

Giuffrida, Leonardo M.; Rovigatti, Gabriele

**Article — Published Version**

## Supplier selection and contract enforcement: Evidence from performance bonding

Journal of Economics & Management Strategy

**Provided in Cooperation with:**

John Wiley & Sons

*Suggested Citation:* Giuffrida, Leonardo M.; Rovigatti, Gabriele (2022) : Supplier selection and contract enforcement: Evidence from performance bonding, Journal of Economics & Management Strategy, ISSN 1530-9134, Wiley, Hoboken, NJ, Vol. 31, Iss. 4, pp. 980-1019, <https://doi.org/10.1111/jems.12492>

This Version is available at:

<https://hdl.handle.net/10419/266660>

**Standard-Nutzungsbedingungen:**

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

**Terms of use:**

*Documents in EconStor may be saved and copied for your personal and scholarly purposes.*

*You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.*

*If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.*



<http://creativecommons.org/licenses/by-nc-nd/4.0/>

# Supplier selection and contract enforcement: Evidence from performance bonding

Leonardo M. Giuffrida<sup>1</sup>  | Gabriele Rovigatti<sup>2</sup>

<sup>1</sup>ZEW Mannheim, MaCCI, and CESifo, Mannheim, Germany

<sup>2</sup>Bank of Italy, Rome, Italy

## Correspondence

Leonardo M. Giuffrida, ZEW Mannheim, MaCCI, and CESifo, Mannheim, Germany.

Email: [leonardo.giuffrida@zew.de](mailto:leonardo.giuffrida@zew.de)

## Abstract

We analyze an important but little-studied institution for balancing supply risk in the management of procurement operations: performance bonding. By adding the surety as a third party that guarantees contract fulfillment between supplier and buyer, performance bonding aims to streamline the purchasing process by influencing both contractor selection in the bidding phase and contract enforcement during project execution. Using the data on US government procurement from 2005 to 2015 and exploiting an exogenous variation in the threshold for its application to construction contracts, we find that performance bonding improves contract outcomes by 10.5% and 3.7% in terms of delays and extra costs, respectively. Net of bond premia, which by law are included in the award amounts, this effect translates into savings of about 4% in the budget for federal construction projects and 16% for mid-size projects. We provide suggestive evidence on the effectiveness of selection and monitoring by sureties as driving channels.

## 1 | INTRODUCTION

Efficient contract procurement is a complex task (Myerson, 1981). The greatest challenge that procurement managers face is supply risk. Sellers have (i) private information on their production costs, which they do not fully disclose in their bids, and (ii) incentives to behave opportunistically once the contract is awarded. Such information asymmetry between buyer and supplier paves the way for the emergence of adverse selection in the award phase and moral hazard during contract execution (Bajari & Lewis, 2014), which, in combination with intrinsic cost uncertainty in the bidding phase, can lead to amendments to the budget and planned schedule and ultimately, to efficiency losses. Both problems are amplified in the context of public procurement due to the interplay between contract rigidity and project size (Beuve et al., 2019; Fugger et al., 2019).

In this paper, we analyze the performance bonding, which is a form of insurance scheme designed to streamline the management of project operations and offset the supply risk. It involves the issuance of bonds to the buyer by private sector providers—that is, the surety companies, henceforth *sureties*—on behalf of the supplier. Through the bond, the surety guarantees that should the supplier fail to fulfill contractual duties, either the project will be completed—for example, by selecting a new supplier—or the buyer will recover the contract value. Liability in the event of inadequate contract performance gives the surety a strong incentive to reduce information asymmetry throughout the entire procurement process. To this end, sureties screen potential suppliers during the bidding phase to assess their riskiness

This is an open access article under the terms of the Creative Commons Attribution-NonCommercial-NoDerivs License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.

© 2022 The Authors. *Journal of Economics & Management Strategy* published by Wiley Periodicals LLC.

(i.e., a decreasing function of their “quality”) then charge a premium accordingly. In turn, premia are passed on to the bids, making them higher but also more informative on vendor’s characteristics, which enhances the selection of the best supplier by the buyer. During the execution phase, the sureties also bring in an additional layer of monitoring that assists the buyer in overseeing project operations. As long as the selection and monitoring performed by sureties works as expected, bonds are seldom claimed, and a better final contract performance can more than compensate for the higher awarding amounts.

Although little is known about it in the management literature, performance bonding is a tool that US procurement managers often rely on in private procurement. The institution is also traditionally adopted at all levels of national public contracting, the world biggest. Considering federal purchases only, procurers obligated \$5.6 trillion in government spending between 2005 and 2015 and 59% of contracts between 2005 and 2015 were completed behind schedule, incurred additional costs, or both.<sup>1</sup> The massive volume of goods and services procured compels contracting officers to minimize both delays and cost overruns, which are the most critical post-award metrics in both project management (Herweg & Schwarz, 2018) and empirical economics (Decarolis, 2014). Performance bonding is considered by practitioners as a powerful supply chain practice to alleviate such concerns because it brings in the benefits of the outsourcing. In fact, a critical component of procurement management is the ability to assess the suitability of potential suppliers and leverage their skills—and sureties, when involved, provide external support for both supply chain management tasks. Outsourcing offers significant organizational advantages in the management of complex projects (Handley & Angst, 2015), and is an efficient alternative for the direct provision of public goods (Adida & Bravo, 2019; Roach & Cohen, 2013). In support of this view, the literature provides a twofold argument: first, private entities can deliver public services efficiently; second, relying on external providers can help managers overcome the rigidity of public processes (e.g., in terms of information acquisition). Both effects are behind the accelerated pace of private outsourcing being implemented by governments around the world under the influence of revenue constraints and efficiency considerations.<sup>2</sup> Project management in public procurement is no exception, with contracting authorities increasingly using external providers for a wide range of services and often establishing partnerships and long-term relationships with suppliers.<sup>3</sup>

Although the effectiveness of contract performance bonds has been discussed in theoretical papers (Birulin, 2020; Calveras et al., 2004; Engel et al., 2006; Wambach & Engel, 2011), robust empirical analyses are lacking.<sup>4</sup> This paper provides the first evidence of the effectiveness of performance bonds in the context of public works contracting by estimating their benefits in reducing delays and cost overruns. We focus on US federal contracts awarded between 2005 and 2015 and take advantage of a legislative change that occurred in fiscal year (FY) 2011, which shifted up the threshold for the application of the Miller Act, the law that requires performance bonds on construction projects. Our reduced-form analysis, supported by a series of robustness checks, shows that projects without performance bonds deteriorate by 10.5% and 3.7% in terms of delays and cost overruns, respectively.<sup>5</sup>

Ideally, we would contrast the positive effects on the ex-post cost outcomes with the upfront costs that the government must bear to purchase the bonds, that is, the premia. Although we do not observe the premium, we rely on aggregate figures per bin of contract size to perform back-of-the-envelope calculations. We use the estimates of the effect of performance bonds on cost overruns and the aggregate data on average premia to build a counterfactual analysis and quantify the deadweight losses induced by the reform. The associated net savings amount to as much as 16% for mid-size contracts (i.e., worth \$100,000 to \$150,000). We find that performance bond exemption has cost US taxpayers approximately \$40 million until FY 2015. Because the FY 2011 reform in the public construction sector also meant the simultaneous introduction of more discretion and less paperwork for procuring agencies—along with the waiver of performance bonds for construction only—we add this effect to the counterfactual analysis. This additional exercise yields an “almost-zero” reform cost (i.e., gains from removing such public oversight compensate losses from waiving performance bonding). The straightforward implication of our findings is that a more effective reform should retain both the benefits of performance bonds and more discretion for public procurement managers for mid-size projects. The limitation of these back-of-the-envelope calculations is that waiving the need for performance bonds might have made agencies able to achieve lower procurement prices (or set shorter deadlines) at the tender stage. Given our lack of tender-stage information about the counterfactual award price absent the premium, we are unable to draw conclusions regarding total welfare implications of the reform.

Through a combination of descriptive evidence and more formal empirical exercises, we also provide supporting evidence for two possible channels underlying our baseline results. First, when no performance bond is required, we show a suggestive pattern of adverse selection in the pool of successful contractors, especially for limited liability firms. This finding is consistent with the judgment-proof theory—which states that less-solvent limited liability firms bid

more aggressively in the auction, so that the competitive mechanism selects undercapitalized firms adversely—and the mitigating role that surety bonds play (Calveras et al., 2004). Second, sureties are found to be effective in monitoring contractor activities during contract execution. The same firm shows dramatically worse contract performance on similar projects without performance bonds. All in all, these driving factors provide the first suggestive evidence of the role of sureties in improving both supplier selection and contract enforcement in the government procurement process.

The road map of the paper is as follows. Section 2 reviews the relevant literature. The performance bonding institution and the underlying theoretical background are discussed in Sections 3 and 4, respectively. Section 5 describes the data, while Section 6 outlines the empirical strategy and the main identification concerns. In Section 7, we present the results and discuss the channels of the empirical findings. Section 8 provides the conclusions.

## 2 | RELATED LITERATURE

This paper contributes to three strands of literature. First, this study speaks to the research thread of project management of public operations. Currently, much of the project management scholarship focuses on project success factors, and several studies have tried to gain insight into understanding projects' key success criteria (Davis, 2014; Shao et al., 2012). Nevertheless, these papers sparingly relate to the public sector. We contribute to this study field by investigating the effectiveness of a regulation strategy in improving public operations—that is, government construction projects. In this respect, our work closely relates to Calvo et al. (2019) and Carril (2021), who causally assess the effects of red tape on projects' time and cost overruns. In our paper, we consider the same setting—that is, the US federal procurement—and we leverage the same bylaw adjustment—that is, the upward shift of the Simplified Acquisition Procedures (SAPs) cutoff, which occurred at the beginning of the FY 2011. However, we instrumentally control for public oversight only to identify the impact of performance bonds on construction projects. We find that public oversight enacted by public offices negatively affects project outcomes, which is in line with their results. We complement their analysis by focusing on performance bonding as an alternative strategy to improve operations of public infrastructures. We also propose a mix of descriptive and more formal empirical exercises to test both the selection and monitoring channels separately. Other papers focusing on methods to enhance public operations deal with funding stability (Smirnoff & Hicks, 2008), optimal design of crashing policies (Goh & Hall, 2013), incentive-aligning contracts (Chen & Lee, 2017), and transparency (Bauhr et al., 2019).

This paper also contributes to the empirical literature on the regulation of public procurement markets. This scholarship 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 tender outcomes (Bajari & Lewis, 2014; Branzoli & Decarolis, 2015; Decarolis et al., 2016; Krasnokutskaya & Seim, 2011; Marion, 2007); and (ii) papers focusing on ex-post tools for enhancement of contract outcomes (Calzolari & Spagnolo, 2009; Coviello et al., 2018). Our paper combines both approaches and highlights the role of performance bonding both as a bidders' selection mechanism and as a means to increase the monitoring of contractors. This study also relates to the rising economic analysis of the impact of different designs and institutions on procurement outcomes.<sup>6</sup>

Finally, this paper speaks to the rich theoretical literature studying the implications for procurement when cost uncertainty interacts with limited liability (Board, 2007; Burguet et al., 2012; Calveras et al., 2004; Chillemi & Mezzetti, 2014; Engel et al., 2006; Waehrer, 1995; Wambach & Engel, 2011; Zheng, 2001). The main insight from this scholarship is that limited liability bidders can default with cost uncertainty because the limited liability makes bidders not fully internalize potential losses and bid more aggressively. This effect is stronger for financially weaker participants at the auction. Starting with Calveras et al. (2004) (henceforth CGH), performance bonds (often called surety bonds in a more comprehensive interpretation, e.g. including a bid bond) are presented as a potentially effective tool to reduce such inefficiencies. Recently, Birulin (2020) argues that these guarantees are effective also when cost overruns are considered. This paper provides empirical support to this long-standing discussion for the first time by identifying an overall positive impact of performance bonds across the different stages of the procurement process. In Section 4, we further discuss the theory on performance bonds while specifying the testable hypotheses.

### 3 | INSTITUTIONAL SETTING

In this section, we motivate the need for performance bonds in the construction sector and outline its scheme in the context of the United States. We also explain how performance bonds are designed to shift the issues plaguing the construction business away from the purchasing manager. We then turn to the US government procurement setting, in which the performance bonding is required to protect federal construction projects from the same systematic risk affecting private constructions, but to a larger extent. We conclude by presenting how a bylaw introduced as of FY 2011 excluded some of these projects from requiring performance bonds and how we will exploit this contingency to estimate the effect of performance bonding on contract outcomes.

#### 3.1 | Performance bond's role in constructions

In procurement management, purchasing physical goods entails strategic considerations for the buyer to ensure the maximum benefit in recognizing needs, the awarding procedure, contract writing, competition, and optimal ex-post rating of the delivery. Procuring services adds an additional layer of complexity as the level of the contractor's commitment during the execution phase is ex ante uncertain. Once the contract is signed, there is room for opportunistic behavior that might force the buyer to modify the planned schedule or budget, with subsequent delay time and cost overrun. When dealing with the procurement of construction projects (such as the building or maintenance of a road, bridge, or tunnel), buyers also face uncertainty related to completion costs and business factor dynamics. In essence, unexpected adverse shocks can hit contractors throughout the execution of the work. The above issues make gambling on contractors in construction projects a risky decision, possibly leading to profit erosion for the buyer and, ultimately, losses. In the worst-case scenario, contractors go bankrupt halfway through the job, leaving the work incomplete and the buyer with no party to claim for damages.<sup>7</sup> To avoid such lose-lose scenarios, buyers should enhance the procurement process of construction projects by better selecting the least risky sources (ex-ante), monitoring the contractors' operations, and securing tools to ensure completion, possibly within the planned deadline and budget (ex-post).

In the United States, performance bonding is a commonly-used institution in both public and private construction businesses as a tool for mitigating the above risks. It is based on the issuance of an obligation (the performance bond) by a surety—that is, a specialized insurance company—which ensures a buyer, in case the contractor is not able to fulfill the contractual obligations.<sup>8</sup> This risk-transfer scheme is used as (i) guarantee for the buyer that contractors are capable of performing the required tasks ex ante, (ii) a layer of contractors' monitoring during the execution phase, and (iii) a prompt solution if the contractor fails to execute the contract under the agreed terms. The details of US performance bonding are presented in the next subsection.

#### 3.2 | Performance bonding and the procurement process

##### 3.2.1 | Outset

When required by the buyer, the issuance of a performance bond is the potential contractor's responsibility. Sureties operate in a competitive market, and they have their own set of underwriting rules and prices; hence, firms competing for a construction project can consult different sureties and select the one that agrees to issue the performance bond. To guarantee that the potential contractor is capable of fulfilling the contract, sureties run prequalification checks based on general assessment of firms' experience, credit history, and finances.<sup>9</sup> Sureties, thanks to their access to firms' detailed information—including soft information—can usually evaluate the construction firm's ability to fulfill the contract provisions better than the buyer.<sup>10</sup> US sureties have evaluated constructors' performance for more than a century, and their expertise, experience, and objectivity in pre-qualifying contractors are the performance bond's most valuable attribute.<sup>11</sup> Being liable in case of contract mismanagement, the surety will back up only those potential contractors that can demonstrate having the prospective capacity to perform a specific project at a specific point in time.

Upon completion of the prequalification process, the surety provides the potential contractor with the performance bond premium, which amounts to an actuarially-based fee that typically ranges from 0.5% to 3% of



the contract value. The premium varies depending on characteristics such as the size, type, and duration of the project, plus, notably, an idiosyncratic component that reflects how the characteristics of the contractor that emerged from the selection process match the contract needs.<sup>12</sup> A small fixed fee for the screening service is normally also included as a price of the bond. The fixed fee for the prequalification is to be paid to the surety always, irrespective of the job award.<sup>13</sup> Importantly, the bond is issued *before* the contract award but becomes effective *only* in case the buyer finally selects the bonded supplier.

### 3.2.2 | Execution

In addition to being scrutinized, the contractor undergoes a monitoring process in case it is awarded the contract. Sureties use accurate accounting systems to gather and analyze information regarding bonded contractors systematically.<sup>14</sup> Depending on the time elapsed since the end of the last FY statement, the surety may ask for an interim financial statement to periodically show how the current year is progressing. It also has the legal right to access information on payments made by the contractor. The latter has to prepare a schedule for the work in progress, listing each job and including, among others, the total contract price, the approved change orders, the amount billed till date, the costs incurred till date, a revised estimate of the cost to complete, the estimated final gross profit, and the anticipated completion date.

### 3.2.3 | Default

Despite the technical and financial assistance provided by the performance bonding scheme, contractor default is sometimes an unavoidable circumstance in business. In the event of bankruptcy, the buyer must formally declare the contractor in default. The surety conducts an impartial investigation before settling any claim to protect the contractor's legal recourse if the buyer improperly declares default. The surety's options in case of proper default, instead, must be spelled out in the bond. These options may include (i) the right to re-bid the job for completion, (ii) bring in a replacement contractor, (iii) providing financial and technical assistance to the existing contractor, or (iv) pay the penal sum of the bond (e.g., the original contract amount). The fact that buyers have been shielded from risk is stressed by the insurance amounts that sureties have paid over time: more than \$10 billion due to contractor default since 1995, according to The Surety & Fidelity Association of America.<sup>15</sup>

## 3.3 | Performance bonds in US federal constructions

### 3.3.1 | The Miller Act (40 USC Chapter 31)

In private sector projects, performance bonds are requested by the buyers under several factors that typically include, but are not limited to, the signed contract amount. On the other hand, public works follow more rigid rules that leave less room for discretion and are designed to minimize the risks for the collective welfare. The set of rules governing the US federal procurement process are gathered in the Federal Acquisition Regulation (FAR). Public construction projects are on average bigger and involve a huge chunk of the government budget. Hence, cost overruns as well as delays in lead time and missed completion are at the taxpayers' expense. The US government requires the issuance of performance bonds whenever the contracting officers procure works from the private market above a pre-specified value threshold.

Performance bonding was initially introduced in 1894 through the Heard Act and later replaced by the Miller Act in 1935. The Act is now included in the 40 US Code Chapter 31, Subchapter III. Its implementation is required in federal procurement by FAR 28.102 and compels firms bidding on federal construction projects to be backed up by a performance bond, which becomes binding for the winning firm only once the contract is awarded. The Department of Treasury lists the sureties that are qualified to back up federal construction projects.<sup>16</sup> FAR 28.102-2 establishes an amount of 100% of the contract amount as a rule to determine the size of the performance bond, implying full insurance of the awarding contract value.<sup>17</sup> Importantly for us, FAR 28.106-1 requires potential sellers to fill out and present, along with their bid, a specific form detailing the terms of the bond agreement with the surety. Among the

many other pieces of information, the form reports the penal sum of the bond and its premium (both the rate in \$ thousand and the \$ total amount). Unfortunately, we cannot access these documents, but we safely assume that the contracting officers handling the call for bids requiring performance bonds observe these pieces of information for all potential contractors. Figures D2 and D3 report the blank module contractors need to fill out and convey to the federal buyer.

Once the performance bond is activated, its coverage remains in force until the end of the contract or when it is called upon by the contracting authority. Before modifying any contractual term, the procurement officer and contractor shall obtain the consent of the surety, which typically gathers information on the contractor's conduct, project delivery, and failure perspectives before consenting, as specified in the previous subsection. To discuss the actual implementation of performance bonding, we need to present the concept of SAP.

