,110	0.3	0.5	0	102
Avg Time Overrun	150.5	200.7	86.5	507	129.2	128.1	99	33
Cost Performance	0.9	0.2	1	963	0.9	0.1	1	91
Num Cost Amendments	1.1	1.5	1	1,110	0.6	1.0	0	102
Prob Cost Revision	0.5	0.5	1	1,110	0.4	0.5	0	102
Avg Cost Overrun	39.0	99.8	10.8	447	30.7	36.2	17.3	26
Contract Value	123.8	14.6	123.5	1,110	123.2	14.3	123.5	102
# Contractual Days	164.7	201.3	111	1,110	173.1	179.6	114.5	102
Offers received	3.6	10.6	3	1,110	2.6	1.7	3	102

Notes: the table reports descriptive statistics for both the Public Works Upper Control Group (upper panel) and the Public Works treatment group (lower panel), before (left side) and after (right side) the threshold revision. *Time* and *Cost Performance* are relative measures of performance - bounded 0 to 1; *Num Time* and *Num Cost Amendments* count the number of amendments per contract, while the relative *Prob* is a binary variable that takes value 1 in case of any amendment occurs; *Avg Time* and *Avg Cost Overrun* account for the average extra time or extra cost and is defined only for contracts which had at least one amendment; *Contract Value* is expressed in US$ thousands; *Offers Received* report the number of offers received per tender.

Table A.3: Number of contracts by department and year - Upper control group and treatment group

	Pre-Treatment						Post-Treatment				
	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015
Dept of Agriculture	556	257	205	208	169	354	214	177	151	166	22
Dept of Commerce	2,718	2,437	2,166	2,313	2,574	2,721	2,479	2,037	1,790	1,739	382
Dept of Defense	755	829	946	1,171	1,317	1,192	959	1,685	1,124	994	110
Dept of Education	48	63	91	66	76	77	43	54	33	20	1
Dept of Energy	63	57	48	60	64	68	58	70	59	54	8
Dept of Health and Human Services	252	412	552	1,069	1,104	1,482	807	218	69	30	11
Dept of Homeland Security	47	43	54	44	31	60	30	35	28	30	3
Dept of Housing	113	185	186	185	177	257	185	179	156	130	24
Dept of Interior	1,431	1,560	1,589	1,513	2,064	2,646	1,504	1,440	1,260	1,332	337
Dept of Justice	852	791	933	881	1,007	1,124	1,051	1,063	998	1,097	122
Dept of Labor	2,159	7,750	13,152	14,188	15,173	14,545	13,675	13,397	9,639	8,642	1,232
Dept of State	130	112	103	119	96	103	103	89	91	142	12
Dept of Treasury	274	336	275	354	333	396	249	262	235	238	41
Dept of Transportation	2,086	2,128	2,207	2,222	2,154	2,317	2,388	2,056	1,862	1,799	234
Dept of Veteran Affairs	2,489	2,374	2,537	2,658	2,400	2,522	2,582	2,627	2,186	2,121	478
Environmental Protection Agency	288	251	263	245	249	292	207	79	79	62	5
General Services Administration	2,707	2,017	1,939	1,854	2,152	3,008	2,313	2,252	1,947	1,988	282
NASA	1,623	1,781	2,534	2,508	2,558	2,674	3,199	3,259	3,284	3,252	862
Nuclear Regulatory Commission	231	171	196	198	261	434	393	381	317	332	69
National Science Foundation	1,406	1,392	1,262	1,267	1,226	1,535	1,759	1,687	1,575	1,641	445
Office of Personnel Management	1,241	1,210	971	793	778	791	692	772	587	553	75
Small Business Administration	774	977	1,301	1,511	1,605	1,684	1,670	1,759	1,424	1,278	294
Social Security Administration	4,554	4,727	6,923	7,065	7,617	8,328	7,972	7,053	6,747	5,322	1,740
N	26,797	31,860	40,433	42,492	45,185	48,610	44,532	42,631	35,641	32,962	6,789

Notes: Number of contracts by year/department; Upper control group and treatment group. Pre-treatment period: January 2005 to October, 2010; post-treatment periods: October 2010 to September 2015.

