CESifo WORKING PAPERS

11142
2024

May 2024

Targeted Bidders in Government Tenders

Matilde Cappelletti, Leonardo M. Giuffrida

CESifo

Impressum:

CESifo Working Papers
ISSN 2364-1428 (electronic version)
Publisher and distributor: Munich Society for the Promotion of Economic Research - CESifo GmbH
The international platform of Ludwigs-Maximilians University's Center for Economic Studies and the ifo Institute
Poschingerstr. 5, 81679 Munich, Germany
Telephone +49 (0)89 2180-2740, Telefax +49 (0)89 2180-17845, email office@cesifo.de
Editor: Clemens Fuest
https://www.cesifo.org/en/wp
An electronic version of the paper may be downloaded
· from the SSRN website: www.SSRN.com
· from the RePEc website: www.RePEc.org
· from the CESifo website: https://www.cesifo.org/en/wp

CESifo *Working Paper No. 11142*

Targeted Bidders in Government Tenders

Abstract

A set-aside promotes a more equitable procurement process by restricting participation in government tenders to small or disadvantaged businesses. Yet its micro-effects on tender outcomes (competition and contract efficiency) and targeted firm performance entail trade-offs, which we evaluate empirically using a decade of US federal procurement data. At the tender level, we employ a two-stage approach. First, we use random forest techniques to compute the propensity score for a tender being set aside based on rules implementation. Second, we employ the scores in an inverse probability weighting framework. We find that set-asides prompt more competition—implying that the rise in participation of targeted firms more than offsets the exclusion of untargeted ones—and inefficiency, measured by cost overruns and delays. We argue that adverse selection and moral hazard are mechanisms behind contract inefficiency. We then study the targeted firm behavior to uncover whether long-run benefits mitigate short-run drawbacks. We compare businesses differentially exposed to a set-aside spending shock through an event study framework. We find mixed evidence on firm growth.

JEL-Codes: H320, H570, L250.

Keywords: set-aside program, public procurement, firm dynamics, random forest, propensity score, event study.

Matilde Cappelletti
University of Mannheim & ZEW
Mannheim / Germany
matilde.cappelletti@zew.de

*Leonardo M. Giuffrida**
ZEW Mannheim / Germany
leonardo.giuffrida@zew.de

*corresponding author

First version: July 13, 2022
This version: April 11, 2024
We would like to thank Filippo Biondi (discussant), Rodrigo Carril, Decio Coviello, Adriano De Leverano, Audrey Guo (discussant), Michael Lechner, Lucais Sewell, Giancarlo Spagnolo, and Janne Tukiainen for their helpful comments and suggestions as well as participants in the Spring Meeting of Young Economist 2022, the ECOBUS Seminar Series, the 47th Simposio de la Asociación Española de Economía-Spanish Economic Association (SAEe), the MaCCI Annual Conference 2023, the 79th Annual Congress of the International Institute for Public Finance (IIPF), the 50th Annual Conference of European Association for Research in Industrial Economics (EARIE), and the CRC Young Researchers Workshop 2023 in Montabaur, where earlier versions of this study were presented. The authors gratefully acknowledge financial support from the Leibniz SAW project "Market Design by Public Authorities." Cheng Chen, Julia Schepp, and Giovanni D'Ambrosio provided excellent research assistance.

"One of the great mistakes is to judge policies and programs by their intentions rather than their results."

Milton Friedman

I Introduction

Economic theory indicates that competition among firms generally leads to more efficient market outcomes. In the realm of public procurement, policymakers traditionally view competition as a means of limiting potential abuses of bureaucratic discretion, poor governance, and bid rigging. Thus, *a fortiori*, competition in public contracting is conceived as an effective instrument for increasing the value of taxpayers' money—making the lack of adequate competition a major concern in policy discussions (e.g., EC, 2017). In an endeavor to foster competition, regulations encourage open procurement auctions to boost firm participation.

Although encouraged, open auctions can backfire in efficiency due to limitations on contract enforcement and informational asymmetries (Hart and Holmstrm, 1986; Laffont and Tirole, 1993). Moreover, the playing field might not be leveled among potential suppliers, as large firms are typically more price competitive. This latter imbalance may inhibit smaller firms from securing public contracts and hinder their growth opportunities. Governments acknowledge this equity-related drawback inherent to open auctions and explicitly intervene with programs designed to support small businesses, given their relevance to the economy.[1]

In an attempt to nurture small firms, a government procurement regulation involves "setting aside" certain auctions for targeted businesses. For instance, small-business set-asides exclude large firms *in toto* from participating in specific tenders, which are set aside solely for small businesses (with no further designation). Set-asides are also used to achieve social objectives, such as promoting small *and* disadvantaged businesses (henceforth, disadvantaged businesses). An example of the latter is the set-aside for service-disabled veteran-owned small firms.

However, fostering equity through set-asides entails trade-offs on actual competition and efficiency in terms of contract outcomes and firm dynamics (see Section III). The empirical existing literature has yet to conclusively assess these dimensions (e.g., Athey et al., 2013; Denes, 1997; Marion, 2007, 2009; Nakabayashi, 2013; Tkachenko et al., 2019). In particular, unresolved issues include the generalizability of findings across heterogeneous procurements and set-aside types, as well as the implications for the execution stage of contracts and the long-term dynamics of winners. This paper aims to fill these gaps by analyzing short- and long-run effects of set-aside programs, concerning procurement and firm-related outcomes, respectively.

Our setting is US federal procurement from 2008 to 2018, where the government allocates more than a quarter of its budget through set-asides, ranging from \$50.1 billion to \$81.2 billion yearly over the period of our analysis. This setting is ideal for shedding light on the unanswered issues for three reasons. First, it encompasses different types of set-asides. Second, the database contains

[1] For instance, small businesses account for two-thirds of employment in OECD countries (OECD, 2019).

detailed information on procurement categories and contract execution stage, and tracks supplier firms over time. Third, it allows us to leverage a set-aside spending shock.

We evaluate the effectiveness of this regulation by first analyzing contract-level responses to set-asides at two different stages of the procurement process. At the award stage, we examine how set-asides influence the number of bids. Answering this question is not as straightforward as it may initially appear. By restricting the participation of untargeted firms (i.e., firms that are large or not disadvantaged), the set-aside provides incentives to place a bid to the targeted firms that would not have competed for the same contract in an open tender. The net response on the number of bids is thus an empirical question, depending on the number of targeted and untargeted businesses in each market and entry costs for targeted firms in both scenarios.