### 3.3.2 | SAPs and threshold

SAPs (FAR Part 13) are a set of streamlined procedures designed to reduce the administrative burden only for federal procurement projects whose estimated value lies below the Simplified Acquisition Threshold (SAT). Public projects in the United States are typically subject to formal checks conducted by the agency to lower the risks of their opportunistic behavior but increase, sometimes dramatically, the amount of red tape. These sets of checks are collectively labeled as “public oversight” by Calvo et al. (2019), and we follow the same terminology. The SAPs exempt federal contractors from public oversight and, in the case of public construction projects, from performance bonding as well.<sup>18</sup> The next section provides a thorough discussion on public oversight and its exemption implications.

### 3.3.3 | The FY 2011 SAT bylaw

In this paper, we exploit a bylaw of the SAT to inform our identification strategy. Indeed, Law 41 USC 1908 requires the government to review acquisition-related thresholds every five FYs for inflation. Notably, “to review” does not necessarily imply “to adjust”: the choice to move thresholds depends on several factors other than the change of the Consumer Price Index in the previous five FYs, including political and economic considerations. As a result, the SAT was raised from \$100,000 to \$150,000 as of October 1, 2010.<sup>19</sup> Figure 1 provides a stylized timeline of the outlined framework. The left panel represents the FY 2005–2010 for construction and all other government-procured projects (“non-constructions” from now on), right panel refers to the 2011–2015 period for the same projects.<sup>20</sup> The horizontal dotted line represents the SAT, moving upward as of FY 2011; the grid identifies awarded projects worth \$100,000 to \$149,999, while the background colors refer to oversight application (gray, dark, and pale) or exemption (white). In the case of construction projects, the exemption includes both public oversight and performance bonding, while for non-constructions, the exemption was for public oversight

	Pre-2011		Post-2011	
	Construction	Non-Construction	Construction	Non-Construction
Above \$150k	Public Oversight and Performance Bonding	Public Oversight	Public Oversight and Performance Bonding	Public Oversight
\$100-150k	Public Oversight and Performance Bonding	Public Oversight	None	None
Below \$100k	None	None	None	None

**FIGURE 1** Reform timing. Projects subject to (in gray) and exempted from (in white) oversight (i.e., performance bonds for constructions, public oversight for all) before and after October 2010. The \$100,000–150,000 class (grid) identifies the treatment group, that is, those projects subject to oversight before but not after the reform. The upper control group (i.e., “Above \$150k”) includes projects that are always exposed to oversight (i.e., always gray, both construction and nonconstruction) while the lower control group (i.e., “Below \$100k”) consists of projects that are never exposed (always white)

only. Over the period considered, construction projects valued above \$150,000 (below \$100,000) are always (never) subject to performance bonding and public oversight, while nonconstruction projects of the same amount are subject to (exempted from) public oversight only.

## 4 | THEORETICAL BACKGROUND

In this section, we set out the theoretical foundations of the impact of performance bonds in the procurement process. We start by acknowledging the existing contribution of Calvo et al. (2019)—corroborated recently by Carril (2021)—that the US federal procurement process is subject to public oversight above the SAT. While helping to restrain vendors' misconduct, public oversight introduces a burden in terms of both time and cost due to its intrinsic characteristics. In fact, to comply with solicitation rules and produce the required paperwork, contractors must divert resources away from contract-specific tasks; the need for public approval hampers contracting officers' leeway. This feature is extensively recognized by the FAR itself when introducing the SAPs.<sup>21</sup> On top of public oversight, firms competing for federal construction projects are *also* required by the Miller Act to obtain performance bonds, which entails an additional layer of supervision exerted by the sureties in the government procurement process.

Due to the nature of performance bonds, two channels may drive their impact on procurement contract performance. Specifically,

- Hypothesis *H1*: Firms subject to performance bonding provisions are structurally different due to the selection effect induced by sureties—*selection effect*;
- Hypothesis *H2*: Being covered and monitored by a surety gives firms more incentives to complete projects under the original terms and conditions—*monitoring effect*.

Note that, net of costs of bonds, both channels lead to expected improvements in project outcomes, which also implies conflicting predictions on the pool of winning firms. On the one hand, according to *H1*—a channel originally proposed by CGH—we should observe a different pool of winning firms, with and without performance bonding. In particular, when contractors are not subject to the pre-bidding screening process enacted by the sureties, we should observe a higher share of low-quality winners.<sup>22</sup> The latter are more likely to win at the expense of “good types” with no performance bonding, since their lower suitability does not translate into higher premia, which are known to the contracting officer. Hence, following the CGH model, moving from a performance bonding regime to one without ex-ante selection causes the pool of winning firms to include more financially unsound firms. As a result, the average quality and projects' performance deteriorates (*Selection Effect*).

On the other hand, *H2* implies that performance bonding improves contract outcomes by tackling the room for moral hazard during the operational phase of a project without affecting the ex-ante selection of contractors. According to it, winning firms exert higher effort and are incentivized to perform better when monitored by the sureties.<sup>23</sup> In the absence of performance bonding, an issue of opportunism arises, and contractors tend to perform worse. Indeed, according to *H2*, the pool of bidders does not necessarily change with or without selection. In line with *H1*, and conditional on *H2*, under the performance-bonding regime, the contractors, *ceteris paribus*, are expected to be less likely to renegotiate contract terms. In other words, when moving from a performance-bonding regime to one without ex-post monitoring, the pool of winning firms does not necessarily change, but the average quality and performance of contractors decrease (*Monitoring Effect*).

*H1* and *H2* are not competing—that is, the gross effect of performance bonding amounts to a combination of the two channels—and we expect both to be relevant in the public procurement market. The role designed by the law for sureties is meant to maximize both effects, but it is twofold: on the one hand, the pre-selection reduces the adverse selection and, on the other hand, monitoring lowers the moral hazard without significant countervailing effects.<sup>24</sup> The goal of our empirical strategy is to assess whether the two channels jointly contribute to the gross effect of performance bonding.



## 5 | DATA

In this section, we present our source of data and its content and detail the data management and sample selection process that leads us to the working sample of procurement contracts.

### 5.1 | FPDS data set

The data we use is sourced from the Federal Procurement Data System (FPDS), a database to which federal contracting officers in the United States must submit complete reports on procurement actions, as required by the FAR.<sup>25</sup> It contains virtually all contracts as well as every following activity, both supply- and service-based, that have been awarded by the government and, for the period under analysis, exceed an individual transaction value of \$2,500.<sup>26</sup> The data set also includes several variables related to the transaction itself, including agency and contractor characteristics, in addition to solicitation and project information, such as the signature, award, and insertion dates, and the procurement category (i.e., service or supply and their sub-classifications). From the latter information, we also define *Object* as the procurement category.<sup>27</sup>

Importantly for this study, we observe the type of solicitation procedures used, which reveals whether a contract is awarded through SAPs (i.e., no performance bonding) or other awarding procedures (e.g., sealed bidding or negotiation). Using this information, we build the binary variable *SAP*, indicating whether the contract is waived from public oversight (and performance bonds for constructions) or neither. The information provided by *SAP* crucially supports our identification strategy. Ideally, we would want to observe the engineers' estimated value (i.e.,  $EV_i$ ), which is the information used by the contracting officer to assign the performance bond treatment to a contract. However, this information is not reported to the FPDS. Nonetheless, we can overcome the issue by combining the information provided by (i) *SAP*, which points out projects with  $EV_i$  below the SAT, with (ii) the award contract value.<sup>28</sup> The version of the FPDS employed dates back to November 30, 2021.

### 5.2 | Data management

We start by splitting the raw data (i.e., the whole universe of transactions between the federal government and private suppliers) into two main groups: base contracts records and amendment records. The former refers to the first transactions between procurement officers and vendors and correspond to our unit of observation whose reported characteristics represent the baseline agreement information (i.e., the contract). The latter accounts for all the revisions, modifications, or corrections to existing contracts. Each contract is identified through a unique ID (i.e., the "PIID"), which is also used to mark all its future alterations. Therefore, we can track the entire history of each contract and link each contract to its revisions (if any). Amendment records are classified according to the reason for contract modification, which is reported alongside the extra cost and the 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.<sup>29</sup> We use the *in-scope* amendments to build the outcome measures of our empirical analysis presented below.<sup>30</sup>

#### 5.2.1 | Performance indexes

First, we define (i) *Time Overrun*, representing the days in excess of a project's initial lead time, measured as the difference 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 all renegotiated amounts. *Time Overrun* and *Cost Overrun* are widely used proxies for contractual performance in both project management and economics; 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 (e.g., material cost fluctuations, adverse unexpected natural conditions). Given that high-value contracts are the minority in our sample, 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 to only adequately reflect

the contract performance.<sup>31</sup> In Appendix A, we show the robustness of our results to the inclusion of out-of-scope amendments.

To compare the two overrun measures with the initial expected outcomes—that is, the lead time and cost of completion reported in the original contract—we specify two convex indexes for contract performance, such as

$$\text{performance}_i^{\text{type}} = \frac{\text{expected outcome}_i^{\text{type}}}{\text{expected outcome}_i^{\text{type}} + \text{overrun}_i^{\text{type}}},$$

where  $i$  refers to the contract and  $\text{type} = [\text{time}, \text{cost}]$ . By design, such a metric maps the couple  $[\text{expected outcome}_i^{\text{type}}; \text{overrun}_i^{\text{type}}]$  to the interval  $(0, 1]$ , with an increasing performance approaching 1, which is in the case of no overruns.<sup>32</sup> Not surprisingly, the two performance measures are positively correlated (i.e.,  $\approx 50\%$ ).<sup>33</sup> The FPDS data set includes a number of other variables from which we build the controls in our regressions. *Constr* is an indicator for construction projects (i.e.,  $\text{objects} \in [Y1, Z1, Z2]$ ); *Fixed Price* indicates whether contracts are priced with a fixed price or cost-plus format, that is, if the supplier is paid a fixed amount regardless of costs incurred, or if the supplier is entitled to obtain compensation in proportion to its costs plus a mark-up; *Small* is an indicator for a small-business vendor;<sup>34</sup> *Negotiation* is a dummy variable for projects awarded through negotiated procedures; and *Bureau Size*, which is the cumulative value of contracts an agency has awarded in the current FY for the same service or construction category.

### 5.3 | Sample selection and descriptives

We restrict our sample to those projects awarded through competitive solicitations as the treatment effect would otherwise not be observable.<sup>35</sup> For similar reasons, we focus on projects whose tasks are such that the vendor can influence the outcome metrics through effort. Supply contracts usually do not allow for renegotiations: hence, for these projects, our measure of performance does not proxy outcome quality whatsoever, and we exclude them from the analysis.<sup>36</sup> The same rationale applies to the service subcategory “Lease or Rental of Equipment, Structures, or Facilities.”<sup>37</sup> To keep a balanced time-window around the SAT update, we rule out observations before January 1, 2005, and cover the FYs 2005 to 2015. We also drop projects related to certain commercial items that make use of SAPs for the acquisition of services for amounts greater than the oversight threshold. This cleaning process yields a sample of 247,353 contracts and 33,731 unique firms.<sup>38</sup>

Two sets of projects—the two solid-colored segments 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*). The reform date and the two treatments cluster the sample into six distinct groups: the treatment group, counting all projects—constructions included—valued between \$100,000 and \$149,999, which are also subject to public oversight before but not after the reform; upper and lower control groups, consisting of all projects valued more than \$150,000 or less than \$100,000, respectively; and construction treatment, as well as upper and lower control subgroups, including construction projects only, which are subject to performance bonding as well public oversight and have the same monetary cutoffs.

In Table 1, we report summary statistics of key features in our working sample conditioning on the upper and lower control group and the treatment group. Furthermore, within each subsample, we condition on pre- or after-reform. This descriptive analysis hinging on sample split allows us to gauge both cross-group time-specific (i.e., vertical) and within-group over-time (i.e., horizontal) differences. First, a vertical comparison shows how average and median contract value and duration mechanically depend on the group by design (e.g., the upper control group contracts are bigger and longer than the treatment group’s). On average, contract performances are similar in the lower and treatment group and better than the upper group. Upper control group variables are noisier because of the more heterogeneous composition of contracts in terms of size (i.e., from \$150k up to \$ tens of millions). Fixed-price pricing is more frequent the smaller the contract; it is consistently lower in the upper control group. A horizontal comparison displays how the size and duration of contracts in the upper group diminish after the reform, and the standard deviation decreases dramatically. This is likely due to the mechanical exclusion of the biggest contracts in the latest years of our sample, as we condition the contract on being completed by the download date. On average, performance indexes are quite stable over time in any group.

TABLE 1 Summary statistics

<b>(a) Upper control group</b>								
	<b>Before</b>				<b>After</b>			
	<b>Mean</b>	<b>SD</b>	<b>Median</b>	<b>N</b>	<b>Mean</b>	<b>SD</b>	<b>Median</b>	<b>N</b>
Expected cost	2000.5	9988.1	420.5	42,117	1357.0	6490.5	320.3	56,687
Expected time	452.2	2609.8	364	42,117	403.5	481.7	364	56,687
Total cost	3459.7	18,638.6	902.8	42,117	2390.7	17,292.9	581.7	56,687
Total time	838.3	881.5	561	42,117	659.6	650.6	398	56,687
Time Performance	0.7	0.3	0.75	42,117	0.7	0.3	1	56,687
Cost Performance	0.7	0.3	0.91	42,117	0.8	0.3	1	56,687
Fixed price	0.7	0.4	1	42,117	0.8	0.4	1	56,687
Number of offers	5.5	19.7	2	41,870	5.6	15.0	2	43,147
<b>(b) Lower control group</b>								
	<b>Before</b>				<b>After</b>			
	<b>Mean</b>	<b>SD</b>	<b>Median</b>	<b>N</b>	<b>Mean</b>	<b>SD</b>	<b>Median</b>	<b>N</b>
Expected cost	71.9	15.1	70.0	30,385	86.6	28.0	80	39,611
Expected time	284.7	1004.7	152	30,385	313.3	414.6	274	39,611
Total cost	174.9	2186.2	96.4	30,385	243.7	5476.6	117.1	39,611
Total time	455.0	1095.8	238	30,385	492.1	555.7	364	39,611
Time Performance	0.8	0.3	1	30,385	0.8	0.3	1	39,611
Cost Performance	0.9	0.3	1	30,385	0.9	0.3	1	39,611
Fixed price	0.9	0.3	1	30,385	0.9	0.3	1	39,611
Number of offers	4.1	13.5	2	30,236	4.4	11.8	2	31,665
<b>(c) Treatment group</b>								
	<b>Before</b>				<b>After</b>			
	<b>Mean</b>	<b>SD</b>	<b>Median</b>	<b>N</b>	<b>Mean</b>	<b>SD</b>	<b>Median</b>	<b>N</b>
Expected cost	122.0	15.1	121.1	11,508	122.8	15.8	121.3	2722
Expected time	282.0	415.8	192.5	11,508	387.6	386.6	364	2722
Total cost	317.8	472.7	166.9	11,508	193.3	219.7	149.4	2722
Total time	533.3	639.6	355	11,508	507.0	482.4	365	2722
Time Performance	0.8	0.3	1	11,508	0.9	0.3	1	2722
Cost Performance	0.8	0.3	1	11,508	0.9	0.2	1	2722
Fixed price	0.9	0.3	1	11,508	0.9	0.2	1	2722
Number of offers	4.9	15.3	3	11,485	2.4	5.4	1	1587

Note: Descriptive statistics for the control groups (both *upper* and *lower*, above panel) and the treatment group (below panel), before and after the bylaw are reported. *Time Performance* and *Cost Performance* are relative measures of performance—bounded 0 to 1; *Fixed Price* indicates whether the pricing type of the contract is fixed-price (vs. a cost-plus type).

## 6 | EMPIRICAL STRATEGY

In this section, we present the methodology we use to estimate the causal effect of performance bonds on the different contract outcomes. Further, we provide a discussion of the main potential threats to our empirical approach.

## 6.1 | Methodology

We shall exploit the threshold adjustment to separately identify the effect of public oversight and performance bonding exemption. In principle, we would want to randomly assign the performance bonding 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 nonexperimental methods that mimic it under reasonable conditions.

In our setting, construction projects above the SAT are exposed to both public oversight and performance bonding, while nonconstruction projects are subject only to public oversight. Two treatments are thus nested one onto the other. Construction projects above the SAT are exposed to both public oversight and performance bonding, while nonconstruction projects are subject only to public oversight. The grid in Figure 1 identifies the two treatments: for construction projects, the treatment results in the exemption from both types of oversight, while for nonconstruction, it reduces to exemption from public oversight only. Segments without grids identify upper (gray) and lower (white) control groups. Hence, the very definition of treatment is ambiguous, as it encompasses the effect of the waiver of both public oversight and performance bonding. The latter is relevant to treated construction projects only, but its effect is estimated jointly on the whole sample and cannot be disentangled via a plain difference-in-differences (DD) estimator.<sup>39</sup>

To deal with two nested treatments, we rely on a difference-in-difference-in-differences (DDD) approach, which nests two DD models in a single equation and, while controlling for the relative differences between treatment and control groups, isolates the two intertwined average treatment effects under testable assumptions.<sup>40</sup> Specifically, we estimate the following linear model:

$$\begin{aligned}
 Y_{it} = & \alpha + \theta_{public}(Waiver_{it} \times Post_{it}) + \theta_{pb}(Waiver_{it} \times Post_{it} \times Constr_i) \\
 & + \beta_1 Waiver_{it} + \beta_2 Constr_i + \beta_3(Constr_i \times Waiver_{it}) + \beta_4(Constr_i \times Post_{it}) \\
 & + \gamma X_{it} + \delta_t + \varepsilon_{it},
 \end{aligned} \tag{1}$$

where  $i$  refers to the contract and  $t$  stands for the FY.  $Y_{it}$  represents the contract outcomes (i.e., *Time Performance* or *Cost Performance*).  $Waiver_{it}$  indicates the treatment group and captures the differences across the treatment and control groups before the policy change.<sup>41</sup>

$Post_{it}$  is a dummy variable for projects awarded after the reform (i.e., FY 2011–2015) and captures aggregate factors affecting in  $Y_{it}$  even in the absence of a policy change. The interaction term  $Waiver_{it} \times Post_{it}$  captures the outcome trend in the treatment group for constructions and nonconstructions jointly. The triple interaction term  $Waiver_{it} \times Post_{it} \times Constr_i$  indicates the construction projects subject to performance bonding exemption. The coefficients of interest are  $\theta_{public}$ , representing the average treatment effect of the exemption from public oversight, and  $\theta_{pb}$ , capturing the effect of the exemption from performance bonding.<sup>42</sup>  $X_{it}$  stands for the list of covariates presented in Section 5.2 (i.e., small dummy, fixed-price dummy, negotiation dummy, office size) that might correlate with our treatments. For instance, according to the selection channel, small firms might be encouraged to participate in treated tenders when performance bonds are waived. Selection of the awarding procedure also depends on project size according to FAR Parts 13 to 15. Moreover, the tendency of employing a fixed-price contract format might be procurement-category-specific and correlate with treatment statuses. A similar argument applies to the time-varying bureau budget, which might drive contract splitting into lots. We treat our data as a pooled cross-section and use *upper control group*—the set of construction projects always subject to both public oversight and performance bonding and that of nonconstruction projects always subject to public oversight—in the baseline and main robustness specifications. We are agnostic about which of the two available control groups is the best candidate. In Appendix A, we show the robustness of our results to the use of the *below control group* contracts.