Table A.4: Number of contracts by agency and year - Lower control group and treatment group

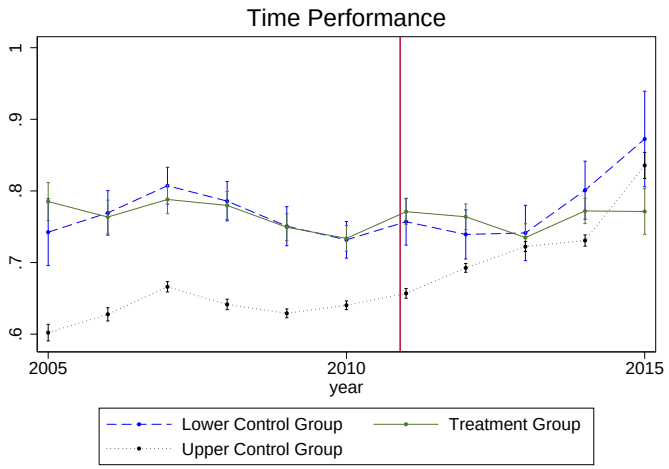
	Pre-Treatment						Post-Treatment				
	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015
Dept of Agriculture	635	361	313	217	251	346	321	283	224	236	32
Dept of Commerce	2,194	1,995	1,869	1,856	2,023	1,949	1,865	1,516	1,362	1,419	495
Dept of Defense	1,006	1,016	1,135	1,129	1,177	1,187	1,019	1,198	937	925	136
Dept of Education	51	50	62	50	63	93	44	55	42	23	4
Dept of Energy	37	35	39	36	62	55	37	57	53	51	6
Dept of Health and Human Services	575	727	950	1,571	1,631	1,392	508	177	79	43	14
Dept of Homeland Security	44	39	53	38	53	51	27	42	36	33	5
Dept of Housing	152	210	192	213	165	456	271	237	189	122	21
Dept of Interior	1,445	1,719	1,845	1,856	2,248	2,573	1,804	1,873	1,559	1,583	667
Dept of Justice	1,100	953	1,051	1,037	1,080	1,201	1,076	1,071	1,014	1,016	146
Dept of Labor	974	5,707	11,453	11,974	11,998	11,751	10,765	9,861	7,840	7,473	1,654
Dept of State	53	77	43	49	23	60	39	35	37	79	6
Dept of Treasury	221	267	222	214	246	248	201	219	181	189	51
Dept of Transportation	1,981	2,135	2,404	2,287	2,287	2,048	1,862	1,697	1,524	1,510	290
Dept of Veteran Affairs	1,853	1,554	1,495	1,480	1,356	1,264	1,480	1,438	1,200	1,310	489
Environmental Protection Agency	146	139	171	188	153	155	98	45	34	28	4
General Services Administration	2,601	2,334	2,373	2,341	2,382	2,882	2,393	2,196	1,891	1,893	411
NASA	1,264	1,643	1,858	1,965	2,113	2,254	2,666	2,516	2,493	2,950	1,021
Nuclear Regulatory Commission	198	149	148	150	159	241	212	203	166	181	37
National Science Foundation	1,361	1,424	1,328	1,402	1,509	1,905	1,770	1,701	1,621	1,701	766
Office of Personnel Management	1,357	1,182	817	565	580	521	515	406	355	310	57
Small Business Administration	465	618	1,048	972	1,040	1,135	1,132	1,034	899	753	219
Social Security Administration	4,162	4,906	6,799	6,568	7,054	6,874	6,126	5,015	4,741	4,054	1,890
N	23,875	29,240	37,668	38,158	39,653	40,641	36,231	32,875	28,477	27,882	8,421

Notes: Number of contracts by year/department; Lower control group and treatment group. Pre-treatment period: January 2005 to October, 2010; post-treatment periods: October 2010 to September 2015.