At the execution stage, we investigate whether the program affects efficiency, resulting in savings or additional costs for the government. The implications for this contract stage are ambiguous due to conflicting hypotheses regarding supplier selection and post-award behavior, including variations in bidder distribution and supplier incentives between set-asides and open tenders. We study the effect of set-asides on extra costs and delays as distinct metrics for contract (in)efficiency, following the related empirical literature (e.g., Decarolis, 2014; Hart and Moore, 1988; Herweg and Schwarz, 2018). The rationale is that public contracts are intrinsically rigid and any amendment captures Williamson (1971)'s transaction costs, which are economically sizable (Bajari et al., 2014; Beuve et al., 2019; Spiller, 2008).

For our contract-level analysis, we rely on the universe of service and construction contracts since they include an execution stage. We consider two treatment groups of targeted firms, i.e., small businesses (under the "small business set-aside" or SBSA) or all categories of disadvantaged businesses (which we pool under the "disadvantaged business set-aside" or DBSA). The control group consists of contracts that are not set aside (i.e., open procedures). Set-asides are used in the vast majority of sectors—defined as the most granular procurement category in our data—ensuring consistent comparisons in our analysis.

Although the Federal Acquisition Regulation (FAR) §19.502 mandates the set-aside of contracts below a certain dollar threshold (the "simplified acquisition threshold"), it also makes clear that the set-aside is ultimately at the discretion of the agency.[2] Indeed, the agency's decision to set aside ultimately hinges on two key factors. First, the agency will set-aside a contract upon a positive assessment of at least two potential "competitive bidders" among the targeted firm set (the so-called Rule of Two). The regulation requires contracting personnel to explicitly report the evaluation of market research by filling out *ad hoc* forms, which are typically reviewed by the internal small business office, which must approve them prior to acquisition.[3] Thus, the role of private or soft information in the decision is minimal. Second, the agency's decision is influenced on the stringency of its annual set-aside spending target. The attainment of this target is evaluated annually by the Small Business Administration, and each agency receives scorecards on the goal

[2]This is reflected in our data, where we observe that about 19% (30%) of contracts below the threshold are awarded through a small business (any type of) set-aside.

[3]Documentation and procedural requirements for small business considerations vary by agency. In the Defense acquisition, this research is documented in the DD 2579 Form (DFARS §219.201(10)(B)). In the appendix, we report a blank version of this form.

achievement, implying greater future oversight over the agency's procurement activities in case of non-compliance (Dilger, 2024). Thus, agencies are incentivized to comply. We employ a propensity score-based method that, exploiting the wealth of our data, approximates the agency's regulatory constraints and information set to predict each set-aside decision.

The agency retains some margin of discretion in implementing these rules. This bureaucratic discretion is not problematic yet an advantage for our approach for two reasons. First, the validity of the propensity score relies on the *common support* assumption, which requires that the treatment assignment cannot be deterministic (Caliendo and Kopeinig, 2008). This is true in our context due to discretionary decision-making. Second, discretion is the source of heterogeneity that leads to some contracts being treated while others, though similar, remain untreated (Gibson and McKenzie, 2014). However, such discretion does not depend on unobserved information since agencies must explicitly justify why they do or do not set aside a contract.

We estimate the propensity score of being set aside for each contract with a random forest (first stage)[4], and use an inverse propensity weighting (IPW) estimation to elicit the effect of set-asides on contract outcomes (second stage). The IPW removes the differences in key observable characteristics while giving more weight to those observations with a low probability of receiving the treatment—i.e., the tender being set aside. We argue that, in our setting, the IPW augmented with random forest estimates causally the impact of set-aside tenders on the procurement outcomes. We perform several exercises testing the identification assumptions that support our claim of causality.

We find that both SBSAs and DBSAs increase the number of bids by approximately one unit. Given that the control group receives an average of 3.6 bids, the effect of an additional bid in set-aside auctions indicates that the increase in competition extends beyond the simple mechanical enforcement of the Rule of Two. Hence, we find evidence that set-asides can level the playing field, inducing more competitive procurement overall (Alcalde and Dahm, 2024). Moreover, SBSAs increase cost overruns by 16.2% (or $54,820 per contract) and delays by 5.2% (or 9 days per contract). DBSAs are more disruptive, although not significantly so: Cost overruns and delays increase by 30.9% (or $104,800) and 8.9% (or 16 days), respectively. The results for execution-stage performance hold with alternative outcome variables based on adjustment costs.

We lay out two non-competing asymmetric-information channels through which set-aside tendering negatively affects contract performance. First, set-asides select targeted suppliers that may lack quality or experience (adverse selection). Second, set-asides may disincentivize suppliers to perform optimally, knowing they have repeated opportunities within the program (moral hazard). We provide suggestive evidence of adverse selection for SBSA and moral hazard for DBSA

[4]Specifically, in the first stage, we employ a machine learning approach for prediction of propensity scores (Goller et al., 2020; Lippmann, 2021). Among the different machine learning techniques, we choose the random forest for three main reasons. First, it requires only two tuning parameters and it is not sensitive to their specification. Second, it outperforms the traditional method for estimating the propensity score (i.e., the logit) in predicting the out-of-sample treatment (e.g., Lee et al., 2010). Third, it enables us to include a wealth of predictors, thereby fully exploiting the richness of our available data and achieving covariate balance. On this dimension, the random forest is again superior to the logit. We provide a battery of statistical tests corroborating these arguments, which confirm the validity of our identification approach. In Section V.1, we present recent examples in the literature of the adoption of random forest for propensity score estimation and further predictive purposes.

through additional exercises based on winner fixed effects regressions, alternative winner-level outcomes (e.g., incumbency status), and effect heterogeneity on procurements differentially exposed to screening and monitoring.

Shifting the focus to targeted firms, we explore whether the documented spending inefficiency is accompanied by improved firm's performance in the long run. Set-aside contracts provide targeted businesses with an opportunity to scale up their activities in line with the ultimate equity goal of the program: to allow targeted firms to grow and eventually compete equitably with other businesses. Yet the program could disincentivize targeted firms from scaling up insofar as this would cause them to lose favorable contract opportunities.