## 6.2 | Identification concerns

The correctness of our identification strategy hinges on the assumption that, absent the treatment, the control and treatment group trends would have hold parallel. Although such assumption cannot be formally tested, we are able to test the presence of pre-trends in the outcome variables. However, being the DDD a difference between two nested DDs makes the interpretation of the parallel trends somewhat different: The identification of  $\theta_{pb}$  requires no trend in the *gap*

between construction and nonconstruction contracts, absent the treatment.<sup>43</sup> In Figure 2, we propose a graphical representation of both pre-trend tests in a simplified setting. In panel (a), we report the test of no pre-trends in the difference between the treated (purple) and the control group (blue) among nonconstruction contracts—in other words, it is enough to show that  $\delta = -0.5$  does not change over time before the treatment, as it would ensure the correct identification of  $\theta = 1$  with a DD estimator. Things become more involved when considering the identification of  $\beta = 0.5$  through the analysis of construction contracts. In this case, the relevant statistics is the gap between the treated (green line) and the control group (red line) difference among construction contracts (i.e.,  $\gamma = -0.3$ ) and the same difference among nonconstruction ( $\delta$ )—in other words, we must show that  $\delta - \gamma = -0.5 + 0.3 = -0.2$  does not display any relevant trend before the treatment.

In the spirit of Autor (2003), we jointly validate the above assumptions by estimating the fixed-effects version of Equation (1) (i.e., leads and lags model):

$$y_{it} = \delta_t + \sum_{j=-m}^q \theta_{public,j} Waiver_{it+j} + Constr_t + \sum_{j=-m}^q \theta_{pb,j} Waiver_{it+j} \times Constr_t + X_t + \epsilon_{it}. \tag{2}$$

The specification includes  $m = 6$  leads and  $q = 4$  lags. A graphical representation of the results is reported in Figure 3:  $t = -1$  is the base year and  $t^*$  is FY 2011. Results confirm that there is no differential pre-trend between public oversight (solid red line) and performance bonding (solid blue line) effects—for both *Time Performance* and *Cost Performance*—before FY 2011, as  $\Delta_{\delta}^{type}$  is not statistically different from zero for any  $t < 0$ . When it comes to  $t > 0$ , the dynamics of  $\theta_t^{type}$  is already informative on the single  $\theta_{type}$ 's that we aim to estimate in Equation (1). Specifically, relaxing the single pre-post structure of the DDD displays how performance bonding and public oversight effects develop after the bylaw. In the absence of performance bonds, the treated contracts experience increasingly worsening outcomes; conversely, in the absence of public oversight, contracts increasingly improve their outcomes. The visual evidence highlights how these patterns apply in particular for *Cost Performance*. The diminishing (increasing) effect over time is interesting by itself; however, for the ease of presentation and the sake of clarity, and to ensure a direct interpretation of the heterogeneity analyses and the robustness checks that we propose below, in the rest of the paper, we will show one summary DDD estimate per treatment.

A very convenient feature of our data is that we can run the model on two equally valid sets of control groups, and switching from one to the other, as long as the parallel trend assumption holds, should not alter the DDD estimates. In fact, as shown in Appendix A, our results are robust to the choice of either groups or a narrower partition of them. In Figure 3, we propose compelling visual evidence for the development of our estimates using a set of more granular control groups (the gray-shaded lines) alternatively. Specifically, we group *all* control contracts—separately for construction and nonconstruction sub-treatment groups and for both time- and cost-dimension outcomes—in the

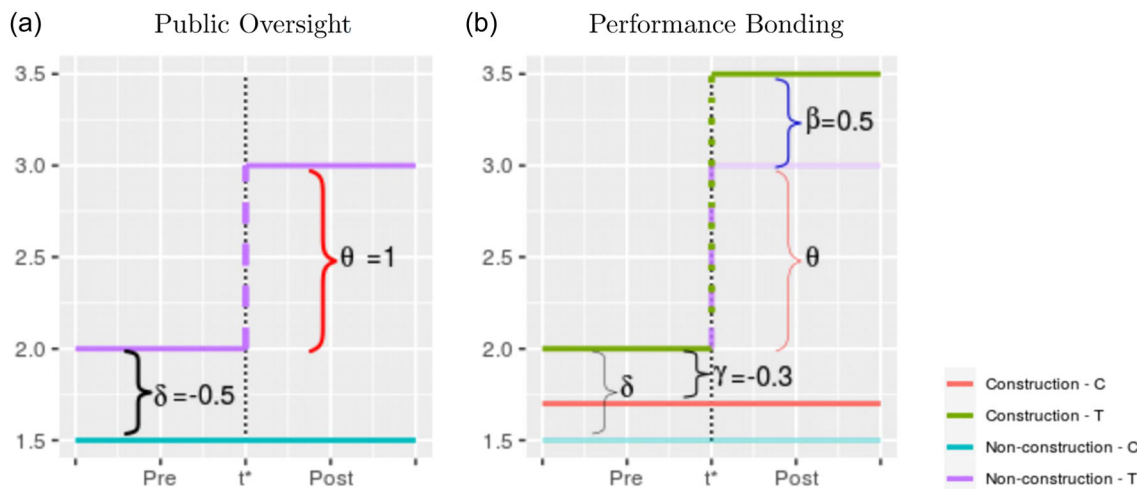
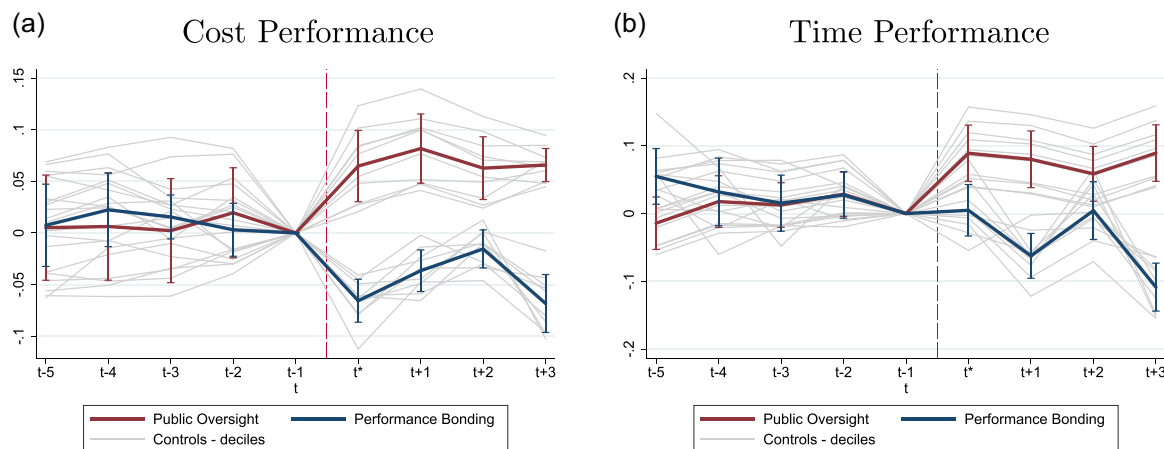


FIGURE 2 Pre-trends in DDD: Graphical example. Graphical representation of pre-trend tests for DDD. In (a), we report the case for public oversight with a plain DD pre-trend test; in (b), we provide the elements needed to ensure the correct identification of the DDD parameter. DD, difference-in-differences; DDD, difference-in-differences. [Color figure can be viewed at [wileyonlinelibrary.com](http://wileyonlinelibrary.com)]





**FIGURE 3** Leads and lags analysis with decile-based control grouping for Cost Performance (a) and Time Performance (b). Visual representation of the coefficients of the leads and lags DDD analysis for public oversight (red line) and performance bonding (blue line). Gray lines represent control groups obtained using deciles of the distribution of contract values. We plot the 95% confidence intervals of the estimated coefficients. [Color figure can be viewed at [wileyonlinelibrary.com](http://wileyonlinelibrary.com)]

deciles of their respective award value distribution, and estimate the Equation (2) for each subsample. The figure, on the one hand, confirms the robustness of the baseline analysis (i.e., solid lines) irrespective of the definition of the control group. On the other hand, the lack of pre-reform jumps in any subsample also speaks against the possible effect of selection and sorting after controlling for observables, which might play a stronger role closer to the borders of the treatment and control regions.<sup>44</sup>

The above exercise confirms the formal suitability of the DDD in our setting. There is, however, a remaining concern on the correctness of our identification strategy: the identification of the performance bonding parameter, in fact, is obtained by implicitly assuming that the public oversight effect is the same for nonconstruction contracts (which allows for its estimation) and construction contracts. This is not necessarily the case: if the public oversight effect differs in construction from all the other sectors in our data, we would obtain biased estimates of  $\theta_{pb}$ .<sup>45</sup> To address this issue, in Appendix C, we identify a subset of construction contracts that are subject to public oversight, but potentially not to the Miller Act provisions above the SAT—namely, federal projects executed abroad. A plain DD analysis on those contracts yields a reliable estimate of the public oversight effect in construction, which is in line not only with previous literature but also with our baseline model results.

Another possible issue in our empirical framework stems from the fact that that, as already mentioned in Section 5.1, we do not explicitly observe  $EV_i$ . Since we rely on a combination of *SAP* projects with their award value to track contract treatment/control status, we cannot identify the lower bound of treated projects: this exposes our treatment group sample to the risk of spurious contamination. To alleviate such identification threats, we test the robustness of our results via a sanitary check around the SAT in Appendix A, and show that the award amount is a good proxy for  $EV_i$  and that the misclassification of projects to the treatment group is residual (Carril, 2021).

It is also 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 exerts an ex-ante selection on potential contractors, affecting the pool of winners on the quality dimension.

## 7 | RESULTS

This section describes our empirical findings. We start by showing the estimates for the gross effect of performance bonding; then, we present evidence suggesting that both selection and opportunism drive the effect of bonds. We then switch to the back-of-the-envelope calculations of the net effect of performance bonds and the cost of the SAT bylaw reform. Please refer to the Appendices A and B for a rich analysis of the robustness of our results to the sample definition, the outcome variable, and against the possibility that the estimates capture differential shocks in the construction industry.

## 7.1 | Gross effect

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 (1). For the sake of clarity, we only report  $\hat{\theta}_{pb}$  and  $\hat{\theta}_{public}$ .<sup>46</sup> Column 1 reports the estimates of Equation (1)'s parameters, without the additional controls, while specifications 2 to 5 enrich the model by iteratively including the controls listed in Section 6.1 plus an increasing number of fixed effects (agency, state, and object). The addition of agency-fixed effects (column 3) seems to solve an omitted variable bias issue in the plain model case, as both treatment effect parameters roughly double in absolute magnitude with virtually

TABLE 2 Baseline results

	(a) Time Performance				
	(1)	(2)	(3)	(4)	(5)
<i>Performance bonding</i>					
$\hat{\theta}_{pb}$	−0.049** (0.023)	−0.046** (0.023)	−0.090*** (0.022)	−0.089*** (0.022)	−0.088*** (0.023)
<i>Public oversight</i>					
$\hat{\theta}_{public}$	0.018 (0.012)	0.029** (0.011)	0.045*** (0.010)	0.044*** (0.010)	0.045*** (0.010)
<i>N</i>	120,175	120,175	120,175	120,175	120,175
<i>R</i> <sup>2</sup>	0.038	0.050	0.109	0.115	0.128
<i>Avg</i> <sub>services</sub>	0.808	0.808	0.808	0.808	0.808
<i>Avg</i> <sub>works</sub>	0.837	0.837	0.837	0.837	0.837
	(b) Cost Performance				
	(1)	(2)	(3)	(4)	(5)
<i>Performance bonding</i>					
$\hat{\theta}_{pb}$	−0.022 (0.014)	−0.012 (0.013)	−0.032** (0.013)	−0.033** (0.013)	−0.035*** (0.013)
<i>Public oversight</i>					
$\hat{\theta}_{public}$	0.019** (0.010)	0.014 (0.009)	0.035*** (0.008)	0.034*** (0.008)	0.040*** (0.008)
<i>N</i>	120,175	120,175	120,175	120,175	120,175
<i>R</i> <sup>2</sup>	0.066	0.142	0.198	0.203	0.218
<i>Avg</i> <sub>services</sub>	0.868	0.868	0.868	0.868	0.868
<i>Avg</i> <sub>works</sub>	0.945	0.945	0.945	0.945	0.945
Fiscal year fixed effects	✓	✓	✓	✓	✓
Controls		✓	✓	✓	✓
Bureau fixed effects			✓	✓	✓
State fixed effects				✓	✓
Object fixed effects					✓

Note: Results of the DDD regression of *Time Performance* (a) and *Cost Performance* (b) on the performance bonding indicator ( $\hat{\theta}_{pb}$ ) and the public oversight dummy ( $\hat{\theta}_{public}$ ). Column 1 reports the results of a plain DDD regression, Column 2 adds controls (*small*, *negotiation*, *fixed price* and *bureau size*). Columns 3–5 include an increasing number of fixed effects (bureau, state, and contract object). In each panel, *Avg*<sub>services</sub> and *Avg*<sub>works</sub> account for the average outcome in the Services and the Public Works treatment group, respectively. Standard errors are two-way clustered at the object and FY level.

\*\**p* < .05; \*\*\**p* < .01.

unchanged standard errors, and model fit increases dramatically. This highlights the crucial role of contracting agencies in dealing with contractors directly and through the sureties.<sup>47</sup> On the other hand, adding fixed effects for the state (Column 4) or object (Column 5, our baseline model) does not appear to alter the estimates substantially. Concerning the latter, much of its effects pass through the contract being construction or non-constructions, which we already control for by research design. We two-way cluster standard errors at the object and FY level. We refer the reader to Appendix C for a discussion of our clustering approach.

To ease the interpretation of the estimated coefficients—that is, in terms of relative impact—we report the mean of each outcome in the treatment groups (i.e.,  $Avg_{works}$  or  $Avg_{services}$ ) at the bottom of each panel. The baseline estimates show that waiving public oversight positively affects contract performance, whereas excluding performance bonds offsets such gains. More specifically, waiving from performance bonding worsen outcomes in terms of both *Time Performance* (10.5%, or  $-0.088/0.837$ ) and *Cost Performance* (3.7%, or  $-0.035/0.945$ ). Instead, waiving public oversight positively impacts contract outcomes, both in terms of *Time Performance* ( $0.045/0.808 \approx +5.6\%$ ) and *Cost Performance* ( $0.04/0.865 \approx +4.6\%$ ); such results are in line with those presented by Calvo et al. (2019) and Carril (2021) on similar data, but with rather different identification strategies. The composite effect of the reform on constructions (i.e., the impact of the joint waive from performance bonds and public oversight) is ambiguous and depends on the outcome considered: contracts slightly worsen in terms of *Time Performance* but slightly improve on *Cost Performance*.

## 7.2 | Net (cost) effect

The performance bonding comes at a cost, as the bond premium is incorporated into the bids by potential contractors and thereby ultimately charged to the agency. To run a cost–benefit analysis, then, it is not possible to abstract from surety costs and focus on the gross effects only. Equipped with reliable estimates of the gross effect, we contrast it with the average premium prices yielding an estimate of the “net” performance bonding cost effect. This exercise is informative on at least two dimensions. First, it provides empirical evidence of the budgetary implications of performance bonding and allows for meaningful counterfactual analysis (i.e., what would happen, in the absence of the bond, to all federal construction projects). Indeed, debating the need for performance bonding in the US government procurement system, some scholars argued that the low bankruptcy rate of contractors would not justify the high surety costs (Gransberg et al., 2014). According to CGH, however, it is the surety itself that endogenously keeps the probability of bankruptcy low—and we empirically test some of their predictions in Section 7.3; on top of that, we show that the surety monitoring decreases cost overruns, proportionally more than the premium costs. Second, the exercise provides a (partial) cost–benefit analysis for mid-size projects that allows us to quantify the estimated performance bond effectiveness per bins of contract cost. One important caveat of this analysis is that the data at hand do not allow us to evaluate cost implications of performance bonds at the tender stage, despite their potentially relevant role in driving bids (for instance, CGH show how performance bonds mitigate underbidding and lower award-stage savings). Given such limitations, we are unable to run a full-fledged welfare analysis. What we propose below is a back-of-the-envelope exercise that yields a “static” cost–benefit figure—that is, we abstract from the strategic interactions between buyers and sellers at the award stage, and only compute the benefits and the costs directly related to the surety interventions.

The implications are that our analysis go beyond the US case and the empirical exercise that we propose in this paper. Several countries have introduced performance bonds in their regulations with a considerable variation in the application thresholds and without proper ex-ante feasibility analyses, leading to structural inefficiencies. Consider the case of Italy, where the performance bonding had been introduced in 2014 just to be withdrawn a year later. Imposed for projects valued more than €75 million, and without any limit to the budget covered by the bond (i.e., it exposed the sureties for an indefinite amount until project completion), the reform discouraged all insurance companies and banks from engaging, which left potential contractors without the compulsory bonds. As a result, absent sureties available to issue the bonds, public works' tenders of the projects above the above threshold were deserted, and the measure had to be immediately retracted.<sup>48</sup>

### 7.2.1 | Net (cost) effect: Performance bonding

The negative effects of the performance bonding waive are ultimately paid by the contracting agencies in terms of higher costs and longer times to completion. Hence, to run a meaningful analysis, we must take such costs into

account, while we discount the total costs by the surety bond premia. Since we do not observe the premium amount in the data, we rely on the official figures provided by the National Association of Surety Bonds Producers (NASBP, 2020) to reckon a back-on-the-envelope-calculation of the net effect of the bylaw reform. According to the latter, the average bond premium changes depend on the contract award amount, moving from 2.5% (for projects up to \$150,000) down to 1.1% for projects valued more than \$3,000,000 (see Figure 4, left panel). Provided with the distribution of premia on the one hand and with the estimated cost effects of the performance bond on the other, we are able to compute the distribution of total costs and benefits. More specifically, for each contract  $i$  we define:

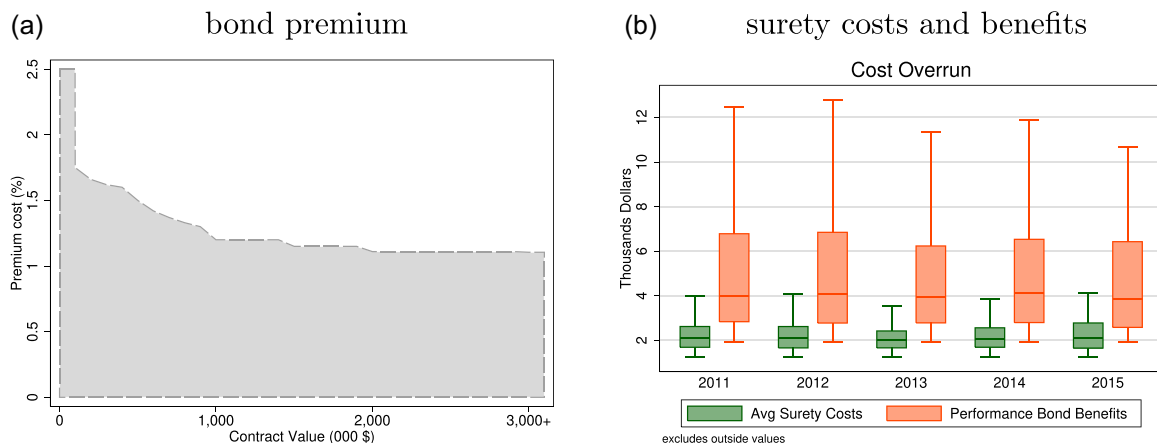
$$\text{premium\_cost}_i = \text{premium}_i \times \text{contract\_value}_i$$

and, focusing on the cost overrun dimension, we define:<sup>49</sup>

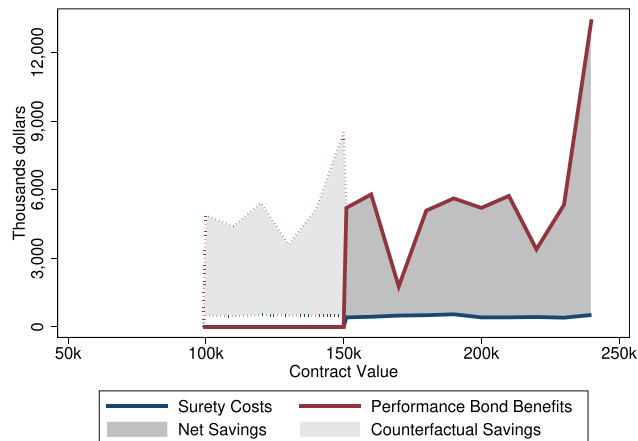
$$\text{premium\_benefit}_i = \text{cost\_overrun}_i - \frac{(1 - \text{cost\_performance}_i - 0.035) \times \text{contract\_value}_i}{(\text{cost\_performance}_i - 0.035)}$$

That is, the difference between the actual overrun and the counterfactual figure that we would have observed with the performance bonding at play.<sup>50</sup> In Figure 4(b), we contrast the distribution of the two measures for projects valued \$100,000–\$150,000, year by year after the reform to provide a glimpse of the net effect of the performance bond within the reform bandwidth. Although slightly overlapping with the tails, the plot confirms that the median potential savings per contract—the horizontal lines inside the profit figures (red boxes)—almost always lie above the *upper adjacent values* of the costs (green boxes).<sup>51</sup> Moreover, the median benefits amount to as much as twice the median costs *per contract*.