A.3 DDD - graphs for constructions

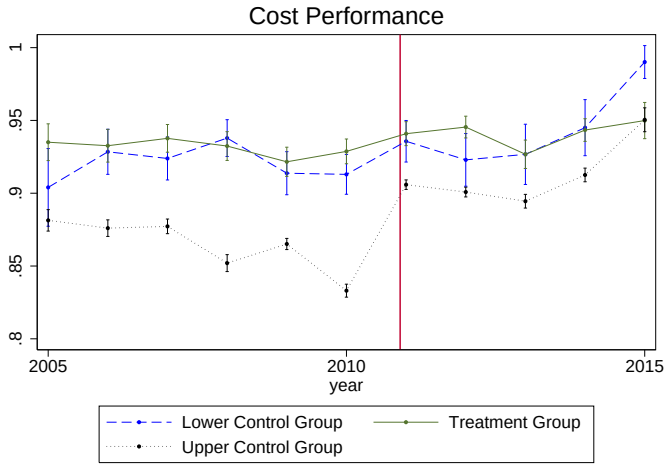
To verify whether the parallel trend assumption is also reasonable in the subsample of constructions, we plot the relative subsample yearly average time series of time performance and cost performance for the treatment and both control groups in Figures (A.6) and (A.7), respectively.

Figure A.6: Time Performance: Yearly Averages - Public Works



Notes: Trends in yearly averages of time performance for treatment, upper and lower control groups. The vertical line corresponds to October, 2010.

Figure A.7: Cost Performance: Yearly Averages - Public Works



Notes: Trends in yearly averages of cost performance for treatment, upper and lower control groups. The vertical line corresponds to October, 2010.

A.4 Agencies and Clusters

Table A.5 report results on public and private oversight treatment effects for baseline DDD models of *Time Performance* and *Cost Performance* when standard errors are clustered at various levels: no clusterized robust (columns 1-2) - used in the empirical analysis -, *contractingoffice * year* (3-4), *contractingoffice * object* (columns 5-6), *contractingoffice * year * object* (columns 7-8), *contractingoffice * year * state* (columns 9-10) and *contractingoffice * year * state * object* (columns 11-12). DDD average treatment effects result in being robust to these different clusterization levels for both public and private oversight.

Table A.5: Baseline model with different cluster level

	robust		off*year		off*obj		off*year*obj		off*year*state		off*year*state*obj	
	T	C	T	C	T	C	T	C	T	C	T	C
$\hat{\theta}_{public}$	0.0553***	0.0452***	0.0553***	0.0452***	0.0553***	0.0452***	0.0553***	0.0452***	0.0553***	0.0452***	0.0553***	0.0452***
	(0.00595)	(0.00486)	(0.0144)	(0.0142)	(0.0181)	(0.0155)	(0.0127)	(0.0106)	(0.0102)	(0.00870)	(0.00942)	(0.00793)
$\hat{\theta}_{private}$	-0.0689***	-0.0387***	-0.0689**	-0.0387*	-0.0689*	-0.0387*	-0.0689**	-0.0387**	-0.0689**	-0.0387**	-0.0689**	-0.0387**
	(0.0257)	(0.0142)	(0.0281)	(0.0218)	(0.0390)	(0.0226)	(0.0336)	(0.0191)	(0.0286)	(0.0175)	(0.0283)	(0.0169)
N	99,578	100,861	99,578	100,861	99,578	100,861	99,578	100,861	99,578	100,861	99,578	100,861
# Clusters			1052	1052	1466	1466	8511	8511	14560	14560	43098	43098

Notes: results of various DDD models of *Time Performance* - odd columns - and *Cost Performance* - even columns - on public and private oversight treatment indicators plus controls for firm size, auction type, whether the firm is a limited liability company, whether the contract was signed during the last week of the fiscal year, bureau, year, state, and contract object fixed effects. Standard errors are clustered at various levels: no clusters (columns 1-2), *contractingoffice * year* (3-4), *contractingoffice * object* (5-6), *contractingoffice * year * object* (7-8), *contractingoffice * year * state* (9-10) and *contractingoffice * year * state * object* (11-12). For each clusterization level, *# Clusters* report the relative number of clusters in the data.