To study this trade-off, we exploit a massive, unexpected, and permanent increase in spending for a DBSA program—i.e., the service-disabled veteran-owned business set-aside—that occurred in 2009 to provide evidence on whether set-asides might be considered an industrial or subsidy policy. We use an event study approach that exploits differential prior business exposure to this specific set-aside program as a source of plausibly exogenous variation.

We find weak evidence that incumbents grow in scale after being exposed to the demand surge. Specifically, firms grow larger as they sell a broader selection of categories and experience a weak increase in out-of-state sales. However, they also become more dependent on the set-aside program in terms of the procurement amount secured. This suggests that if they grow in size, they restrict their growth to the maximum eligible size for continued set-aside opportunities. If they improve in competitiveness, they do so along other dimensions. While none of these firm-level results is a perfect test in isolation, we believe that in combination, they support the conclusion that the program, on top of generating short-term, contract-level inefficiency, is not associated with long-term, firm-level efficiency—at least when looking at business performance in the government procurement market and service-disabled veteran-owned set-asides.

Related literature This paper builds on three different strands in the literature. First, our work is related to the broad scholarship on the determinants of public procurement efficiency.[5] More specifically, we speak to the studies examining the relationship between competition and efficiency in procurement. Although economic theory indicates that more competition leads to better outcomes in equilibrium, empirical studies reveal mixed evidence on the monotonicity of such relationship due to the influence of perceived market competition on bidding behavior (e.g., Fischbacher et al., 2009). On the one hand, this stream of literature has mostly focused on buyer-induced competition level. Coviello et al. (2017) and Kang and Miller (2022) provide contrasting evidence to the common belief that open competition—and consequently low discretion—is always the best tool for boosting value for taxpayers' money. They find that increasing buyers' discretion on winner selection can have a positive effect on procurement outcomes. Calzolari and Spagnolo (2009) provide a theoretical background to explain these empirical results. They find that buyers optimally choose constrained auctions that threaten the exclusion of suppliers with poor past

[5] Examples include the award mechanism (Decarolis, 2018), end-of-year spending rush (Liebman and Mahoney, 2017), role of buyers (Decarolis et al., 2020, 2021; Best et al., 2023), external audits (Gerardino et al., 2017), industry consolidation (Carril and Duggan, 2020), and centralized purchasing (Bandiera et al., 2009).

performance when non-contractable quality is critical. When non-contractable quality is marginal, the public buyer optimally chooses open competition combined with the threat of switching to restricted competition to eliminate poor-performing suppliers. We augment this literature with evidence on the impact of set-aside programs on competition.

On the other hand, existing work explores the trade-off between competition and ex-post performance (Bajari et al., 2014; Decarolis, 2014; Spulber, 1990). In particular, Carril et al. (2022) show that competition's benefits and costs depend on the complexity of the purchase. Our work examines the trade-off between competition and procurement outcomes from a different perspective. Our results show that set-aside policies, which impose a mechanical constraint on participant type, are associated with an actual increase in participants from targeted potential entrants who offset the loss of excluded untargeted firms and a deterioration in execution-stage outcomes.

Second, our paper relates to studies that examine policies mechanically altering competition in public tenders. Such policies are grouped into two categories: bid preferences and set-asides.[6] Krasnokutskaya and Seim (2011) show that preference programs increase procurement costs. Nonetheless, tailoring the program to different project categories marginally increases costs since the effect is extremely heterogeneous. Marion (2007) and Marion (2009) also find a cost increase from such programs. Rosa (2019) finds that, under such programs, affiliation—i.e., the dependence between the bidder's cost and that of its competitors—makes procurement more expensive and reduces efficiency. In contrast, existing studies on set-aside programs provide mixed results. Using US Army Corps data from 1990-1991, Denes (1997) finds no evidence that set-asides for small businesses correlate with increases in the cost of government contracts. Athey et al. (2013) find that small business set-asides produce a decline in revenue and efficiency for the Forest Service. More recently, Tkachenko et al. (2019) provide evidence that set-asides induce lower award prices for homogeneous goods.

Our paper is closely related to Nakabayashi (2013). Using Japanese public construction contracts, the author finds that eliminating the program would lead to a counterfactual 40% decline in small business participation, with the negative effects of reduced competition outweighing the efficiency loss from small businesses winning contracts. Our results confirm this participation effect in a broader scope of US federal procurements. We add by showing an unexplored dimension of the inefficiency effect. While Nakabayashi (2013) finds an increase in government costs due to differential bidding by small businesses at the award stage, we contribute by showing that set-asides increase government costs at the execution stage due to adverse selection and moral hazard.

Third, this paper contributes to the empirical literature examining the impact of demand on firm performance (Foster et al., 2016; Pozzi and Schivardi, 2016). Focusing on public procurement demand, the current literature shows that firms exposed to a demand shock experience a persistent increase in revenue and employment (Fadic, 2020; Ferraz et al., 2015; Gugler et al., 2020; Lee, 2021). A positive shock also increases capital investment (Hebous and Zimmermann, 2021), facili-

[6] "Bid preferences in procurement auctions allow firms from an identifiable group an advantage in bidding against unfavored firms" (Marion, 2007). An example of bid preference is giving a bidding discount to small firms in auctions to increase their chances of winning. Notably, unlike set-asides, the untargeted firms are not excluded from competition *in toto*, but instead do not benefit from bid discounts.

tates access to external borrowing (di Giovanni et al., 2022; Goldman, 2019), promotes innovation (Czarnitzki et al., 2020), and prolongs survival (Cappelletti et al., 2022). Our work complements this empirical literature in two ways. First, we focus on procurement-specific firm outcomes, such as the delivery of multiple product variants to the government. Second, we also exploit a persistent yet unexpected surge in demand to study targeted firms' dynamics. Most of the works use identification strategies that mainly rely on temporary shocks, namely the random awarding of a contract. One exception is provided by Coviello et al. (2022), who exploit a legislative change permanently affecting the fiscal leeway of Italian municipalities. They find that businesses facing a persistent decline in public demand respond by cutting capital.