In Figure 5, we plot the empirical distribution of performance bond-induced saving—obtained by summing all net effects for each contract value class—for projects between \$50,000 and 250,000 awarded after the reform. For all projects in the \$50,000–100,000 interval, no surety costs were imposed before or after the reform; on the other hand, they did not even enjoy any potential benefit of the performance bonding. For projects above \$150,000, total benefits (solid red line) outweigh the total costs (solid blue line); total savings, represented by the gray area between the two lines, are positive in each point of the distribution. A similar pattern characterizes the counterfactual distribution of the savings; we plot it in light gray for projects between \$100,000 and 150,000. It represents the additional costs incurred (dotted blue line), and the benefits enjoyed (dotted red line) should sureties have backed up the projects. Finally, this counterfactual approach allows us to quantify the losses due to the reform of the performance bonding threshold—in graphical terms, the integral of the area bounded by the dotted lines, and the actual distribution—to  $\approx$  \$40 million until FY 2015.<sup>52</sup>



**FIGURE 4** Cost-benefit analysis: Performance bond costs (a) average bond premium (in percent of total contract value). (b) distributions of surety costs (green boxes) and performance bond benefits (red boxes), 2011–2015. The plot reports, for each figure, the upper and lower adjacent values, the interquartile range—that is, the limits of the boxes—and the median value [Color figure can be viewed at [wileyonlinelibrary.com](http://wileyonlinelibrary.com)]



**FIGURE 5** Performance bond savings. distribution of counterfactual and net savings due to performance bonding for projects valued \$50,000 to 250,000 [Color figure can be viewed at [wileyonlinelibrary.com](http://wileyonlinelibrary.com)]

### 7.2.2 | Net (cost) effect: Reform

Using a similar approach, we are able to assess the net effect of the bylaw reform (i.e., shifting up the SAT) in the public works sector. To do that, we make use of the estimate of the public oversight waive effect—that is,  $\hat{\theta}_{public}$ —which captures the effect of granting more discretion to contracting officers and less paperwork to projects within the \$100,000–150,000 range. We first compute the effect of the reform ( $\hat{\theta}_{reform} = 0.005$ ) by summing up the point estimates for performance bonding ( $\hat{\theta}_{pb}$ ) and public oversight ( $\hat{\theta}_{public}$ ). The F-test on the joint significance of  $\hat{\theta}_{pb}$  and  $\hat{\theta}_{public}$  rejects the null hypothesis of zero effect.

We can also compute the effect of the reform by accruing the average effect on all projects awarded within the \$100,000–150,000 value range. The increased discretion granted to contracting officers has overcompensated the costs of performance bonding waiving due to the same bylaw reform to around \$0.6 million to US taxpayers during FY 2011–2015. The economically negligible effect spurs from the very small magnitude of  $\hat{\theta}_{reform}$ .

Procurement offices differ dramatically in the ability to manage projects and contractors, and their competence helps select the better contractors, design better contracts, and avoid misconduct more efficiently (Baltrunaite et al., 2018). Hence, we investigate whether the competence of the contracting agency is a driver of the overall reform effect. Decarolis et al. (2020) propose a measure of contracting agency's quality defined as the persistence of project outcomes awarded by the same agency and for the same object. For each project awarded, they measure the weighted average outcome of the projects awarded by the same bureau in the past. A similar rationale applies to our framework, and we build a similar measure of past performance for each contract  $i$  as  $past\ performance_{ik} = \frac{\sum_{j_k < i_k} w_{jk} \times cost\ performance_{jk}}{\sum_{j_k} w_{jk}}$ , where  $k$  stands for the contracting agency,  $performance_{jk}$  is the *cost performance* achieved in the  $j$ th project awarded by  $k$  and  $w_{jk}$  are weights for smoothing past observations. In Table 3, we compute the net effect of the reform on four subsamples of the projects grouped by quartiles of the distribution of *office past cost performance* at the time of the contract signature. The exercise shows that  $\hat{\theta}_{reform}$  is rather heterogeneous across contracting agencies' quality: It moves from gains (+) to losses (–) as the average past performances increases.<sup>53</sup>

## 7.3 | Selection, monitoring, or both?

Via the DDD approach, we identify, in a reduced-form fashion, the gross effect of performance bonding on contract outcomes. However, the estimates alone do not help distinguish whether this effect is due to sureties' selection, monitoring, or both as outlined in Section 4. To provide suggestive evidence on  $H1$  and  $H2$  separately, we propose two empirical exercises. Specifically, we provide a piece of descriptive evidence for selection and a firm-fixed model specification for monitoring.

### 7.3.1 | Selection

To explore the role of selection played by sureties, we would ideally like to observe the bid distribution in federal construction tenders, the limited/unlimited nature of bidders, their default risk, and evaluate how this composition changes with and without performance bonding. In the absence of such granular information, we adopt an alternative



TABLE 3 Reform counterfactual analysis: cost effect

Quartiles of public office past performance index				
	1st quartile	2nd quartile	3rd quartile	4th quartile
$\hat{\theta}_{reform}$	0.094***	0.026	-0.013**	-0.023***
<i>F</i> -test	(0.00)	(0.14)	(0.08)	(0.00)
Per contract (\$000)	9.86	1.07	-4.43	-5.76
Per contract (%)	0.7	0.08	-0.33	-0.4
Total effect (\$000)	10,311	1,877	-11,196	-1615

Note: Counterfactual cost analysis of the reform by quartiles of past performance distribution. Columns 1–4 report the estimated effect of the reform by quartiles of the distribution of contracting agencies' past performance. *F*-test on joint significance reported in parentheses. For each quartile, it also displays the average *per contract* effect in \$ thousands and percentage contract value, and the total value within the treatment band.

\*\* $p < .05$ ; \*\*\* $p < .01$ .

TABLE 4 Mean differences in past performance: Limited versus unlimited winners

	Time			Cost			
	(1) $\mu^{unlim}$	(2) $\mu^{lim}$	(3) $\Delta$	(4) $\mu^{unlim}$	(5) $\mu^{lim}$	(6) $\Delta$	
(a) Unlimited vs. versus limited before the treatment							
All Winners	0.68	0.73	-0.04 (0.00)	0.76	0.81	-0.05 (0.00)	
(b) Unlimited versus limited after the treatment							
Stayers	0.69	0.73	-0.04 (0.00)	0.78	0.75	0.04 (0.00)	
Entrants	0.67	0.63	0.04 (0.00)	0.76	0.72	0.04 (0.00)	
Exiters	0.67	0.73	-0.06 (0.00)	0.75	0.81	-0.06 (0.00)	

Note: Mean comparison between all limited versus unlimited before the bylaw (panel a), or unlimited and limited *Stayers*, *Entrants*, and *Exiters* (panel b). For each group, we report the average values of past performance, the difference in the mean and the *p*-value of a *t*-test on such difference—in parenthesis—for both time and cost (columns 1–3 or 4–6, respectively).

approach hinging on the behavior of *winning* firms with and without bonds across treatment and control groups of contracts. We start by partitioning firms in the construction treatment subgroup into three clusters: *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 \$150,000 afterward; *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 afterward.

According to CGH (i.e., *H1*), the pool of winning limited companies must change with the treatment. This prediction would be reflected in lower-quality limited liability firms in the pool of entrants and higher-quality ones in the pool of exiters given that the absence of premia does not counterbalance their aggressive bidding strategy.<sup>54</sup> Conversely, *H2* does not entail any structural adjustment in the features of winning firms. According to this hypothesis, sureties do not exert any ex-ante selection and affect the outcomes only through the ex-post supervision of work progress. In this scenario, the same firms are awarded projects 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 *Stayers*, *Entrants*, and *Exiters* should be similar in observable characteristics.

In Table 4, we report the mean comparisons of past performance in terms of time (Columns 1 to 3) and cost (Columns 4 to 6) between limited and unlimited firms. To test the dynamic effects of the reform we look separately at

all winners *before* the reform (panel a) and at *Stayers, Entrants* and *Exiters after 2011* (panel b). We can summarize the evidence as follows:

- Limited versus unlimited with sureties: When subject to the performance bonding obligations, the selected limited liability firms show higher quality than their unlimited counterparts (panel a). This is consistent with the existence of a “penalty” for limited liability companies which favors unlimited firms, irrespective of their quality;
- Limited entrants versus stayers/exiters: The limited companies that enter the market *after* the reform are rather poor performers with respect to both the pool of Stayers and the pool of Exiters—up to  $-13\%$  (time) and  $-11\%$  (cost);
- Unlimited entrants versus stayers/exiters: the unlimited companies that enter the market *after* the reform, on the other hand, perform in line with both the Stayers and the Exiters, showing virtually no difference in past performances;
- Limited versus unlimited Entrants/exiters: in line with previous points, the limited Entrants are of lower quality, whereas the limited Exiters show much better performances than their unlimited counterparts.

These descriptive pieces of evidence suggest that a selection channel for limited liability companies is in place through the performance bonding.

### 7.3.2 | Monitoring

After assessing that selection is a channel at stake, we provide suggestive evidence also on the concurrent monitoring effect (i.e., *H2*). To do so, in Column 2 of Table 5, we estimate Equation (1), controlling for firm fixed effects. Our goal is

TABLE 5 Monitoring

Panel (a): Time Performance			
	Baseline (1) Base	Monitoring	
		(2) firmFE	(3) PMFE
$\hat{\theta}_{pb}$	$-0.088^{***}$ (0.023)	$-0.050$ (0.032)	$-0.130^{***}$ (0.049)
<i>N</i>	120,175	70,758	60,248
$R^2$	0.128	0.371	0.376
$Avg_{services}$	0.808	0.808	0.808
$Avg_{works}$	0.837	0.837	0.837
Panel (b): Cost Performance			
	(1) Base	(2) firmFE	(3) PMFE
$\hat{\theta}_{pb}$	$-0.035^{***}$ (0.013)	$-0.054^{**}$ (0.021)	$-0.067^{**}$ (0.029)
<i>N</i>	120,175	70,758	60,248
$R^2$	0.218	0.421	0.430
$Avg_{services}$	0.868	0.868	0.868
$Avg_{works}$	0.945	0.945	0.945
Firm FE		✓	✓
Prod × Multi FE			✓

Note: Column 1 report baseline estimates from Table 2. Column 2 includes contractor fixed effects and Column 3 adds up a control for project complexity-fixed effects for the interaction of procurement categories with a dummy indicating whether a project is expected to last more than 1 year. Standard errors are two-way clustered at the object and FY level.

\*\* $p < .05$ ; \*\*\* $p < .01$ .

to identify whether the very same firm performs differently with or without performance bonds. Relative to the baseline estimates (reported in Column 1), we must restrict the analysis to firms featuring at least two completed projects, thus reducing the estimation sample to 70,714 contracts and 9771 firms. The estimates on time performance hold to the baseline, while they almost double in magnitude for cost performance. These results suggest that the same contractor once waived from the surety oversight, performs worse in terms of delays and extra costs.

Although the main result is robust to the inclusion of contractor fixed effects, this does not rule out the selection of weaker contractors as a channel through which waiving performance bonding affects outcomes. Indeed, holding contractors constant merely controls for the overall quality of contractors, not for how suited a particular contractor is for a specific contract. Thus, results from Column 2 might be driven by the exemption of performance bonds leading to a mismatch between contracts and contractors. To rule out this possibility, ideally we would like to run the same model by adding a control for project complexity and see whether the very same contractor performs equally on very similar projects. In the absence of a measure of complexity valid for all the procurement categories and industries we span, we are to rely on a proxy variable. The FPDS reports the binary variable “Multi-Year Contract,” indicating that the contract is issued under specific Congressional multi-year authority for specific programs applicable to complex and long purchasing process.<sup>55</sup> Thus, in Column 3, we enrich the model with fixed effects for the interaction of procurement categories with the multi-year dummy to control for specific subindustry-level idiosyncrasies of the variable definition. In other words, we want to test not only whether the same firm operates better under the supervision of a surety but whether this also applies to projects of similar procurement-wise complexity. Qualitatively, the results hold again. The treatment effect—over both dimensions of project metrics—is twice the magnitude of the baseline when fixing the supplier and detailed project-level characteristics. This robust finding suggests that the impact of sureties in the supply chain is not merely loaded into the selection of the contractors—as suggested above—but also involves all the additional activities in the operational phase of the project. Consequently, performance bonding seems to be impactful on the procurement process both for supplier selection and contract enforcement.

## 8 | CONCLUSIONS

The rising pressure on the US government to contain the public deficit forces public entities affected by capacity shortages (Warren, 2014) to explore alternative options to increase their effectiveness in providing high-quality public goods. Moreover, our paper comes at a time when economies have been severely affected by the COVID-19 pandemic, and governments plan to increase demand for infrastructures development to boost recovery. The purpose of our work is to provide public managers and other stakeholders with a quantification of the benefits of performance bonds on the procurement of infrastructure through an application to field data. We show that performance bonds sizeably improve project outcomes by curtailing extra costs and delays at the cost of a relatively low premium.

Our findings suggest that the procurement managers should consider performance-bond-type tools in two general scenarios. On the one side, third parties—broadly speaking, private sector providers—should have a better knowledge of the market or, likewise, be able to convert soft and unverifiable information (e.g., managerial practices, past performances on private projects) into hard and verifiable data. On the other hand, the underlying business should be intrinsically risky and the purchase relatively complex. We demonstrate that performance bonding streamlines public projects management by providing support to contracting officers in the above scenarios by enhancing supplier selection and contract enforcement.

Our results apply to a market that includes more than 19,000 contractors with 1500 service and construction task categories performed all over the US territory over 11 years. Our results prove to be robust within different subcategories of constructions (which sureties may have a different experience with), states (where availability of sureties differs), and years (controlling for business cycles and their implications for the national construction market). Nonetheless, we are careful about the generalizability of our findings as our results hinge on the specific features of our institutional setting. We cannot claim that performance bonds, with everything else equal, are always desirable tout court; yet, we are confident that our results can be replicated outside the United States in countries with a similarly developed insurance market, in sectors that are similarly complex to constructions and where third parties with superior industry knowledge are available, and with the same rules with which the US writers comply.

The implications of our study foster the ongoing discussion by strategy literature on how public organizations and private firms interact to create and share value. The rigidities of public contracting can be powerful obstacles to the value creation of governance arrangements, as emphasized in recent contributions on value creation and appropriation

in public and nonprofit organizations (Cabral et al., 2019). We show that the interplay between public agencies and private service providers creates value along the supply chain. On the one hand, providing public goods of higher quality builds state capacity by generating a cascade of beneficial effects (e.g., boost citizens' trust, free public budget, and resources). On the other hand, we are ignoring the profits that surety companies gain from their business. By focusing on the direct effect of performance bonding, we remain agnostic about how much of the overall created value the government and the sureties share.

We propose a back-of-the-envelope analysis to reckon the aggregate impact of the SAT bylaw: The analysis yields an almost-zero effect—that is, the gains from removing public oversight roughly compensate losses from waiving performance bonding. The straightforward implication of our findings is that a more effective reform should retain both the benefits of performance bonds and more discretion for public procurement managers for mid-size projects. We are anyways careful about the complete welfare evaluation of performance bonds in public procurement. First, the limitation of our back-of-the-envelope calculations is that waiving the need for performance bonds might have made agencies able to achieve lower procurement prices at the tender stage. Given our lack of information about the counterfactual award price absent the premium at the tender stage, it is difficult to draw conclusions regarding total welfare. Moreover, a complete evaluation would also require considerations on competition and firm dynamics. Sureties' premia reward those firms capable of performing the project under consideration, conditioning the capability assessment for measures like past performance in public as well as private projects. In this sense, small and entrant firms could be negatively affected by performance bonding—at the very least, data on newcomers do not exist, and sureties are forced to use aggregated data, or no data sources at all, for their assessment.<sup>56</sup> Although the prequalification process does not exclude those categories a priori, it can effectively amount to a market entry barrier in the short run. In the long run, the process might educate and guide excluded firms about the bonding process and prepare them to be more successful in future tenders.

## ACKNOWLEDGMENTS

An earlier version of this paper circulated under the title *Can the Private Sector Ensure the Public Interest? Evidence from Federal Procurement*. We are grateful to the Editor Ramon Casadesus-Masanell and the anonymous Co-editor at the *Journal of Economics & Management Strategy*, plus the two anonymous referees for their valuable suggestions and constructive comments throughout the review process. We thank Francesco Decarolis and Giancarlo Spagnolo for their precious advice in the earlier stage of the project. For the comments received, we also thank Zareh Asatryan, Jean Beuve, Nicolas Fugger, Stefano Gagliarducci, Juan-José Ganuza, Vitalijs Jascisens, Sofia Lundberg, Calvin Luscombe, Neale Mahoney, Stéphane Saussier, Hidenori Takahashi, and Achim Wambach, plus the colleagues and participants at ZEW, Bank of Italy, Einaudi Institute for Economics and Finance, University of Mannheim, University Pompeu Fabra, Italian National Research Council-GRAPE, European Association for Research in Industrial Economics, University of Rome Tor Vergata, ICEE, WEEE, Società Italiana di Economia Pubblica, University of Bozen/Bolzano, Jornadas de Economía Industrial, LAGV, EEEP Chair, and the Association for Public Economic Theory seminars and conferences. We also thank the Società Italiana di Economia Pubblica for the SIEP Best Paper Award 2017. The views expressed herein are those of the authors and do not necessarily reflect those of the Bank of Italy. All errors are our own. Open Access funding enabled and organized by Projekt DEAL.

## ORCID

Leonardo M. Giuffrida  <http://orcid.org/0000-0001-8397-799X>

## ENDNOTES

<sup>1</sup> Source: *US Aspending.gov*, accessed on November, 2019.