Despite this long list of contributions, the evidence in the literature on how set-asides affect the procurement process is inconclusive. We shed light on the following four dimensions, which remain unresolved to date. First, it is unclear whether the estimated effects can be generalized to a broader range of procurement categories. Second, relevant existing work focuses exclusively on ex-ante procurement outcomes such as the award amount. Third, there is no evidence on the effect of other types of set-asides.[7] Fourth, the long-term effects of such programs are unexplored. Hence, it is unclear whether the policy goal of leveling the playing field between small and large firms is achieved. We contribute to all of these unanswered dimensions. First, we use a large dataset on service and construction contracts from 11 years of government contracting across the US. Second, we evaluate the impact in terms of execution-stage contract outcomes. Third, we study more restrictive set-asides (i.e., the DBSAs). Fourth, we exploit the unexpected increase in spending for a particular set-aside program combined with an event study analysis to assess the long-term implications of this policy for winning firms.

The rest of the paper unfolds as follows. Section II provides the necessary institutional information on the US government's set-aside programs. In Section III, we outline a theoretical background to our empirical research. Section IV presents and describes the data. We pool the contract-level analysis in Section V. In particular, the related identification concerns and strategy are explained in Section V.1, while findings are presented and discussed in Section V.2. Section VI reports our empirical exercise for assessing the firm-level implications in the long run of set-asides for target firms. Section VII concludes.

II The US set-aside programs

In this section, we present the institutional framework of set-aside programs in US federal procurement. First, we outline the categories of set-asides and the targeted firms. Second, we detail the implementation of set-asides in practice.

[7]A recent exception is Carril and Guo (2023), which investigates the impact of a veteran-owned small business set-aside program on various firm and procurement outcomes. Other works on disadvantaged business programs focus on the effect of subcontracting requirements on targeted firms. See, for example, De Silva et al. (2012), De Silva et al. (2017), and Rosa (2020).

II.1 Set-aside programs in the US government procurement

The set-aside program has a long history in federal acquisitions, dating back to the 1950s when the US Congress passed the first Small Business Act, which established that a "fair share" of federal contracts had to be set aside for small businesses. The SBSA has by far the largest budget for its set-aside program. Currently, the federal government aims to allocate 23% of its contracting budget to small businesses.[8] In federal government contracting, revenues (or employment) below a certain threshold define a company as "small" for a specific tender category.[9] To participate in a specific SBSA tender, firms self-certify that they meet the size requirements for "small business".

In addition to SBSA, the government seeks to award at least 3-5% of the budget to disadvantaged firms. Similar in spirit and implementation to SBSAs, DBSAs target specific subsets of small businesses. These programs are more restrictive than SBSAs in that they impose further requirements on the prospective vendor. For example, to be eligible for the service-disabled veteran-owned small business set-aside program, a sub-category of DBSA, a service-disabled veteran with a service-connected disability must own a majority stake in the small business and be in a management position.[10]

II.2 Set-asides in practice

According to FAR §19.502-2 (a), "Each acquisition of supplies or services that has an anticipated dollar value above the micro-purchase threshold, but not over the simplified acquisition threshold, shall be set aside for small business [...]."[11] In practice, while contracts below the simplified acquisition threshold are generally earmarked for set-aside, the agency retains the discretion not to ()to) set aside a contract below (above) the threshold by justifying their decision. In particular, during the procurement planning stage, the agency decides whether to set aside a contract for small business concerns by considering two factors, namely (i) agency-specific set-aside annual spending and (ii) the Rule of Two, i.e., the number of targeted businesses in the marketplace deemed capable of performing the work at a competitive price.

The first factor implies that the more below (above) the agency is from its set-aside annual target, the higher (lower) the pressure to set-aside. Agencies must consider this aspect when drafting tenders, as the Small Business Administration evaluates their achievement of spending targets on an annual basis and issues scorecards. For agencies that fail to meet these goals, the consequences are generally not punitive, but include increased oversight by the Small Business Administration and Congress, the development and implementation of improvement plans to address shortfalls, and potential pressure from the legislative or executive branch to encourage small

[8]The Small Business Administration works annually with each agency to establish the specific annual goal to ensure that the federal government meets the overall goal of 23% small business participation in contracting.

[9]We provide more definition on the definition of "small business" in Appendix B.

[10]For details on the other types of set-asides, see Appendix B. We will discuss the service-disabled veteran-owned small business set-aside again in Section VI.

[11]Note that the simplified acquisition threshold was set at $100,000 until 2010 and $150,000 from 2011 for the rest of the period of our analysis.

business participation.[12]

The second factor is commonly known as the Rule of Two, which states that the agency must have a reasonable expectation of receiving a competitive bid from at least *two* small businesses to set aside a contract. Indeed, FAR §19.502-2 (a) clearly mandates the set-aside prescription, stating that it is applicable "[...] unless the contracting officer determines there is not a reasonable expectation of obtaining offers from two or more responsible small business concerns that are competitive in terms of fair market prices, quality, and delivery." The price, quality, and effective delivery chances associated with the bid must be expected to be fair and reasonable. The following conventional methods are recommended to prove such expectations: Reviewing recent procurement history, conducting market surveys, or obtaining expert advice from an agency small business specialist.[13] Critically, agencies must document their assessment of small business participation expectations and submit them to internal small business offices for approval prior to acquisition. Within the Department of Defense, for example, this process is standardized through the use of the DD Form 2579. A blank version of the DD Form 2579 is provided in the Online Appendix for reference.

III Theoretical background

In this section, we set out the theoretical foundations of the impact of set-aside in the procurement process. We divide the discussion into contract- and firm-level implications according to the structure of our empirical analysis, for which we postulate trade-offs. We empirically investigate these trade-offs in Sections V and VI.

Contract-level effects We group the number of bidders (to study tender competition), extra costs, and delays (to study post-award efficiency) under contract-level outcomes. We start with the hypotheses behind a competition effect.

To represent possible drivers of the effects of set-aside programs on competition, we postulate the following opposing hypotheses.

- Hypothesis *H.a1*: setting aside contracts diminishes the number of bids: the increased participation from targeted firms does not compensate for the missed participation from untargeted firms—*negative competition effect*;
- Hypothesis *H.a2*: setting aside contracts boosts the number of bids: the increased participation from targeted firms more than compensates for the missed participation from untargeted

[12] "There are no punitive consequences for not meeting these goals. However, the [...] scorecards and GSA's Small Business Goaling Report are distributed widely, receive media attention, and heighten public awareness of the issue of small business contracting. For example, agency performance as reported in the [...] scorecards is often cited by Members during their questioning of federal agency witnesses during congressional hearings." (Dilger, 2024, p.30).

[13] For detailed information on practical guidance for setting aside a contract, visit `https://www.sba.gov/partners/contracting-officials/small-business-procurement/set-aside-procurement`. In addition, a small business self-certification database (i.e., Dynamic Small Business Search) is available to agencies, which may use this database for consultation but must carefully consider a small business bidder's self-certification before awarding a contract.

firms—*positive competition effect.*

A negative effect would align with Li and Zheng (2009), who argue that more *potential* bidders result in less profitable entry due to increased competition and a lower winning chance. In contrast, a positive effect would be consistent with Cantillon (2008), who claim that reduced asymmetry in bidders' production costs leads to fiercer competition, as set-asides likely induce more symmetric value distributions. Thus, depending on the prevailing effect, set-asides could either encourage or discourage entry, influencing actual competition in procurement tenders. The latter effect would reinforce the equity impact of set-asides, confirming the entry of targeted firms, whereas the former could be attributed to the exit of larger firms only. The net outcomes depend on the mix of targeted and untargeted potential participants and the entry costs for targeted in the specific procurement market.[14]

However, how does excluding untargeted firms from bidding and reserving the participation to targeted bidders affect the efficiency of the procurement process? This question is important from an economic policy perspective. If targeted bidders deliver better quality, set-aside programs may lead to cost savings for taxpayers and improvements in public good provision, on top of aiming to improve equity in the process. However, if this relationship works the other way, excluding untargeted bidders could backfire and lead to waste. Looking at procurements allowing for an execution stage (i.e., post-award contract modifications), we will consider realized outcomes of set-aside contracts vis-à-vis expected outcomes in order to perform an overarching assessment of the policy on procurement efficiency. We envision four standalone hypotheses. Specifically,

- Hypothesis *H.b1*: Smaller firms receiving contracts might leverage their intimate knowledge of local resources and conditions, leading to more efficient resource allocation—*advantageous selection effect*;

- Hypothesis *H.b2*: Set-aside policies might incentivize smaller or less experienced firms to exert greater effort or commitment to prove their capabilities. This heightened dedication could enhance, e.g., project management and effort—*commitment effect*;

- Hypothesis *H.b3*: Set-aside policies could inadvertently attract firms that are less capable or experienced, leading to inefficiencies in contract execution—*adverse selection effect.*

- Hypothesis *H.b4*: Firms benefiting from set-aside policies might feel less pressure to perform optimally, knowing they face repeated opportunities under the program—*moral hazard effect*;

On the one hand, irrespective of an improved or adverse selection effect dominating, it is reasonable to hypothesize that bidder distribution in set-asides is different from open tenders. This is justified by the restriction to participation in set-asides for targeted firms and the likely refrain

[14] Our competition hypotheses implicitly build on the assumption that the set of participants in set-aside tenders is endogenous. Participants in set-aside tenders are either targeted incumbents who will definitely participate in the auction, or potential targeted entrants whose participation rates are endogenously determined to ensure that their expected payoff derived from participating matches their outside options. Excluding large, untargeted firms can therefore achieve the goal of attracting more targeted entrants to the auction (Jehiel and Lamy, 2015).

of some targeted firms from participating in open tenders. On the other hand, set-asides impact supplier incentives. Winners benefiting from set-aside policies might feel less (more) pressure to exert effort, knowing they face differential bidder composition and buyer's monitoring; this could lead to less (more) efficient contract execution, resulting in more (less) cost overruns or delays for the same winner across set-aside and non-set-aside procedures (henceforth NSA or open procedures). In the former case, a commitment effect would be at play; in the latter case, a moral hazard effect would apply. *H.b1* and *H.b2* would imply positive performance effects, while *H.b3* and *H.b4* negative effects.

Firm-level effects The declared overarching goal of the US set-aside program is to enable small (or disadvantaged) businesses to enter the procurement market so they can expand, thrive, and eventually compete on a level playing field with competitors. In the long run, we can imagine set-asides having both positive and negative effects on firm performance. We need to consider the underlying trade-offs for the government.

- Hypothesis *H.c1*: the opportunities provided by set-aside contracts could empower small businesses to expand their size and become more competitive—*market-enhancement effect*;
- Hypothesis *H.c2*: persistent reliance on the set-aside program could dissuade firms from expanding to a point where they lose eligibility for participation—*deadweight-loss effect*.

IV Data

In this section, we first present the data source and sample selection, then we describe the outcomes of interest and provide some stylized facts from our working sample.

Federal Procurement Data System (FPDS) The FPDS is a publicly available database that tracks the vast majority of US federal procurement spending.[15] The dataset provides a wealth of information, including over 200 variables on contract, seller, and buyer characteristics. Examples of contract-level information include the value of the contract, the start and end dates, the number of bids received, the place of performance, the type of product or service purchased, and the type of set-aside. Examples of seller-level information include the seller's ID and headquarters' location as well as the met small business standards, if any. An example of buyer-level information is the identity of the awarding agency. The FAR prescribes contracting agencies to report all awarded contracts with an estimated value above $3,500 (in the period under study) to the FPDS. Any subsequent contract modification, regardless of its amount, must also be reported and categorized in the FPDS.

Sample selection We start with the entire population of contracts from fiscal year (FY) 2008 to 2018. To calculate our metrics for execution-stage procurement outcomes, we need contracts that include an execution phase—i.e., with potential cost overruns and delays. Accordingly, we consider

[15] See `www.usaspending.gov`.

service and construction contracts. This selection implies excluding contracts where renegotiations are not meaningful to their outcomes: research and development, physical deliveries, and leasing.[16] Similarly, we exclude indefinite-delivery contracts as they are based on agreements with a supplier for an indefinite quantity of goods and services over a specified period of time, so delays and additional costs cannot be interpreted as indicators of poor quality performance. We also limit our sample to contracts that are performed within US borders and with a fixed-price format.[17] Finally, for the sake of comparable processes, we exclude very small contracts from the sample, i.e., those with an expected duration of less than 2 weeks *or* expected cost of less than \$25,000.[18] This ultimately leaves us with a sample of 141,199 contracts (38% and 15% of which are SBSA and DBSA, respectively) with a total value of \$125 billion (\$11 billion yearly on average), about 570,000 bids submitted and 54,800 unique winners.