<sup>2</sup> See Barkley (2018) for an overview of the relevant discussion.

<sup>3</sup> Public-private partnerships can improve the efficiency of production, delivery, and maintenance of public goods. See Fabre and Straub (2019) for recent reviews on this topic, focusing mainly on infrastructure projects.

<sup>4</sup> Implications of contract performance bonds in the field are discussed only in technical reports. See, for example, GAO (2013) and Garvin et al. (2011).

<sup>5</sup> More specifically, delays and extra cost relative to the expected duration and cost, respectively are our baseline metrics. Examples of empirical analyses of government efficiency that use direct measures of outcomes, the approach that our work follows, include (Decarolis et al., 2020), Ferraz and Finan (2008), Olken (2007), and Hyytinen et al. (2018).

- <sup>6</sup> Examples include awarding design (Decarolis, 2018), wasteful end-year spending (Liebman & Mahoney, 2017), buyers' role (Best et al., 2017; Decarolis et al., 2021, 2020), external audits (Gerardino et al., 2017), industry consolidation (Carril & Duggan, 2020), and the effects of centralized purchase agreements (Bandiera et al., 2009).
- <sup>7</sup> For instance, out of the 1,424,124 US contractors in business in 2007, only 969,937 were still in business in 2009—a 31.9% exit rate (NASBP, 2020).
- <sup>8</sup> Sureties in the United States are subsidiaries or divisions of insurance companies. Both performance bonds and insurance policies are different kinds of risk transfer mechanisms. Performance bonds offer assurance that the contractor is capable of completing the contractual tasks on time, within budget, and according to contractual specifications.
- <sup>9</sup> More in detail, according to the National Association of Surety Bond Producers and Surety and Fidelity Association of America, the most commonly checked criteria are the references and reputation of the firm and its employees; the ability to meet current and future obligations; profitability history; the degree of matching between previous experience and contract requirements; availability of the necessary equipment to do the work or the ability to obtain it; workload; management practices; financial strength; credit history; and an established bank relationship and line of credit. See <http://suretyinfo.org/>.
- <sup>10</sup> On top of that, for public buyers, there are limits to the amount of soft information that can be translated into “hard” data (Albano et al., 2017), while sureties, being private companies, are not constrained by the law in the amount of soft information that can be used during their business activities.
- <sup>11</sup> In the US construction industry, construction bonds generate two-thirds of the total surety premia written and 70% of total revenues. See <https://www.nasbp.org/home>.
- <sup>12</sup> The existing theoretical literature on performance-bond-like tools (Birulin, 2020; Calveras et al., 2004; Engel et al., 2006; Wambach & Engel, 2011) assumes that the potential contractors include the premium amount in their bids. When it comes to federal constructions, the regulation requires explicitly reporting the bond and premium amount to the contracting officer via filling out an ad hoc form (see the next subsection). Hence, it is the buyer that ultimately pays the premium.
- <sup>13</sup> Sureties usually offer flat rate fees depending on the amount of coverage requested by the potential contractor. Fees are included in the premium when the firm is awarded the contract—and, therefore, paid by the buyer—or directly paid by the construction firm in case the firm is not awarded the contract.
- <sup>14</sup> The most common method employed is percentage-of-completion accounting, whose goal is to measure the results of bonded work performed during the accounting period. The percentage of contract values recognized as revenue is typically done on a cost-to-cost percentage-of-completion method.
- <sup>15</sup> In 2009, the surety industry paid more than \$84 million in losses on private construction and more than \$1.6 billion since 1995.
- <sup>16</sup> Sureties writing on federal projects must possess a certificate of authority from the Department of Treasury, which conducts a thorough financial review of the surety and sets a single bond size limit for that surety (FAR 28.202). The updated list of official writers is available at <https://www.fiscal.treasury.gov/surety-bonds/list-certified-companies.html>.
- <sup>17</sup> The amount of the bond cannot exceed the specified value of that guarantee. A performance bond is not a general insurance policy and, if cashed by the principal, the payment amount does not exceed the signed contract value.
- <sup>18</sup> Note that bidders cannot influence the contract regime by setting their bids below the SAT, given that it applies to the ex-ante valuation of the contract value (or “anticipated dollar value,” according to FAR terminology), irrespective of the actual award value.
- <sup>19</sup> 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. As a side note, there has been no adjustment to procurement thresholds after the reviews in 2005 and 2015.
- <sup>20</sup> The FY ends in September. The threshold revision was, therefore, enforced from FY 2011 on, that is, as of October 2010.
- <sup>21</sup> Such red tape is acknowledged by FAR 13.002 itself, which explicitly prescribes SAPs to reduce administrative costs; improve opportunities for small or disadvantaged firms to obtain a fair proportion of government contracts; promote efficiency and economy in contracting; and avoid unnecessary burdens for agencies and contractors.
- <sup>22</sup> In CGH terms, low quality means for a firm to be smaller and of more limited liability. Indeed, limited liability companies are more willing to bid aggressively because they face, by definition, limited risks. As a result, they are more likely to underbid, face unexpected shocks, go bankrupt, or renegotiate contract terms. CGH then show that surety bonds indeed mitigate the problem of abnormally low tenders. CGH assume that the cost of the surety bond is proportional to its size. Wambach and Engel (2011) modify their framework and assume in a first step that surety bonds are priced fairly—the common benchmark case in the insurance literature. Then, in contrast to CGH, the problem becomes analytically solvable, and full insurance or even over-insurance is optimal, that is, the project is always finished either by the contractor or by the surety. In a second step, they introduce a risk loading (unfair premia) and show that full insurance or even over-insurance might still be optimal.
- <sup>23</sup> Sureties are entitled to file claims against the contractor and have power to perform an actual oversight of contract execution before any claim is filed. Most importantly, the constant interaction of sureties and public contractors proves to be a crucial feature in this market. Indeed, monitoring is a successful tool in hindering moral hazard in repeated games, where cooperative equilibria are sustainable.



- <sup>24</sup> For example, the red tape in the case of public oversight.
- <sup>25</sup> The FPDS data is already used and described in the literature. See, for instance, Liebman and Mahoney (2017) and Kang and Miller (2016), and Decarolis et al. (2021).
- <sup>26</sup> Data is gathered by contracting offices of 23 federal agencies. In the Appendix D, we report the number of projects per government department/FY composing our working sample, depending on different specifications of the control group, respectively.
- <sup>27</sup> The information on the procurement category of each awarded contract is sourced from the FPDS variable “Product or Service Code,” which is composed of two alphabetic and two numeric digits. The *Object* variable coincides with the first digit of “Product or Service Code.”
- <sup>28</sup> Consider two projects, A and B, whose observed contract value is \$105,000, both awarded before the threshold revision. Suppose the (unobservable to us) engineers’ estimated value of A, that is,  $EV_A$ , is \$110,000, while  $EV_B = \$95,000$ . According to the contract value information alone, they should both be subject to oversight as both are above the pre-FY 2011 SAT. However, we observe that  $SAP_A = 0$  and  $SAP_B = 1$ . Therefore, we can proceed to the correct identification and avoid any source of bias in the estimates by labeling A as below SAT and B above SAT.
- <sup>29</sup> 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.
- <sup>30</sup> Before initiating a modification, the contracting officer must determine if the proposed effort is within the scope of the existing contract or is a new acquisition outside of the scope. A new requirement outside of the scope of the existing contract must be processed as a new acquisition. Contract scope means, in simple terms, that the contemplated change must be generally related to the work originally contracted for. If a contract was awarded for the design only of an automated information system, it could not be later modified to have the contractor provide and install hardware. See also Decarolis et al. (2020) for a further discussion on the distinction between in-scope and out-of-scope negotiation.
- <sup>31</sup> Spiller (2008)’s argument unfolds as follows: given the formal, bureaucratic nature of public contracting, any term’s 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 the adaptation costs in their construction data in 8%–14% of the winning bid. Recently, Beuve et al. (2019) found that public procurement contracts feature more rigidity clauses than private procurement contracts and that the use of rigidity clauses in public contracts rises when political risks are more salient.
- <sup>32</sup> In Appendix A, we display the robustness of our results to different definitions of contract outcomes in terms of delay and cost overruns. More specifically, we employ the plain overrun as well as a linear transformation obtained using the ratio of overrun and initial expected outcome. Both variables, however, feature a non-negligible number of outliers that drive the estimated parameters (one contract shows a time overrun of 1,644 times the initial expected time to completion) that we are forced to exclude from the sample. The convexity of the “performance” transformation flattens the outliers toward zero and allows us to run the estimations without losing data.
- <sup>33</sup> This feature of our data differs from Decarolis (2014), which finds a nearly zero correlation between time and cost renegotiations and no evidence of a nonlinear relationship. However, the author stresses 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 leads to shorter time and greater cost overruns. We are not able to reproduce these results since the FPDS does not contain similar information.
- <sup>34</sup> The Small Business Authority (SBA) labels small firms based on the particular service category that the contract belongs to, as well as on contractors’ characteristics (revenues, number of employees, etc).
- <sup>35</sup> We consider as competitive a lot for which the extent of competition is labeled “Full and Open” and whose participation is not set aside to any specific group of firms. In noncompetitive tenders, the participation criteria restrict the competition ex ante to dimensions other than quality (e.g., Athey et al., 2011).
- <sup>36</sup> The typical supply (i.e., physical goods) contract shows a 0 value in *time/cost overruns* and a unit value in both performances.
- <sup>37</sup> According to our procurement categorization, services and objects included in the sample are: R&D; 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. Due to their peculiar nature, we test the robustness of results to the exclusion of R&D the Appendix A. Construction objects—composing our dummy *Constr* following the definition of “Construction” in the Miller Act—are instead: Constructions of structure/facilities; maintenance, repair, alteration of structures Facilities.
- <sup>38</sup> The firm ID variable is missing in approximately 53% of the projects in our sample.
- <sup>39</sup> As discussed in Appendix C, in our data, there is a clear indication of sorting before different procedural thresholds, and we cannot implement a regression discontinuity analysis. This is also justified in this context as firms self-select into different tenders.

- <sup>40</sup> In Appendix C, we discuss the identification assumptions underlying our DDD approach, showing that the public oversight parameter that we estimate is qualitatively and quantitatively comparable to one estimated in a plain DD setting on a subset of constructions projects not subject to performance bonding. See Berck and Villas-Boas (2016) for an overarching discussion on the DDD methodology.
- <sup>41</sup> Specifically,  $Waiver_{it}$  is a binary variable indicating whether the contract value lies between \$100,000 and \$150,000. In the pre-reform period, the variable is set to zero for all contracts in the treatment band procured via SAP. After the reform, the variable is set to zero for all contracts in the treatment band which are not procured via SAP.
- <sup>42</sup> Please refer to Appendix C for identification tests on  $\hat{\theta}_{public}$ .
- <sup>43</sup> See Section 5 of Olden and Møen (2020) for the analytic proof.
- <sup>44</sup> For a discussion on the role of selection into the treatment issue in our setup, we refer to Appendix B.
- <sup>45</sup> Consider the case in which  $\theta_{public} = X$  in all sectors but construction, where it amounts to  $-X$ , and assume  $\theta_{pb} = 0$ . In such a case, our model would yield a badly biased  $\hat{\theta}_{pb} = -2X$ .
- <sup>46</sup> Full tables are available from the authors upon request.
- <sup>47</sup> From a purely econometric point of view, the interpretation of these changes hinges on the existence of agency-specific intercepts: as shown in Decarolis et al. (2020), the persistency of agency's performance matters for contract execution.
- <sup>48</sup> For details, see <https://www.lavoce.info/archives/51317/opere-pubbliche-cerca-garanzie/> (in Italian).
- <sup>49</sup> It is hard to quantify the “value of time” per contract, to also include the savings in terms of time overrun driven by performance bonds. One straightforward way would be to compute the average value of a day (i.e.,  $\frac{\text{estimated\_cost}_i}{\text{estimated\_time}_i}$ ), but that would include all costs, including administrative and input expenses, in the *per day* cost evaluation. Hence, we decided to only include the direct measures of costs: in this sense, our estimates of the net benefits reach a lower bound of actual savings.
- <sup>50</sup> Note that to get the overrun, we revert the performance measure, that is,  $Overrun = \left(\frac{1}{Performance} - 1\right) \times Expected\ Outcome$ .
- <sup>51</sup> We define the upper adjacent value as the largest observation, which is less than or equal to the third quartile plus  $1.5 \times$  interquartile range.
- <sup>52</sup> The actual estimate is \$39,699,000 and is a static figure, in the sense that it does not take into account the dynamic interactions that we described in the previous sections, and has been estimated on projects whose performance had already been waived both from the monitoring and the selection effect of the sureties. In this respect, we consider it an upper bound of the actual value.
- <sup>53</sup> The leading role of the awarding agency is also suggested by the impact of the agency fixed effects in our baseline analysis.
- <sup>54</sup> Limited liability firms can bid more aggressively because their financial responsibility in case of loss is bounded. For more details on this, see CGH.
- <sup>55</sup> The key distinguishing difference between a multiple year (directly sourceable from the data) and a multi-year contract is the latter authorize performance beyond the period, for which funding has been appropriated, without having to exercise options for each program year (FAR 17.103). Irrespective of the *actual* duration of the contract, a multi-year contract is therefore already expected to last more than one FY.
- <sup>56</sup> The EU Member States have substantial leeway to regulate the issue of financial guarantees under domestic public procurement law. The reluctance of setting up a joint practice hinges on the idea that performance bonding may give rise to restrictions of competition by preventing the participation of relatively more financially constrained bidders, particularly SMEs. The characteristics of the US performance bonds—based on capacity and firm suitability more than on financial guarantees—are instead compatible with European regulations. See Sanchez-Graells (2015) for a review of the debate on bid bonds, performance bonds, and other sureties or financial guarantees in European public procurement law.
- <sup>57</sup> As remarked in the body of the text, we only observe an indicator variable for the valuation being above or below the threshold.
- <sup>58</sup> Due to the unboundedness of the variable, a 1:1 comparison between point estimates would be meaningless. However, we record that the implied percentage changes concerning the average outcome values are of the same order of magnitude, with a different sign.
- <sup>59</sup> In the analysis, aggregate the series at the yearly level, and use the average values.
- <sup>60</sup> See the related discussion in Section 6.2.
- <sup>61</sup> An instance of bunching in procurements just below legislative thresholds is presented in Palguta and Pertold (2017).

## REFERENCES

- Abadie, A., Athey, S., Imbens, G. W., & Wooldridge, J. (2017). *When should you adjust standard errors for clustering?* (Technical report). National Bureau of Economic Research.
- Adida, E., & Bravo, F. (2019). Contracts for healthcare referral services: Coordination via outcome-based penalty contracts. *Management Science*, 65(3), 1322–1341.

- Albano, G. L., Cesi, B., & Iozzi, A. (2017). Public procurement with unverifiable quality: The case for discriminatory competitive procedures. *Journal of Public Economics*, 145, 14–26. <https://doi.org/10.1016/j.jpubeco.2016.11.004>
- Athey, S., Levin, J., & Seira, E. (2011). Comparing open and sealed-bid auctions: Evidence from timber auctions. *The Quarterly Journal of Economics*, 126, 207–257.
- Autor, D. H. (2003). Outsourcing at will: The contribution of unjust dismissal doctrine to the growth of employment outsourcing. *Journal of Labor Economics*, 21(1), 1–42.
- Bajari, P., Houghton, S., & Tadelis, S. (2014). Bidding for incomplete contracts: An empirical analysis of adaptation costs. *American Economic Review*, 104(4), 1288–1319.
- Bajari, P., & Lewis, G. (2014). Moral hazard, incentive contracts, and risk: Evidence from procurement. *The Review of Economic Studies*, 81(3), 1201–1228.
- Baltrunaite, A., Giorgiantonio, C., Mocetti, S., & Orlando, T. (2018). Discretion and supplier selection in public procurement. *The Journal of Law, Economics, and Organization*, 37(1), 134–166.
- Bandiera, O., Prat, A., & Valletti, T. (2009). Active and passive waste in government spending: Evidence from a policy experiment. *The American Economic Review*, 99(4), 1278–1308.
- Barkley, A. (2018). *Cost and efficiency in dynamic government outsourcing: Evidence from the dredging industry*. [Doctoral dissertation, Carnegie Mellon University]. <https://doi.org/10.1184/R1/6715430.v1>
- Bauhr, M., Czibik, A., de Fine Licht, J., & Fazekas, M. (2019). Lights on the shadows of public procurement: Transparency as an antidote to corruption. *Governance*, 33(3), 495–523. <https://doi.org/10.1111/gove.12432>
- Berck, P., & Villas-Boas, S. (2016). A note on the triple difference in economic models. *Applied Economics Letters*, 23(4), 239–242.
- Bertrand, M., Duflo, E., & Mullainathan, S. (2004). How much should we trust differences-in-differences estimates? *The Quarterly Journal of Economics*, 119(1), 249–275.
- Best, M. C., Hjort, J., & Szakonyi, D. (2017). *Individuals and organizations as sources of state effectiveness, and consequences for policy* (National Bureau of Economic Research Working Paper 23350). <http://bit.ly/2dlCTyo>
- Beuve, J., Moszoro, M. W., & Saussier, S. (2019). Political contestability and public contract rigidity: An analysis of procurement contracts. *Journal of Economics & Management Strategy*, 28(2), 316–335.
- Birulin, O. (2020). Optimality of simple procurement auctions. *International Journal of Industrial Organization*, 70(C), 102610.
- Board, S. (2007). Bidding into the red: A model of post-auction bankruptcy. *Journal of Finance*, 62(6), 2695–2723.
- Branzoli, N., & Decarolis, F. (2015). Entry and subcontracting in public procurement auctions. *Management Science*, 61(12), 2945–2962.
- Burguet, R., Ganuza, J.-J., & Hauk, E. (2012). Limited liability and mechanism design in procurement. *Games and Economic Behavior*, 76(1), 15–25.
- Cabral, S., Mahoney, J. T., McGahan, A. M., & Potoski, M. (2019). Value creation and value appropriation in public and nonprofit organizations. *Strategic Management Journal*, 40(4), 465–475. <https://doi.org/10.1002/smj.3008>
- Callaway, B., & Sant'Anna, P. H. (2021). Difference-in-differences with multiple time periods. *Journal of Econometrics*, 225(2), 200–230.
- Calveras, A., Ganuza, J., & Hauk, E. (2004). Wild bids. gambling for resurrection in procurement contracts. *Journal of Regulatory Economics*, 26(1), 41–68.
- Calvo, E., Cui, R., & Serpa, J. C. (2019). Oversight and efficiency in public projects: A regression discontinuity analysis. *Management Science*, 65(12), 5651–5675.
- Calzolari, G., & Spagnolo, G. (2009). *Relational contracts and competitive screening* (CEPR Discussion Papers).
- Carril, R. (2021). *Rules versus discretion in public procurement* (Economics Working Papers 1765). Department of Economics and Business, Universitat Pompeu Fabra. <https://ideas.repec.org/p/upf/upfgen/1765.html>
- Carril, R., & Duggan, M. (2020). The impact of industry consolidation on government procurement: Evidence from Department of Defense contracting. *Journal of Public Economics*, 184, 104141.
- Castellani, L., Decarolis, F., & Rovigatti, G. (2018). *Procurement centralization in the EU: The case of Italy* (CEPR Discussion Paper No. DP12567).
- Chen, S., & Lee, H. (2017). Incentive alignment and coordination of project supply chains. *Management Science*, 63(4), 1011–1025. <https://doi.org/10.1287/mnsc.2015.2373>
- Chillemi, O., & Mezzetti, C. (2014). Optimal procurement mechanisms: Bidding on price and damages for breach. *Economic Theory*, 55(2), 335–355.
- Coviello, D., Guglielmo, A., & Spagnolo, G. (2018). The effect of discretion on procurement performance. *Management Science*, 64(2), 715–738. <https://doi.org/10.1287/mnsc.2016.2628>
- Davis, K. (2014). Different stakeholder groups and their perceptions of project success. *International Journal of Project Management*, 32(2), 189–201. <https://doi.org/10.1016/j.ijproman.2013.02.006>
- Decarolis, F. (2014). Awarding price, contract performance and bids screening: Evidence from procurement auctions. *American Economic Journal: Applied Economics*, 6(1), 108–132.
- Decarolis, F. (2018). Comparing public procurement auctions. *International Economic Review*, 59(2), 391–419.
- Decarolis, F., de Rassenfosse, G., Giuffrida, L., Iossa, E., Mollisi, V., Raiteri, E., & Spagnolo, G. (2021). Buyers' role in innovation procurement: Evidence from U.S. Military R&D Contracts. *The Journal of Economics & Management Strategy*, 30, 697–720.
- Decarolis, F., Giuffrida, L. M., Iossa, E., Mollisi, V., & Spagnolo, G. (2020). Bureaucratic competence and procurement outcomes. *The Journal of Law, Economics, and Organization*, 36(3), 537–597. <https://doi.org/10.1093/jleo/ewaa004>
- Decarolis, F., Spagnolo, G., & Pacini, R. (2016). *Past performance and procurement outcomes* (Working Paper 22814). National Bureau of Economic Research. <https://doi.org/10.3386/w22814>