Procurement outcomes As an award-level outcome, we consider the *# of offers*. Although we do not observe losing bids, we are interested in whether, *ceteris paribus*, set-aside contracts receive a divergent number of bids. Quantifying this impact allows us to learn about the differential composition and size of potential competition across procedures. As the main execution-stage contract outcome, we focus on *Extra Cost* and *Delay*. Specifically, *Extra Cost (abs.)* is the sum of all cost renegotiations related to a project that exceed the expected budget. We define it as (in \$): $\textit{Extra Cost (abs.)} = \textit{Final Cost} - \textit{Award Amount}$, where the latter term refers to the expected budget of the project (i.e., the contract value) and the former term refers to the actual cost. Similarly, we define $\textit{Delay (abs.)} = \textit{Final Duration} - \textit{Expected Duration}$, which is measured in days and corresponds to the difference between the actual and estimated completion date. We then compute the two measures relative to their benchmarks (i.e., award amount and expected duration) to define our outcome variables. We define the cost performance variable as *Extra Cost* $= \dfrac{\textit{Extra Cost (abs.)}}{\textit{Award Amount}}$ and $\textit{Delay} = \dfrac{\textit{Delay (abs.)}}{\textit{Expected Duration}}$.

We stress again that these metrics are the most common execution-stage metrics in both project management (Herweg and Schwarz, 2018) and empirical economics (Decarolis, 2014). They aim to capture the *quality* of contracts, based on the idea that renegotiations of fixed-price contracts lead to adjustment costs that are suboptimal for all parties involved. Indeed, Spiller (2008) proposes this argument for the first time for the public procurement context and claims that contracts are less flexible than private contracts. As a result, they require more frequent renegotiations and provide weaker incentives to comply with contract terms. Such rigidity leads to higher adjustment

[16] For research and development contracts, we would need to merge procurement data with patent generation data as in Decarolis et al. (2021). The outcome of interest would be the probability of R&D contracts obtaining a patent and the quality of the patents. Instead, relevant outcomes for physical goods are unit prices, as in Best et al. (2023). However, the FPDS withholds this type of unit price information. Finally, for leases or rentals, contract renegotiations are not an indicator of poor outcomes.

[17] The fixed-price format, which accounts for the vast majority of procurement contracts, sets the entire procurement value upfront, unlike cost-plus. Amendments are not mechanically included in the pricing format and need to be negotiated between parties, which implies the transaction cost we rely on for performance metrics.

[18] Note that the \$25,000 selection is a natural choice given that above this threshold contracts have to be posted on a centralized online platform. See Carril et al. (2022) for more details on this threshold. In Appendix D, we show that our results are robust to this sample selection.

costs than private contracts. Bajari et al. (2014) quantifies this claim empirically and finds that adjustment costs in public procurement are high and can account for 7.5% to 14% of the winning bid. Moreover, cost overruns and delay in public services and construction may create important disruption for the affected citizens.

We also build two secondary outcomes measuring the extensive and (a further dimension of) intensive margins of contract amendment. The details and the results on these two variables are presented in Appendix D.

Stylized facts Table 1 provides summary statistics on our outcomes of interest across control and treated groups. On average, treated tenders (i.e., SBSA or DBSA) have worse execution-stage outcomes (i.e., more delays and higher additional costs) than the control group (i.e., NSA). The difference is more pronounced for DBSAs. We find a similar pattern in terms of *Delay*. Note that both types of treated contracts receive more bids on average than non-treated contracts. Table 1 also displays that the average award amount of DBSAs is twice as large as that of SBSA and the former last longer, highlighting asymmetric distributions. From a descriptive viewpoint, participation restriction correlates with worse execution performance and more bids. Our first goal is to identify these effects. We do that in the next section.

*** INSERT TABLE 1 ABOUT HERE ***

Is set-aside used uniformly across procurement categories? Figure 1 shows the within-sector share of set-aside spending across all sectors in our data. We define the sector by considering the procurement category defined by the four-digit product or service code reported by FPDS.[19] This share is fairly uniform across sectors. In addition, the percentage of sectors in which set-asides are never or always used is 13% and 1%, respectively, which together represent $< 1\%$ of our sample of contracts. These statistics are reassuring given that the differential use of set-asides across sectors could be a potential confounding factor in our pooled comparison of procurements.

*** INSERT FIGURE 1 ABOUT HERE ***

V Contract-level analysis

V.1 Methodology

In this section, we describe the identification challenges for our contract-level analysis and our strategy to circumvent them. Then, we explain our two-step methodology, which hinges on random forest techniques to predict the propensity score of the set-aside treatments as a first stage, and utilizes propensity scores to identify our parameter of interest via an IPW as a second stage.

[19] Note that this definition is very granular. For example, the procurement category reported for "Maintenance of schools" will be different from "Maintenance of other educational buildings".

V.1.1 Identification concerns and strategy

At the contract level, our main empirical goal is to identify and quantify how restricting participation to different sets of targeted firms impacts procurement outcomes. Formally, we are interested in estimating the following model:

$$Y_i = \alpha + \beta^{SA}\, SetAside_i^{SA} + \gamma\, \mathbf{C}_i + \epsilon_i, \tag{1}$$

where Y_i and C_i denote our outcomes—e.g., number of bidders, extra costs and delays—and a vector of controls of the contract i, respectively. $SetAside_i^{SA}$ takes the value 1 if the contract is awarded through a set-aside procedure—i.e., $SA = \{SBSA; DBSA\}$—and 0 if no set-aside program is implemented. The parameter of interest in Equation (1) is β^{SA} as we are interested in the effect of SBSA and DBSA programs on the selected outcomes, separately. We proceed with the rest of the analysis using two different subsamples of the dataset, designating one of the two SA groups as the treatment group and maintaining the NSA as the control group.

Failing to consider the reasons why tenders are set aside for small or disadvantaged business would introduce endogeneity into our model as set-aside follows bureaucratic rules,which are implemented with decisional leeway by the agency. The wealth of our data enables us to trace back the information set underlying each agency's idiosyncratic set-side decisions.