- Engel, A., Ganuza, J.-J., Hauk, E., & Wambach, A. (2006). Managing risky bids. In D. Nicola, P. Gustavo, & S. Giancarlo (Eds.), *Handbook of procurement* (pp. 322–344). Stockholm School of Economics.
- Fabre, A., & Straub, S. (2019). The economic impact of public private partnerships (ppps) in infrastructure, health and education: A review. TSE Working Paper 986.
- Ferraz, C., & 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.
- Fugger, N., Katok, E., & Wambach, A. (2019). Trust in procurement interactions. *Management Science*, 65(11), 5110–5127.
- GAO. (2013). *Agencies need to strengthen oversight of multibillion dollar investments in operations and maintenance* (Technical Report GAO-14-66). US Government Accountability Office.
- Garvin, M., Molenaar, K., Navarro, D., & Proctor, G. (2011). *Key performance indicators in public-private partnerships* (Technical Report FHWA-PL-10-029). US Department of Transportation.
- Gerardino, M. P., Litschig, S., & Pomeranz, D. (2017). *Can audits backfire? Evidence from public procurement in Chile* (Working Paper 23978). National Bureau of Economic Research. <https://doi.org/10.3386/w23978>
- Goh, J., & Hall, N. G. (2013). Total cost control in project management via satisficing. *Management Science*, 59(6), 1354–1372.
- Gransberg, D. D., Kraft, P., & Park, H. (2014). Performance bond: Cost, benefit, and paradox for the public highway 3 agencies. *Transportation Research Record*, 2408, 3–9.
- Handley, S. M., & Angst, C. (2015). The impact of culture on the relationship between governance and opportunism in outsourcing relationships. *Strategic Management Journal*, 36(9), 1412–1434.
- Herweg, F., & Schwarz, M. A. (2018). Optimal cost overruns: procurement auctions with renegotiation. *International Economic Review*, 59(4), 1995–2021.
- Hyytinen, A., Lundberg, S., & Toivanen, O. (2018). Design of public procurement auctions: Evidence from cleaning contracts. *The RAND Journal of Economics*, 49(2), 398–426. <https://doi.org/10.1111/1756-2171.12232>
- Kang, K., & Miller, R. A. (2016). *Winning by default: Why is there so little competition in government procurement?* Mimeo.
- Krasnokutskaya, E., & Seim, K. (2011). Bid preference programs and participation in highway procurement auctions. *American Economic Review*, 101(6), 2653–2686.
- Liebman, J. B., & Mahoney, N. (2017). Do expiring budgets lead to wasteful year-end spending? Evidence from federal procurement. *American Economic Review*, 107(11), 3510–3549. <https://doi.org/10.1257/aer.20131296>
- 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.
- McCrary, J. (2008). Manipulation of the running variable in the regression discontinuity design: A density test. *Journal of Econometrics*, 142(2), 698–714.
- Myerson, R. B. (1981). Optimal auctions design. *Mathematics of Operations Research*, 6(1), 58–73.
- NASBP. (2020). *Bond execution guide* (Technical report). National Association of Surety Bond Producers.
- Olden, A., & Moen, J. (2020). *The triple difference estimator* (NHH Department of Business and Management Science Discussion Paper (2020/1)).
- Olken, B. A. (2007). Monitoring corruption: Evidence from a field experiment in Indonesia. *Journal of Political Economy*, 115(2), 200–249.
- Palguta, J., & Pertold, F. (2017). Manipulation of procurement contracts: Evidence from the introduction of discretionary thresholds. *American Economic Journal: Economic Policy*, 9(2), 293–315. <https://doi.org/10.1257/pol.20150511>
- Roach, M., & Cohen, W. M. (2013). Lens or prism? Patent citations as a measure of knowledge flows from public research. *Management Science*, 59(2), 504–525.
- Sanchez-Graells, A. (2015). *Public procurement and the EU competition rules*. Hart Publishing.
- Shao, J., Müller, R., & Turner, J. R. (2012). Measuring program success. *Project Management Journal*, 43(1), 37–49.
- Smirnov, J. P., & Hicks, M. J. (2008). The impact of economic factors and acquisition reforms on the cost of defense weapon systems. *Review of Financial Economics*, 17(1), 3–13. Special Issue on Defense Industries. <https://doi.org/10.1016/j.rfe.2007.02.005>
- Spiller, P. T. (2008). *An institutional theory of public contracts: Regulatory implications* (NBER Working Paper 14152). <https://doi.org/10.3386/w14152>
- Waehrer, K. (1995). A model of auction contracts with liquidated damages. *Journal of Economic Theory*, 67(2), 531–555.
- Wambach, A., & Engel, A. R. (2011). Surety bonds with fair and unfair pricing. *The Geneva Risk and Insurance Review*, 36(1), 36–50. <https://EconPapers.repec.org/RePEc:pal:genrir:v:36:y:2011:i:1:p:36-50>
- Warren, P. L. (2014). Contracting officer workload, incomplete contracting, and contractual terms. *The RAND Journal of Economics*, 45(2), 395–421. <https://doi.org/10.1111/1756-2171.12056>
- Wooldridge, J. M. (2003). Cluster-sample methods in applied econometrics. *American Economic Review*, 93(2), 133–138.
- Zheng, C. Z. (2001). High bids and broke winners. *Journal Economic Theory*, 100, 129–171.

**How to cite this article:** Giuffrida, L. M., & Rovigatti, G. (2022). Supplier selection and contract enforcement: Evidence from performance bonding. *Journal of Economics & Management Strategy*, 31, 980–1019. <https://doi.org/10.1111/jems.12492>

## APPENDIX A: MAIN ROBUSTNESS ANALYSIS

We test the robustness of our baseline DDD findings over two dimensions. First, we check whether the results are robust across different subsamples of contracts we deem particularly relevant to our setting; second, we run the estimation on alternative specifications of contract outcomes to validate the goodness of the baseline proxy.

### A1 | Testing the sample selection

We start this part of the analysis by checking whether the estimated parameters are robust to changes in the estimation sample. On the one hand, we are concerned that big projects in the upper control group could be incorrect controls and drive the results and, on the other hand, there exist possible sources of contamination due to the unobserved engineers' value. Since we do not observe the project's ex-ante valuation, we may misclassify part of the projects to the treatment/control groups.<sup>57</sup> Both concerns are particularly worrying for our methodology, and we want to test our findings against both identification threats. In Table A1, Column 1 reports as a benchmark estimates (i.e., Table 2 Column 5); Columns 2 and 3 report the *Range* and *Sanitary* models, which restrict the sample to all projects valued less than \$500,000 and exclude observations in a 10% window around the \$100,000 and \$150,000 thresholds, respectively, to rule out observations being the most likely source of contamination. Large contracts, that is, those above \$500,000, can be fundamentally different from our mid-size treated contracts and not comparable. Including large contracts can easily add noise to the analysis and outweigh any true pattern of this study.

TABLE A1 Robustness check: Sample selection

Panel (a): Time Performance					
	(1) Base	(2) Range	(3) Sanitary	(4) Hperf	(5) Lower
$\hat{\theta}_{pb}$	-0.088*** (0.023)	-0.062** (0.030)	-0.099*** (0.027)	-0.098*** (0.021)	0.013 (0.024)
<i>N</i>	120,175	70,866	100,333	116,223	27,073
<i>R</i> <sup>2</sup>	0.128	0.107	0.127	0.130	0.144
Avg <sub>services</sub>	0.808	0.788	0.785	0.832	0.861
Avg <sub>works</sub>	0.837	0.784	0.811	0.843	0.894
Panel (b): Cost Performance					
	(1)	(2)	(3)	(4)	(5)
$\hat{\theta}_{pb}$	-0.035*** (0.013)	-0.051*** (0.016)	-0.024 (0.016)	-0.037*** (0.012)	-0.079*** (0.015)
<i>N</i>	120,175	70,866	100,333	116,116	27,073
<i>R</i> <sup>2</sup>	0.218	0.197	0.219	0.191	0.207
Avg <sub>services</sub>	0.868	0.852	0.844	0.878	0.920
Avg <sub>works</sub>	0.945	0.924	0.932	0.945	0.971

Note: Results of various DDD models of *Time*—panel a)—and *Cost*—panel b)—dimensions on public oversight and performance bonding treatment indicators plus controls for firm size, tender type, whether the firm is a limited liability company, whether the contract was signed during the last week of the FY, agency, FY, state, and object fixed effects. *Base* reports the baseline model results; *Range* reports results of the above model run on a sample of projects trimmed at a face value of \$500,000; in *Sanitary* we account for the possible contamination at the \$100,000 and \$150,000 thresholds dropping all projects whose face value lies within a 10% band around; *HPerf* indicates the exclusion of projects with performance lower than 0.1 *Lower* is run with the *lower control group*. In each panel, Avg<sub>services</sub> and Avg<sub>works</sub> report the average outcome in the Service and the Public Works treatment group, respectively. Standard errors are two-way clustered at the object and FY level.

\*\**p* < .05; \*\*\**p* < .01.



In Column 4, we take care of possible outliers in terms of outcomes and exclude projects associated with overruns worth nine times the expected contract values (i.e., performance lower than 0.1) in terms of cost and time, separately. In Column 5, we show the results obtained by running model (1) with the lower control group (i.e., all projects valued between \$50,000 and \$100,000). The number of observations is reduced due to the lower number of control contracts. Results are qualitatively and quantitatively comparable (i.e., contained within the 95% confidence interval of the corresponding baseline estimates), although not significant for *cost performance*. Our results prove to be solid against all the above concerns, with quantitatively and qualitatively comparable estimates with respect to the baselines'.

## A2 | Subsampling

In Table A2, we present the results of our baseline model for  $\hat{\theta}_{pb}$  when replicated in further subsamples of our population of contracts. Column 1 reports the baseline estimates benchmark estimates (i.e., Table 2 Column 5) as a benchmark. In Column 2, we run the baseline model on a narrower time window (2008–2013), reducing the risk that long-term trends in the data possibly drive our results.

Moreover, we want to check whether an additional service-only specific treatment drives our results, mainly concerning the *public oversight* estimates. Hence, we modify the DDD model in Equation (1) to

$$y_{ijt} = \alpha + \beta_1 Waiver_{it} + \beta_2 Post_{it} + \theta_{public}(Waiver_{it} \times Post_{it}) + \beta_3 Constr_i + \beta_4 (Constr_i \times Waiver_{it}) + \beta_5 (Constr_i \times Post_{it}) + \theta_{pb}(Constr_i \times Waiver_{it} \times Post_{it}) + \theta_{pc} T_{it}^{pc} + \gamma X_{it} + \delta_t + \zeta_j + \varepsilon_{ijt}, \quad (A1)$$

where  $T_{it}^{pc}$  is a binary treatment variable active for nonconstruction-only contracts above the threshold. Results of the estimation are reported in column 3: parameter  $\hat{\theta}_{pc}$  (unreported) is not statistically significant.

We are also concerned that contractors might be able to anticipate the October 2010 bylaw. In particular, potential contractors can react to the upward revision of the SAT in two opposite ways, namely by opting to participate less

TABLE A2 Additional robustness checks

Panel (a): Time Performance						
	(1) Base	(2) 2008–2013	(3) PubCheck	(4) No2010	(5) NoA	(6) underBudget
$\hat{\theta}_{pb}$	−0.088*** (0.023)	−0.072** (0.035)	−0.096*** (0.025)	−0.096*** (0.023)	−0.104*** (0.023)	−0.078*** (0.022)
<i>N</i>	120,175	75,453	120,175	104,752	103,520	136,373
<i>R</i> <sup>2</sup>	0.128	0.118	0.128	0.131	0.129	0.114
Panel (b): Cost Performance						
	(1)	(2)	(3)	(4)	(5)	(6)
$\hat{\theta}_{pb}$	−0.035*** (0.013)	−0.045** (0.018)	−0.043*** (0.015)	−0.034** (0.013)	−0.043*** (0.013)	−0.043*** (0.013)
<i>N</i>	120,175	75,453	120,175	104,752	103,520	133,787
<i>R</i> <sup>2</sup>	0.218	0.211	0.218	0.219	0.209	0.132

Note: Results of various DDD models of *Time Performance*—panel (a)—and *Cost Performance*—panel (b)—on public oversight and performance bonding treatment indicators plus controls for firm size, tender type, whether the firm is a limited liability company, whether the contract was signed during the last week of the FY, agency, FY, state, and object fixed effects. *Base* reports the baseline model results; in *Contamin*, we account for the possible contamination at the \$100,000 threshold dropping all projects whose face value lies between \$90,000 and \$110,000; *2008–2013* is self-explanatory and *Public Check* reports results of a modified model with a specific treatment dummy for public on-site monitoring. In model *No2010* we drop all contracts awarded in 2010, to take into account possible treatment anticipation effects, in *NoA* we exclude all *R & D* contracts (contract object A); in *outScope*, we control for the number of out-of-scope amendments related to the contract and, finally, in *underBudget*, we include projects concluded ahead of time and under budget. In each panel,  $Avg_{services}$  and  $Avg_{works}$  account for the average outcome in the Service and the Public Works treatment group, respectively. Standard errors are two-way clustered at the object and FY level.

\*\**p* < .05; \*\*\**p* < .01.

(more) right before the reform and to invest (divest) in her business, depending on her quality. Consider the firm  $i$ , whose productivity level lies below the median—that is, a relatively bad potential contractor. It anticipates the upward threshold revision and learns her increased likelihood of winning a contract in the \$100,000–\$150,000 range after the reform. Hence, it is in her best interest to *decrease* her participation in federal tenders before the bylaw update to avoid workload and congestion once the reform comes into force. For opposite reasons, good types are willing to participate more and counterbalance this effect. A second potential effect involves good types only. To not face competition by bad types, they are incentivized to scale up their business, participate in more valuable tenders, still subject to performance bonding, and take full advantage of their superior quality, passing through lower premia. Although contractors at the margin might have suspected the revision and behaved accordingly, such anticipatory effect has already been shown not to be long-lasting, as suggested by Castellani et al. (2018) and Palguta and Pertold (2017). To rule out this possibility, in Column 4, we drop FY 2010 contracts—that is, the 12-months period before the reform—from the analysis to get rid of most of the short-lasting anticipations, and the results hold.

In Column 5, we re-run the baseline estimation excluding all R&D contracts from the control group due to the very peculiar nature of those projects and their nonstandard outcome assessment measures (see, for instance, Decarolis et al., 2021). In Column 6, we consider the out-of-scope amendments in the computation of our outcome variable—as defined in Section 5.2—and the results are again robust with respect to the baseline's. Finally, in Column 7, we include in the sample all projects concluded ahead of time and under budget: despite their relatively low number, we were concerned that their exclusion could drive the results of our linear model specification. Point estimates, instead, are extremely robust as well as for the other dimension we explore.

### A3 | Alternative outcomes

In Table A3, we test the robustness of the results to the choice of outcome variables, both for time- (columns 1–3) and cost-related measures (4–6). Model *Base* reports the baseline convex  $performance_{it}$  measures, while in the *Relative* model, we use the ratio between the raw measures of overrun (in days and \$ for time and costs, respectively) and the initial schedule/budget, that is,  $relative\_overrun_{it} = \frac{overrun_{it}}{expected\_outcome_{it}}$ . Finally, model *Overrun* features plain overrun metrics.

The main difference between the relative measures and the performances lies in their support. While the former is theoretically unbounded on the positive side (and they range 0–28 and 0–11 for delays and extra costs in our data, respectively), the latter ranges by construction from 0 to 1. To reduce the variable dispersion, we run the estimation a sample that trim out clear outliers. In particular, we exclude all contracts (3598) whose renegotiation amounts are unrealistic—that is, amounting to 10 times or more the initial expected outcome—on either time or cost dimensions. The interpretation of the sign differs too, as performance worsens when moving to zero, whereas the opposite holds true for relative measures. The results of *Relative* models are in line with the baseline, although noisier.<sup>58</sup> Moving to overrun measures, the interpretation of the results becomes harder due to the much higher variance. Indeed, model *Overrun* yields opposite results for delays and extra costs: while the former is consistent with the baseline results, though unbelievably large (amounting to 100+% of the average overrun value), the latter shows an opposite direction. This is due to the high share of extreme values, which characterizes the distribution of overrun—the kurtosis of the distribution of *cost overrun* amounts to 4.2—and drives the point estimate.

TABLE A3 Alternative outcomes

	Time			Cost		
	(1) Base	(2) Relative	(3) Overrun	(4) Base	(5) Relative	(6) Overrun
$\hat{\epsilon}_{pb}$	−0.088*** (0.023)	0.365*** (0.090)	117.815*** (22.033)	−0.035*** (0.013)	0.093* (0.051)	−2.34e+04*** (4,822.707)
$N$	120,175	116,577	116,577	120,175	118,512	118,512
$R^2$	0.128	0.088	0.020	0.218	0.130	0.143

Note: Results of the DDD regression of time—Columns 1–3—and cost-related—Columns 4–6—outcomes on public oversight and performance bonding treatment indicators, plus controls and fixed effects. *Base* is the baseline model on performance measures, *Relative* is the ratio between overrun and initially expected measures, and *Overrun* are self-explanatory. Standard errors are two-way clustered at the object and FY level.

\* $p < .1$ ; \*\*\* $p < .01$ .

Altogether, these results highlight a property of our baseline convex outcome variables, which are run on the working sample (i.e., with outliers). The problem with the relative specification is indeed the sensitivity to outliers and misreports of expected and actual outcomes, which can strongly bias the index in the tails of the distributions. When employing absolute overrun as outcomes, the magnitude of the standard errors grows substantially, to the point of affecting the statistical significance of the estimates. This point clearly shows the weakness of the linear contract outcome measures (both Relative and Overrun) to cope with outliers.