The decision to set aside a contract relies on two main aspects. In Appendix C, we provide evidence that these two aspects are relevant for predicting the agency's decision. First, the agency needs to implement the Rule of Two (as discussed in Section II.2). The higher the number of "targeted competitive firms" available for a particular procurement—and thus in a specific market-time—the greater the likelihood that the agency will set aside the tender. We exploit various nuances related to the availability of targeted competitive firms across different market definitions, as neither dimension is specified in detail in the regulation. Second, the agency has a specific spending target for set-aside contracts during and over the years. The further below (above) its annual target the agency is at the time of award, the more (less) likely it is that the tender i will be set aside. Our data allows us to reconstruct this record as well.

We process the constructed information with a propensity score approach, which allows us to compare set-aside contracts similar in observable characteristics to open procedure contracts. We proceed with a two-stage approach. In the first stage, using random forest techniques, we predict the probability (i.e., propensity score) of set-aside treatment assignment (i.e., SBSA or DBSA) conditional on observed covariates. The estimated propensity scores are used in the second stage to run a weighted version of Equation 1 to identify our coefficients of interest. We discuss the details of the second stage in Section V.1.3. Under specific assumptions (we report them in Section V.1.2 and formally define them in Appendix C), this method allows us to perform an unbiased counterfactual analysis by eliminating key observable differences between the treatment and the control groups. We find no violation of such assumptions in our setting.

Finally, we argue that a propensity score-based approach is appropriate, as our setting exhibits desirable features for this method. In particular, the fact that the agency retains some margin of discretion in implementing set-aside rules is not an issue but rather an advantage for two reasons.

First, discretion prevents the treatment assignment from being deterministic. A deterministic assignment would violate the common support assumption on which the validity of the propensity score relies. Second, discretion is why similar contracts have different treatment statuses (Gibson and McKenzie, 2014). Indeed, agencies must justify why they do or do not set aside a contract (see Section II), making the role of private information in the decision negligible.

V.1.2 First stage: Propensity score of set-asides

In this subsection, we explain the approach used to predict the probability that a contract will be tendered with one of the two set-asides. We do so by using a random forest, following in particular Lee et al. (2010).[20] We also provide information about the treatment predictors and the identification assumption needed for this approach.

Random forest for binary prediction A random forest is built on "decision trees," consisting of a series of yes/no questions to predict the class for each observation—in this case, whether the contract is in the treatment or control group. The random forest grows trees, each time using a different bootstrapped sample of the data. When the outcome is binary, as in our case, the treatment allocation of each observation is predicted by a majority vote. The random forest deals with overfitting by selecting a random subset of features at each split (or node), i.e., each yes/no question. Using decision trees, the variable used at each node is the best *among all variables* in the data. Using a random forest, instead, only a subset of variables is randomly selected at each node, and the selected variable is the best *among the subset*.

Propensity score estimation In this work, we are interested in the probability of each contract being set aside. We define $p_i(X_i)$ for each tender i as the estimated propensity score receiving a binary treatment, where X_i is the vector of covariates.[21] Given the binary nature of the outcome, our random forest aggregates multiple classification trees.[22] To predict the propensity score, we use a classification method with multiple binary, categorical, and continuous covariates as inputs, and we perform two separate classifications for the two subsamples, one classification for the treatment being SBSA and one for the treatment being DBSA. For each subsample, we grow a random forest of 1,000 trees by closely following the practical guidance in Breiman and Cutler (2011) and Liaw et al. (2002) in tuning the parameters.[23] Following Lee et al. (2010), we compute a continuous propensity score for each contract by taking the average of the predicted outcomes of each tree.[24]

[20] Zhao et al. (2016) predicts the propensity score also using a random forest, but uses a slightly different approach. In our approach, we closely follow Lee et al. (2010) for practical application of the algorithm.

[21] Note that a natural alternative as a predictor of propensity scores in our setting is a causal random forest. However, its greatest merit is to allow us to look at heterogeneous effects, which is beyond the scope of this paper.

[22] In the literature, one refers to classification trees when the outcome variable is binary, while regression trees are used to predict continuous variables.

[23] See Appendix C for the exact steps taken in the random forest.

[24] Each of the 1,000 trees in our random forest predicts a classification for the predicted outcome. In other words, each unit receives a vote at each terminal node of each grown tree regarding the class to which it should belong. The average of this prediction is mostly a continuous number $\in [0, 1]$, since each tree is

Why using the random forest for propensity score estimation? The recent literature has shown the merits of machine learning techniques such as classification and regression trees when predicting propensity scores (e.g., Lee et al., 2010). Among these techniques, we choose the random forest for three reasons. First, not only this method requires only two tuning parameters but it is also not sensitive to their specification (Liaw et al., 2002). Second, we find that it outperform the logit—the traditional method for estimating the propensity score—in predicting the out-of-sample treatment. Third, it enables us to include a wealth of predictors, thereby fully exploiting the richness of our available dataset. Hence, the random forest also allow us to achieve covariate balance, while the logit does not. In Section V.2.3, we present the empirical exercises that corroborate these arguments.

To our knowledge, the random forest technique for estimating propensity scores is fairly new. This is evidenced by the recent works of Goller et al. (2020) and Lippmann (2021). While the former primarily aims to refine the propensity score computation methodology, the latter employs a random forest model to estimate a propensity score for augmenting a difference-in-differences analysis. The economic literature reveals additional applications of decision trees and random forests for predictive purposes. For instance, Hersh et al. (2022) apply a random forest model to forecast 3G network coverage, integrating these predictions into a Poisson fixed-effects model. Andini et al. (2018) demonstrate the efficacy of a decision tree model in improving the targeting of tax rebates to individuals. Antulov-Fantulin et al. (2021) utilize a random forest model to predict the bankruptcy of local governments. Furthermore, Andini et al. (2022) also leverage random forests to identify firms eligible for public credit guarantees.

Which variables explain the set-aside decision the most? We select and build variables that (i) replicate the information set to implement the Rule of Two, (ii) reconstruct set-aside spending in the agency from year-start, and (iii) control for contract and agency features. All of these covariates must satisfy the following conditions to be salient predictors: intuitively correlate with both treatment and outcomes simultaneously, are measured prior to treatment, and are orthogonal to the anticipation of treatment. We identify 71 variables the variables that meet these criteria in our data.[25] Then, we allow the random forest to select on the most relevant variables for the propensity score prediction. The agencies seem to adhere closely to the Rule of Two: Our random forest selects the (i) *percentage of small firms within a sector and division* (as defined by the Census Bureau) and (ii) *percentage of contracts previously awarded with set-aside by the same agency* as the two most significant factors for a contract to be set aside.[26]

grown differently as the variables at each node are chosen randomly, and this classifies each unit differently.