## APPENDIX B: TESTING THE “DIFFERENTIAL SHOCK HYPOTHESIS”

In this appendix, we provide a series of empirical exercises to rule out the possibility that our estimates of  $\theta_{pb}$  merely capture differential shocks in the construction industry. In fact, another concern in our setup is the possibility that a shock hit the construction industry in FY 2011—and lasted for a long period of time—and that the estimate of  $\theta_{pb}$  thus captures the time- and cost-dimension effects on contract outcomes of such differential shock instead of an actual treatment effect.

We run the estimations controlling for input prices on *all* contracts in the sample (i.e., including inputs for both constructions and non-constructions). The exercise involves three steps.

- *Link to NAICS.* First, we build a correspondence table linking procurement categories to the 4-digits NAICS codes of the same industry group. As shown in Table B1, there are cases in which multiple NAICS codes refer to the same object and others (e.g., A and B), which involve two distinct objects covering the same NAICS industry.
- *Data Collection.* Second, we gather data on the average value and the distribution of the single most crucial input for most services: labor. The *Bureau of Labor Statistics* in the US issues a yearly report (the Occupational Employment Statistics, or OES) containing employment and hourly wage estimates for over 800 occupations, further aggregated at the NAICS industry group level. We proceeded by collapsing OES data at the procurement category level: in Figure B1, the left panel, we plot the average hourly wage for each object in 2005–2015, indexing the base year to 100. Although all are upward trending, the dynamics differ dramatically across contract types: while some industry groups have experienced an increase in the labor cost of  $\approx 43\%$  (*F-Natural Resource Management*), in sectors like M (*Operation of Structure Facilities*), it only amounted to 6%. A similar, striking variation across contract types is highlighted in the right panel, which plots the overall distribution of hourly wages: these range from an average \$40 (A, B, and D—i.e., research and IT) to as low as \$10 (*Social*). Moreover, all distributions appear to be asymmetrical, with the mean always higher than the median and interquartiles ranging from \$38 to \$5. Like the case of constructions, in some industries, there are physical inputs whose prices potentially drive our results. To account for those cases, we collect data produced in the context of the *Producer Price Index* (PPI) program, 2005 to 2015: it measures the average change in the selling prices of domestic businesses. Typically, it covers the majority of commercial products; however, it is missing for several services. In Figure B2, we plot the monthly series for each procurement category.<sup>59</sup>
- *Analysis.* Third, we control for input prices through five different models, presented in Table B2. We add the average hourly wage per industry group ( $\bar{w}$ , Column 1), its distribution through the first and last deciles (unreported) and quartiles ( $\bar{w}^{25}$  and  $\bar{w}^{75}$ , respectively, in Column 2), the PPI (Column 3), and all the measures together (Columns 4 and 5). The estimates are extremely robust, with two exceptions: the performance bonding parameter loses statistical significance when the PPI enters the equation, alone and with the average wage—this being due to the drop in the number of observations to less than one-third of its original sample size.

TABLE B1 Contract object—NAICS code walkthrough

Object		NAICS	
Code	Description	Code	Description
A	R&D	5417	Scientific Research and Development Services
B	Special Studies	5417	Scientific Research and Development Services
C	Architect/Engineering services	5413	Architectural, Engineering, and Related Services
D	Information technology	5415	Computer Systems Design and Related Services
E	Structure Facilities	2362	Nonresidential Building Construction
–	–	5612	Facilities Support Services
F	Natural Resources Management	8133	Social Advocacy Organizations
–	–	9241	Administration of Environmental Quality Programs
G	Social	8134	Civic and Social Organizations
H	Quality Control	5416	Management, Scientific, and Technical Consulting Services
J	Maintenance of Equipment	8113	Commercial and Industrial Machinery and Equipment (except Automotive and Electronic) Repair and Maintenance)
K	Modification of Equipment	8113	Commercial and Industrial Machinery and Equipment (except Automotive and Electronic) Repair and Maintenance)
L	Technical Representative	4251	Wholesale Electronic Markets and Agents and Brokers
–	–	5419	Other Professional, Scientific, and Technical Service
M	Operation of Structure Facilities	5612	Facilities Support Services
N	Installation of Equipment	8812	Electronic and Precision Equipment Repair and Maintenance)
–	–	8113	Commercial and Industrial Machinery and Equipment (except Automotive and Electronic) Repair and Maintenance
P	Salvage	4231	Motor Vehicle and Motor Vehicle Parts and Supplies Merchant Wholesalers
–	–	4239	Miscellaneous Durable Goods Merchant Wholesalers
Q	Medical	6211	Offices of Physicians
–	–	6212	Offices of Dentists
–	–	6213	Offices of Other Health Practitioners
–	–	6214	Outpatient Care Centers
–	–	6215	Medical and Diagnostic Laboratories
–	–	6216	Home Health Care Services
–	–	6219	Other Ambulatory Health Care Services
R	Support (Administration)	5611	Office Administrative Services
S	Utilities & Housekeeping	2213	
–	–	5617	Services to Buildings and Dwellings
T	Printing, Maps, Publications	3231	Printing and Related Support Activities
U	Education & Training	6114	Business Schools and Computer and Management Training
–	–	6117	Educational Support Services
V	Transportation	4841	General Freight Trucking
–	–	4821	Rail Transportation
–	–	4831	Deep Sea, Coastal, and Great Lakes Water Transportation

TABLE B1 (Continued)

Object		NAICS	
Code	Description	Code	Description
-	-	4832	Inland Water Transportation
-	-	4851	Urban Transit Systems
-	-	4854	Interurban and Rural Bus Transportation
Y	Construction	2362	Nonresidential Building Construction
-	-	2371	Utility System Construction
-	-	2361	Residential Building Construction
Z	Maintenance	8113	Commercial and Industrial Machinery and Equipment (except Automotive and Electronic) Repair and Maintenance

Note: Mapping between procurement category, whose code and description are reported in Columns 1 and 2, respectively, and NAICS codes (3 and 4).

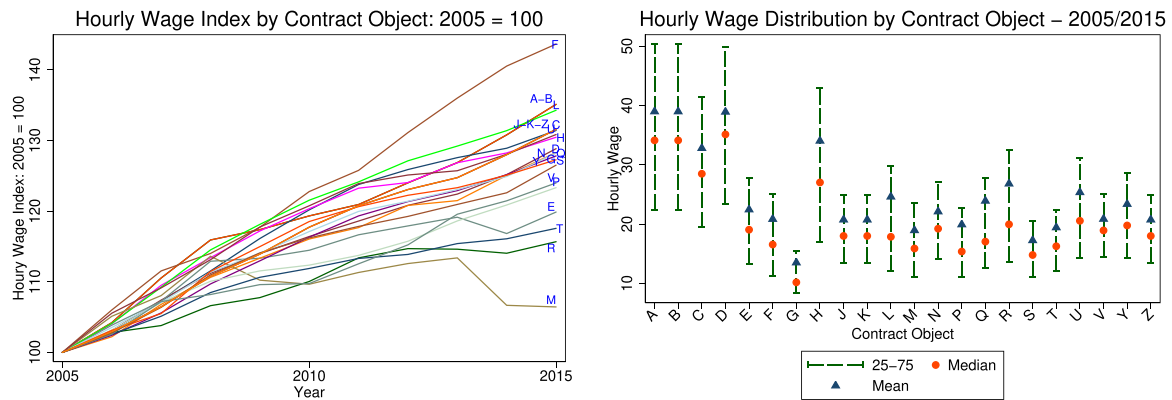


FIGURE B1 Occupation employment statistic-wages dynamics per procurement category. Hourly wages data from the Occupation Employment Statistic, by procurement category. Left panel: time series by procurement category, base year 2005 = 100. Right panel: distribution of wage by procurement category, in 2005 \$. Mean and median values are reported along with the interquartile range. [Color figure can be viewed at [wileyonlinelibrary.com](http://wileyonlinelibrary.com)]

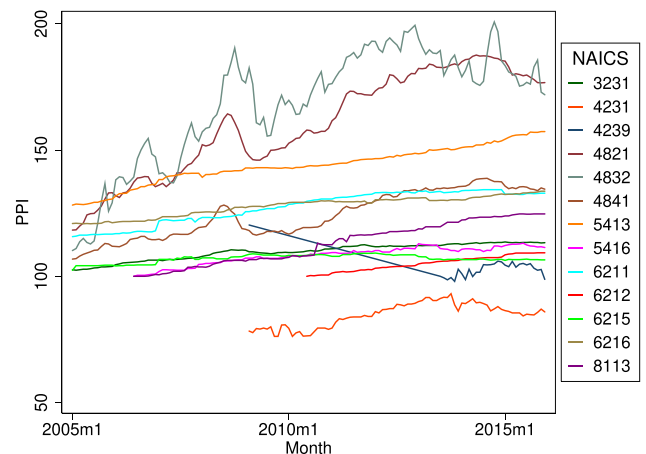


FIGURE B2 Producer price index-bureau of labor statistics. Evolution of Producer Price Index by NAICS code-monthly data from the Bureau of Labor Statistics. [Color figure can be viewed at [wileyonlinelibrary.com](http://wileyonlinelibrary.com)]



TABLE B2 Input costs table

<b>Panel (a): Time Performance</b>					
	(1) Wages – Mean	(2) Wages – Dist	(3) PPI	(4) PPI + Wage	(5) PPI + Wage dist
$\hat{\theta}_{pb}$	–0.090*** (0.023)	–0.094*** (0.023)	–0.082*** (0.031)	–0.082*** (0.031)	–0.084*** (0.031)
$\bar{w}$	0.008*** (0.003)	0.055** (0.025)		0.002 (0.008)	–0.066 (0.064)
$\bar{w}^{25}$		0.104*** (0.031)			0.271*** (0.050)
$\bar{w}^{75}$		–0.053*** (0.014)			–0.013 (0.043)
PPI			0.002* (0.001)	0.002** (0.001)	0.000 (0.001)
<i>N</i>	120,146	120,146	33,402	33,402	33,402
<i>R</i> <sup>2</sup>	0.128	0.130	0.146	0.146	0.148
<b>Panel (b): Cost Performance</b>					
	(1)	(2)	(3)	(4)	(5)
$\hat{\theta}_{pb}$	–0.037*** (0.013)	–0.038*** (0.013)	–0.047*** (0.017)	–0.052*** (0.017)	–0.054*** (0.017)
$\bar{w}$	0.009*** (0.003)	0.039* (0.022)		0.028*** (0.004)	0.105*** (0.039)
$\bar{w}^{25}$		0.011 (0.025)			0.038 (0.034)
$\bar{w}^{75}$		–0.028** (0.013)			–0.072*** (0.025)
PPI			–0.000 (0.001)	0.000 (0.001)	–0.001* (0.001)
<i>N</i>	120,146	120,146	33,402	33,402	33,402
<i>R</i> <sup>2</sup>	0.218	0.219	0.143	0.146	0.147

Note: Results of various DDD models of *Time Performance*—panel (a)—and *Cost Performance*—panel (b)—on a performance bonding treatment indicator plus controls for firm size, tender type, whether the firm is a limited liability company, whether the contract was signed during the last week of the FY, and agency, FY, state, and procurement category fixed effects. In Column 1, we control for input costs using the average wage per NAICS code; in Column 2, we add the first and last deciles (unreported) and the first and last quartiles of the wage distribution. In Column 3, we use the Producer Price Index to control for input prices; Column 4 adds the average wage, and Column 5 features all controls at once. In each panel,  $Avg_{services}$  and  $Avg_{works}$  account for the average outcome in the Service and the Public Works treatment groups, respectively. Standard errors are two-way clustered at the object and FY level.

\* $p < 0.1$ ; \*\* $p < .05$ ; \*\*\* $p < .01$ .

## APPENDIX C: ECONOMETRIC ISSUES

### C1 | Identification: Public oversight in constructions

If the effect of public oversight in construction contracts differs from that in other sectors, our model would yield biased estimates of  $\theta_{pb}$ .<sup>60</sup> To ensure that our estimates are not picking up differential effects of red tape, we identify a sample of construction contracts subject to public oversight but likely not to the performance bonding requirement. More specifically, we resort to the contracts related to projects performed abroad (which we omit in the main sample).

TABLE C1 Overseas constructions

	Time Performance					Cost Performance				
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
<i>Public oversight</i>										
$\hat{\theta}_{public}$	0.070*** (0.027)	0.082*** (0.028)	0.079*** (0.028)	0.076*** (0.027)	0.076*** (0.027)	0.028*** (0.011)	0.021* (0.011)	0.026** (0.012)	0.024** (0.011)	0.025** (0.011)
N	29,966	29,924	29,876	29,866	29,866	33,065	33,020	32,969	32,956	32,956
R <sup>2</sup>	0.024	0.030	0.046	0.142	0.142	0.008	0.014	0.035	0.062	0.063
Fiscal year FE	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Controls		✓	✓	✓	✓		✓	✓	✓	✓
Bureau fixed effects			✓	✓	✓			✓	✓	✓
Country fixed effects				✓	✓				✓	✓
Object fixed effects					✓					✓

Note: Results of various DD models of *Time Performance*—Columns 1–5—and *Cost Performance*—Columns 6–10—on public oversight for constructions performed outside the US territory (i.e., subject to public oversight, but potentially not exposed to the performance bonds). We add iteratively FY fixed effects, controls for firm size, tender type, whether the firm is a limited liability company, whether the contract was signed during the last week of the FY, and agency, country, and object fixed effects. Standard errors are two-way clustered at the object and FY level.

According to the FAR, such projects above the threshold *may be* waived from the issuance of performance bonds, whereas they still have to be subject to public oversight. However, there is no recorded information on whether such contracts have been waived from the performance bonding or not. For this reason, our identification assumption relies on anecdotal evidence provided by the GAO (<https://www.gao.gov/assets/gao-17-683r.pdf>): “Officials from State and NAVFAC confirmed that they have used waivers for overseas construction contract [...] Decisions on waivers, they told us, are based on the risk, visibility, or complexity of the contract.” Hence, even though the exercise that we propose below is the nearest possible to the one requested—given the data constraints—it is exposed to two sources of bias. On the one hand, we pool together contracts with and without performance bonding, with no way to control for the latter—that is, we estimate  $Y_{it} = \alpha + \theta_{public}(treat\_group_{it} \times post_{it}) + \beta_1 treat\_group_{it} + \gamma X_{it} + \delta_t + \eta_{it}$ , where  $\eta_{it} = \theta_{pb}(treat\_group_{it} \times post_{it} \times not\_waived\_it) + \varepsilon_{it}$  reflects the fraction of contracts that are actually subject to the performance bonding (omitted variable bias). On the other hand, the decision to waive is a function of contract complexity, and therefore, of the potential gains from performance bonds—which generates reverse causality.

Despite the above issues, we believe the results of the exercise to be extremely informative because they directly address one of the main concerns on our proposed identification: despite the fraction of contracts subject to performance bonding—whose effect is partially captured by the estimate—the parameters of  $\hat{\theta}_{public}$  reported in Table C1 are positive, significant, and report a magnitude comparable to our baseline estimate. This result is robust to the inclusion of fixed effects and several controls: we take this as a further confirmation of the fact that the effect of public oversight in construction is in line with (i) all other sectors as shown in our baseline exercise, and mostly with (ii) the results of Calvo et al. (2019) and Carril (2021).

## C2 | Identification: Public oversight and performance bonding interactions

The robust estimates of  $\theta_{public}$  in constructions reported above, in line with our baseline and with previous contributions, help us address another identification concern related to the possibility that interactions between public oversight and performance bonding might drive our estimates. In the absence of exogenous variations in the extent of interactions between performance bonding and public oversights, we are unable to directly estimate the extent of such interactions; however, we can characterize them to investigate the magnitude and direction of resulting bias.

Assume that the overall treatment effect is a combination of performance bonding and public oversight plus their interaction (e.g.,  $ATT = \theta_{public} + \theta_{pb} + \theta_{publicpb}$ , where  $\theta_{publicpb}$  is the parameter of the interaction). Our estimates show that the ATT of *waiving* from oversights in constructions is  $\approx 0$  irrespective of any interaction, while we found that  $\theta_{public}$  is positive (see previous section); therefore,  $\theta_{pb} + \theta_{publicpb} < 0$  is the effect that we are capturing. In our baseline model, we assume  $\theta_{publicpb} = 0$ , but it need not necessarily be the case. In fact, performance bonds might be *substitutes* for public oversight, in a way that public procurers, when dealing with surety-backed contracts, restrain from

TABLE C2 Estimated parameters bias under different scenarios of public oversight-performance bonding interaction

	Substitutes (+)	Substitutes (–)	Complements
$\theta_{publicPb} = 0$	✓, unbiased	✓, unbiased	✓, unbiased
$ \theta_{pb}  >  \theta_{publicPb} $	✓, up	✓, down	✓, up
$ \theta_{private}  <  \theta_{publicPb} $	✓, up	X	✓, up

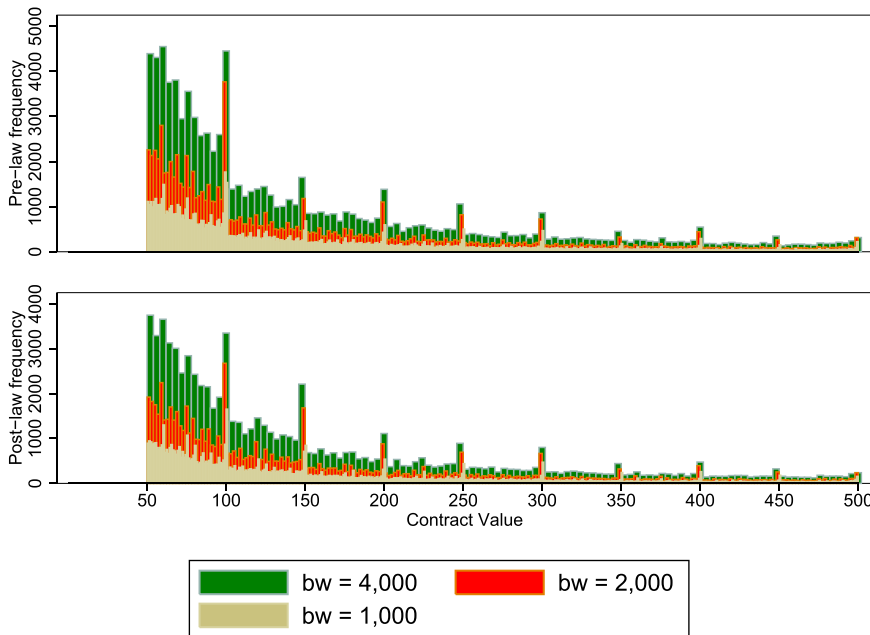


FIGURE C1 Contract frequency: various binsizes. Contract frequencies pre- and post-reform, absolute values. The histogram is plotted for different bin widths: \$4000 (green), \$2000 (red) and \$1000 (yellow). [Color figure can be viewed at [wileyonlinelibrary.com](http://wileyonlinelibrary.com)]

performing their otherwise assigned tasks; on the other hand, the two may be *complements*: in this scenario, surety companies interact with public procurers in ensuring contract completion. Asymmetrically, the former effect can be either positive or negative, depending on sureties being worse/better than procurers at performing the public-oversight tasks; the latter instead is surely performance-enhancing as long as, if that were not the case, it would be optimal for all parties to simply stop interacting ( $\theta_{publicPb} = 0$ ) at no cost.

We are left with three possibilities for each case, as shown in Table C2: (i)  $\theta_{publicPb}$  has, in fact, no real effect on contract performance—in which case, our estimates are unbiased, irrespective of the interaction type; (ii) the “direct” effect ( $|\theta_{pb}|$ ) is greater, in magnitude, than the interaction ( $|\theta_{publicPb}|$ )—hence, our estimates capture an upper bound of the performance bonding effect (substitutes +, complements) or a lower bound (substitutes –), but crucially, the sign is always correct; (iii)  $|\theta_{private}| < |\theta_{publicPb}|$ : in this case, our baseline model still captures the correct sign for substitutes + and complements, but appears to be utterly wrong in the case of substitutes –. The characterization of such a case, however, helps assess the likelihood of its occurrence: in other words, it means that the sureties, on top of their selection and monitoring duties, perform the tasks generally assigned to the public procurers (*substitution*), and do it so badly (the “minus”) as to more than offset their direct effects. The above case is also at odds with a simple argument on the incentives of public officials: even if delegating public oversight tasks to sureties might be beneficial to them (because they would exert lower effort), the contract performance deterioration would reflect, say, in worse career prospects, to the point that it would make it more convenient just to keep the public oversight tasks in the first place—and hence, reverting to a  $\theta_{publicPb} = 0$  situation.

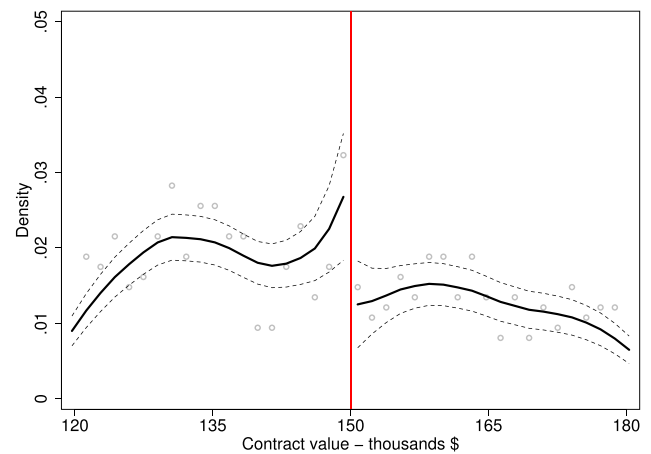
To summarize, based on our knowledge of practitioners' reports and the letter of the regulation, we assume no interaction between the public and the performance bonding in our baseline model. We acknowledge, however, that we cannot rule out the existence of some interaction, nor can we estimate it with the existing data at our disposal. Nonetheless, we can show that even if such interaction effect exists, our model correctly captures the sign of the performance bonding effect, except for very unlikely scenarios.

### C3 | Further discussion on the methodology

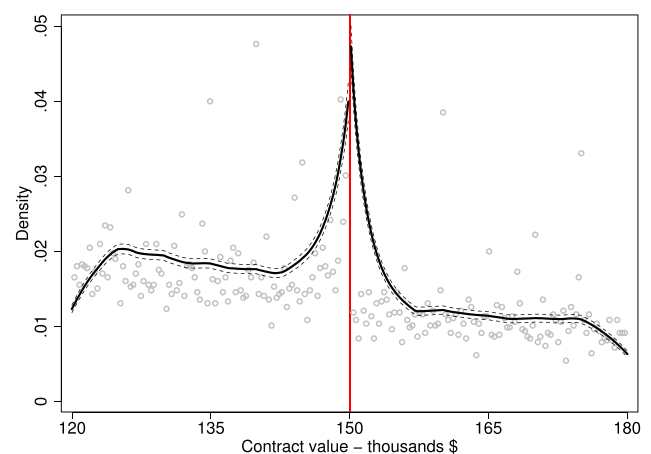
The previous research exploiting bylaws in public procurement to inform an identification strategy often implements a regression discontinuity analysis to analyze whether contract-allocation procedures with different levels of discretion to the procurement officer affect procurement outcomes. As shown in Figure C1, in our data, most of the “round” values—that is, 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 contract values. As a result, due to this likely endogenous sorting of contracts both before and after October 2010, the performance of firms above and below the bylaw may have been different even before the reform. Had we ignored this and performed a regression discontinuity analysis, we could have retrieved severely upward biased estimates due to a pre-existent discontinuity.<sup>61</sup>

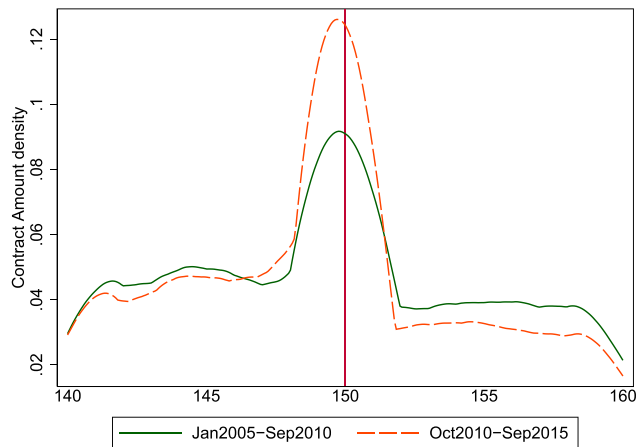
To test for endogenous sorting or discontinuities in the forcing variable, we performed the McCrary (2008) density test for post-law data for both construction (Figure C2) and services (Figure C3). The sharp discontinuity of the running variable at the \$150,000 threshold, highlighted by the graph and confirmed by the significant test results, rules out any possibility of running a usual regression discontinuity analysis 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 Posttreatment period (as Figure C4, displaying the pre- and post-reform contract value density around the \$150,000 threshold, displays). This is confirmed by Figure C5, showing that the pre-reform contract value density around the \$100,000 threshold is higher than the post-reform.

**FIGURE C2** McCrary density test—Construction contracts. McCrary density in a  $\pm\$30,000$  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. Construction contracts only. [Color figure can be viewed at [wileyonlinelibrary.com](http://wileyonlinelibrary.com)]

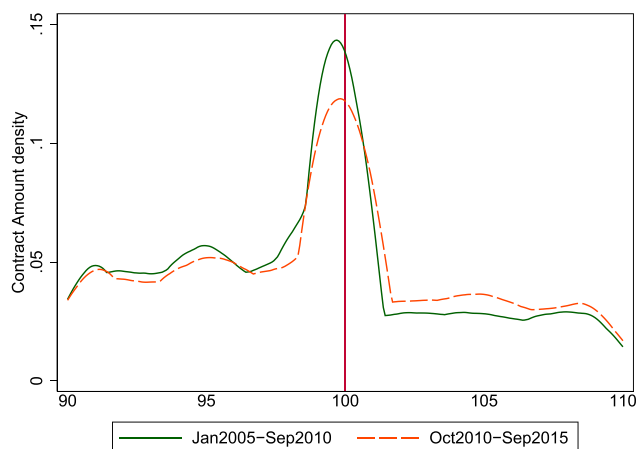


**FIGURE C3** McCrary density test—Full sample. McCrary density in a  $\pm\$30,000$  window around the \$150,000 threshold. The dots represent the density of projects in different intervals of project budget, the solid line represent a kernel estimate of the density, and the two dashed lines are 95% confidence intervals. [Color figure can be viewed at [wileyonlinelibrary.com](http://wileyonlinelibrary.com)]





**FIGURE C4** Pre- and post-reform contract value density. Contract amount density in a  $\pm\$20,000$  window around the \$150,000 threshold before (solid) and after (dotted) the threshold revision. [Color figure can be viewed at [wileyonlinelibrary.com](http://wileyonlinelibrary.com)]



**FIGURE C5** Pre- and Post-reform contract value density—100,000. Contract amount density in a  $\pm\$20,000$  window around the \$100,000 threshold before (solid) and after (dotted) the threshold revision. [Color figure can be viewed at [wileyonlinelibrary.com](http://wileyonlinelibrary.com)]

#### C4 | Discussion on clustered standard errors

Abadie et al. (2017) show that two things matter for clustering: whether there are clusters in the population of interest that are not represented in the sample (i.e., *sampling design issue*) and whether clusters of units rather than individual units are assigned to a treatment (i.e., *experimental design issue*). On the one hand, the FPDS sample is not selected because it amounts to the universe of contracts, so we do not have any sampling concerns. On the other hand, the clustering of standard errors should take the possible experimental design issue into account since, by design, the intertwined treatments are assigned depending on the object (construction or nonconstruction), conditional on the year of the award, rather than at the treated unit level (i.e., contract). In other words, the procurement category determines the assignment to the treatment/control groups, while the timing indicates the actual treatment allocation.

For the sake of our application, it is worth extending the discussion on clustering to DD-like applications. The DD formulation includes fixed effects for cross-sectional units and, for this reason, the DD allows for the selection for treatment based on time-invariant unobserved characteristics. Practitioners typically apply cluster-robust estimators to account for residual correlation within-group. Although it is common practice in DD to use cluster-robust inference procedures (Bertrand et al., 2004; Wooldridge, 2003), we note that the decision whether or not to cluster is usually not obvious and depends on the type of uncertainty one is trying to consider (Callaway & Sant'Anna, 2021). As a result, there is currently no consensus on how to cluster standard errors with DD applications and their extensions, like our DDD. Our approach in this sense is the following: since procurement processes follow predetermined policies and funds allocation, the observations on procurement categories and years are not independent pieces of new information—thus, we must assume that observations within a procurement category are likely to be correlated across time. Hence, the two-way clustering of standard errors at the object and year level addresses the sampling design issue and is consistent with our proposed empirical identification.



## APPENDIX D: FURTHER DESCRIPTIVE TABLES AND THE PERFORMANCE BOND BLANK FORM

- Tables D1 and D2 display the number of projects by department and year before and after the reform when considering the upper and the lower control group, respectively, for the analysis.

TABLE D1 Number of projects by department and FY - Upper control and treatment groups

	Pretreatment						Posttreatment				
	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015
Dept of Agriculture	76	42	58	57	75	153	145	118	85	110	12
Dept of Commerce	635	639	605	660	948	1187	1073	862	857	666	121
Dept of Defense	336	380	404	502	330	384	368	441	307	265	26
Dept of Education	12	17	18	19	27	37	15	26	10	7	1
Dept of Energy	8	13	24	17	19	48	34	38	23	20	5
Dept of Health and Human Services	132	324	462	971	989	1422	717	147	21	10	3
Dept of Homeland Security	23	20	17	13	13	24	13	14	8	17	3
Dept of Housing	12	13	17	24	34	137	100	112	86	94	17
Dept of Interior	462	469	497	398	656	1118	525	482	361	411	146
Dept of Justice	278	286	394	332	365	532	481	387	369	392	44
Dept of Labor	1262	3218	4792	4862	4807	4649	3821	3716	2398	2104	302
Dept of State	41	60	59	65	57	44	50	41	35	36	8
Dept of Treasury	109	101	77	106	105	127	88	104	76	105	15
Dept of Transportation	1209	1156	1036	991	1095	1205	1160	1035	899	844	105
Dept of Veteran Affairs	843	846	1197	1206	1170	1097	1248	1249	1029	887	191
Environmental Protection Agency	47	48	55	44	49	111	48	18	23	12	1
General Services Administration	410	789	736	737	1047	1602	1116	963	896	902	115
NASA	357	374	698	679	991	1469	1812	2000	2010	1993	582
Nuclear Regulatory Commission	144	45	74	67	78	165	148	142	143	129	26
National Science Foundation	22	107	129	187	247	379	405	382	302	252	76
Office of Personnel Management	306	400	442	331	346	397	362	418	291	269	39
Small Business Administration	213	437	726	907	840	1050	1048	1019	836	656	127
Social Security Administration	749	596	787	1227	1963	3354	3223	2972	2823	2325	958
<i>N</i>	7686	10,380	13,304	14,402	16,251	20,691	18,000	16,686	13,888	12,506	2923

Note: Number of projects by FY/department; Upper control group and treatment group. Pretreatment period: January 2005 to October 2010; posttreatment periods: October 2010 to September 2015.

TABLE D2 Number of projects by government department and FY—Lower control and treatment groups

	Pretreatment						Posttreatment				
	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015
Dept of Agriculture	82	70	73	60	124	159	224	157	124	153	17
Dept of Commerce	440	426	367	439	517	706	665	558	581	518	149
Dept of Defense	231	250	293	299	232	209	201	222	222	195	34
Dept of Education	5	8	10	16	17	43	22	27	12	10	2
Dept of Energy	1	5	8	8	15	28	12	24	19	21	4
Dept of Health and Human Services	443	597	851	1470	1476	1314	434	101	31	16	5
Dept of Homeland Security	21	15	23	13	22	30	12	15	12	10	3
Dept of Housing	8	18	23	26	26	265	182	178	125	83	10
Dept of Interior	422	507	530	492	738	1,037	587	619	401	432	348
Dept of Justice	380	295	301	367	381	443	413	371	316	289	43
Dept of Labor	551	1914	3441	3451	3342	3364	2890	2716	1885	1780	356
Dept of State	11	15	21	19	14	30	21	16	14	21	3
Dept of Treasury	93	97	56	49	53	56	83	71	61	69	14
Dept of Transportation	890	933	909	903	953	837	724	716	616	622	146
Dept of Veteran Affairs	542	467	596	594	551	477	601	598	509	485	156
Environmental Protection Agency	18	31	48	30	30	38	9	8	9	8	
General Services Administration	304	759	853	980	1117	1323	1020	861	770	768	149
NASA	193	240	395	452	781	1100	1500	1431	1399	1790	656
Nuclear Regulatory Commission	100	46	58	56	58	104	105	83	70	80	19
National Science Foundation	24	101	133	165	209	361	301	270	203	161	68
Office of Personnel Management	190	200	229	204	243	252	261	194	147	126	28
Small Business Administration	77	213	549	527	490	716	693	549	504	356	65
Social Security Administration	573	586	673	981	1728	2897	2757	2215	1933	1670	968
<i>N</i>	5599	7793	10,440	11,601	13,117	15,789	13,717	12,000	9963	9663	3243

Note: Number of projects by FY/department; Lower control group and treatment group. Pretreatment period: January 2005 to October 2010; posttreatment periods: October 2010 to September 2015.

- Figure D1 presents the composition of treatment and both lower and upper control groups in terms of the categories of service content of projects representing around 97% of observations. The distribution of categories is quite regular across groups. This exercise aims to show that the treatment group truly constitutes a representative sub-sample in our population of projects.
- Figures D2 and D3 report the blank module contractors need to fill out and convey to the federal buyer.

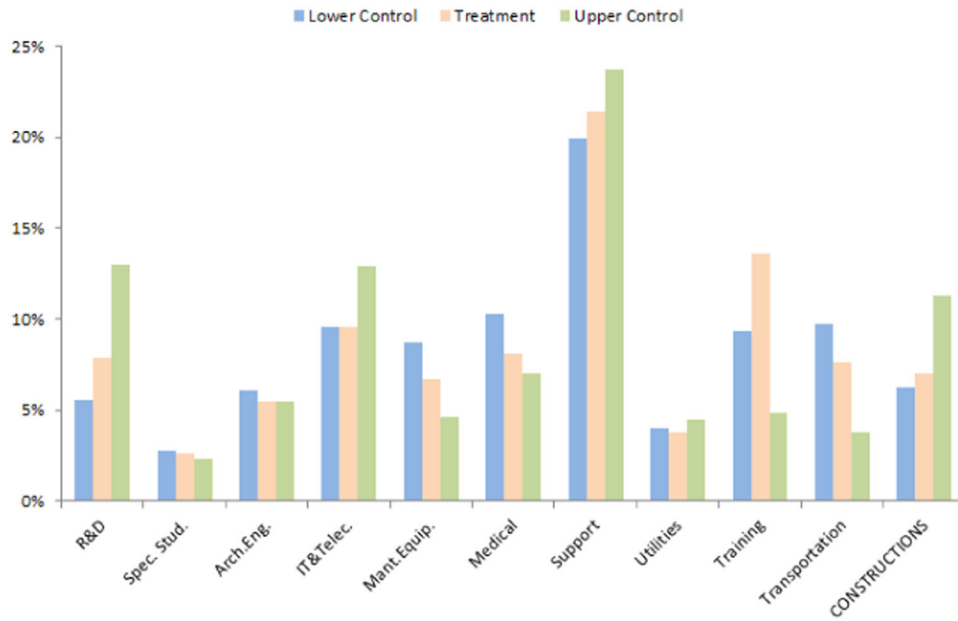


FIGURE D1 Categories of services by treatment group and lower/upper control groups. Number of projects by FY/department; Upper control group and treatment group. Pretreatment period: January 2005 to October 2010; posttreatment periods: October 2010 to September 2015. [Color figure can be viewed at [wileyonlinelibrary.com](http://wileyonlinelibrary.com)]

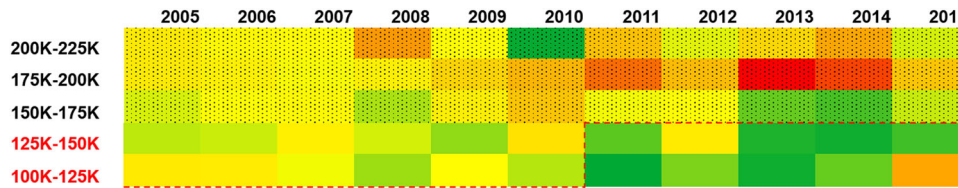


FIGURE D2 Performance bond (blank) form (FAR 28.106-1): Page 1 [Color figure can be viewed at [wileyonlinelibrary.com](http://wileyonlinelibrary.com)]

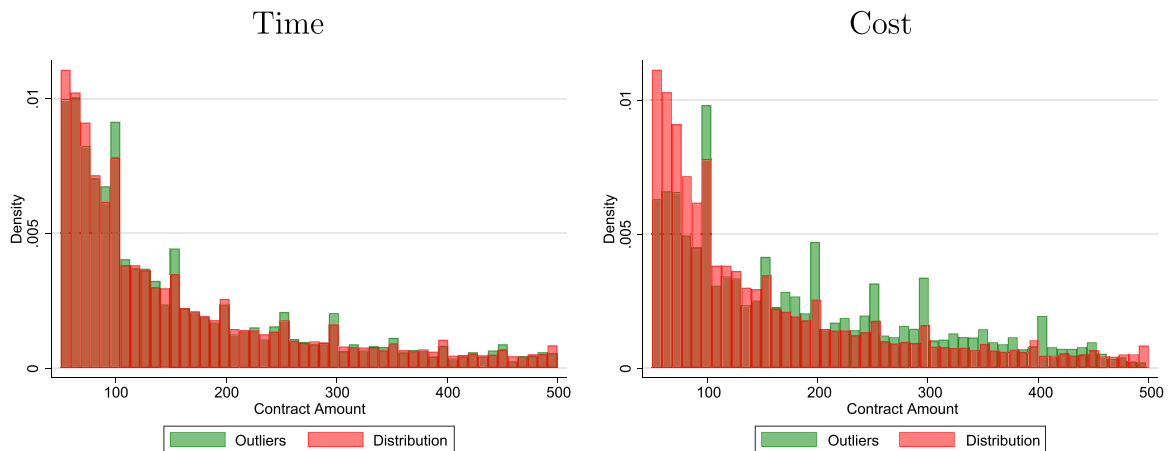


FIGURE D3 Performance bond (blank) form (FAR 28.106-1): Page 2 [Color figure can be viewed at [wileyonlinelibrary.com](http://wileyonlinelibrary.com)]