[25]We provide the full list of these variables in the Online Appendix.

[26]Among other variables, the other most important factors are the (iii) *percentage of competitive contracts awarded to small firms in the division/year*, (iv) *percentage of competitive small firms in the division/sector/year*, (v) *percentage of small firms that also win without set-aside in the sector/year*, (vi) *percentage of small firms that always win without set-aside in the sector/year*, and (vii) *percentage of small firms in the sector/year*. We refer the reader to Appendix C for an in-depth discussion of the logic, grouping, and relevance of first-stage predictors X_i.

Identification assumptions The propensity score approach relies on two main assumptions, which we formally define in Appendix C. First, we need *common support* (or "overlap assumption"), which implies for us that each contract (i) *could* receive both treatment levels and (ii) that we cannot *perfectly* predict the probability of receiving treatment. Both conditions apply in our setting since, on the one hand, contracts of any amount could be set aside; on the other hand, the set-aside choice is not entirely deterministic, as the rules are not binding. As discussed above, the role of private information in this choice is minimal, implying that the agency makes decisions based primarily on hard information about the contact and buyer characteristics. In Section V.2.3, we test this assumption and find that it is verified. Second, we assume *unconfoundedness* ("or conditional independence"), i.e., no unobservable or other observables might influence selection into treatment. In Section V.1.1, we documented the agency's decision-making process and argued that, through our data, we are able to reconstruct all relevant dimensions driving set-aside choices. Moreover, we do not find a violation of this assumption in our data and report the results in Section V.2.3.

V.1.3 Second stage: IPW regression using propensity scores as weights

As a second stage, we employ the inverse of the propensity scores estimated to perform as weights for a weighted OLS regression of Equation 1. The model specification turns into

$$Y_{i,b,t} = \alpha + \delta^{SA}\, SetAside_i^{SA} + \eta\, \mathbf{X_i} + \zeta_{b,t} + \varepsilon_i, \tag{2}$$

with X_i being the vector of first-stage predictors. In our regressions, when indicated, we include interaction fixed effects of agency and year—i.e., $\zeta_{b,t}$—to capture unobserved time-varying reasons for outcome variation and set-aside, e.g., agency budget cycles. However, because of collinearity, this implies that we cannot include the predictors at the agency level used in the first stage in the model. Nonetheless, they are captured due to the dynamic nature of the interaction fixed effects.

We are interested in the average treatment effect on the treated (ATT) contracts. Formally, in the potential outcome framework, the ATT is defined as $\theta_{att} = E(Y(1) - Y(0)|SA = 1)$, where $Y(1)$ is the outcome for the restricted solicitation, and $Y(0)$ is the counterfactual outcome had the contract had not received the treatment (Angrist, 1998; Heckman and Robb Jr, 1985).[27] To estimate an ATT, the IPW cancels out the differences in observable characteristics and gives more weight to those observations that have a low probability of receiving the treatment and vice versa—i.e., by weighting the treated by 1 and the control by $\frac{p_i(X_i)}{1-p_i(X_i)}$. For the standard errors of the estimates, we compute the variance-covariance matrix with a sandwich estimator when weights are used (Dupraz, 2013). Following the clustering recommendations of Abadie et al. (2022), we do not cluster the standard errors as the treatment decision is at the individual (i.e., contract) level. Finally, in Section V.2, we check that the results are not qualitatively sensitive to the chosen approach.

[27]Recently, Decarolis et al. (2021) and Bruce et al. (2019) employ IPW in a procurement context.

V.2 Results

In this section, we first document the effect of both types of set-aside programs on tender and contract outcomes vis-à-vis the counterfactual scenario of no set-aside. We then isolate possible effect channels and discuss the robustness of our findings. In summary, these exercises show that set-aside triggers more competition and execution-stage inefficiency. We provide suggestive evidence that worse execution-stage performance is triggered by adverse selection or moral hazard, depending on the set-aside type.

V.2.1 Baseline results

In Table 2, we report the baseline results of the IPW regressions. The top panel reports the estimated coefficient $\widehat{\delta^{SA}}$ from the Equation (2) for $SA = SBSA$, and the bottom panel reports the results for $SA = DBSA$. First, columns 1 and 2 report the effect on the log number of bids received. We then report the estimates on the two *execution-stage* contract outcomes, i.e. *Extra Cost* (columns 3 and 4) and *Delay* (columns 5 and 6). The odd columns show point estimates with a plain model specification, while the even columns include contract-level first-stage covariates and agency×FY fixed effects. We set the rich model as our preferred specification.

We find that SBSA increases the number of bids by 23.7%—i.e., $(\exp(0.213) - 1) \times 100$. DBSAs have a slightly weaker effect on competition, i.e., +20.2%. The magnitude of these effects amounts to approximately one additional bidder. While the Rule of Two is designed to ensure a minimum level of small business participation, our results show that set-asides attract more bids than the regulatory minimum. The average receipt of 3.6 bids for NSA contracts, our control group (see Section IV), confirms that the increase in bids for set-aside contracts goes beyond mere regulatory compliance, indicating a meaningful increase in competition. Notably, about 20% of SBSAs in our dataset received only a single bid, underscoring the unpredictability of bidder responses to set-aside auctions.

Both set-aside programs cause a deterioration in contract outcomes, namely an increase in *Extra Costs* and *Delays*. Specifically, the SBSA program increases *Extra Cost* relative to the award amount by 3.4 p.p.—i.e., +16.2% or $54,820 per contract—and it increases *Delays* relative to the expected duration by 4.6 p.p.—i.e., +5.2% or 9 days per contract. The DBSA program is more disruptive: *Extra Cost* and *Delays* increase by 6.5 p.p. and 7.8 p.p.—i.e., +30.9% and +8.9% relative to their benchmarks or +$104,800 and 16 days. However, this difference in point estimates between DBSA and SBSA is not significant, as their confidence intervals overlap.

*** INSERT TABLE 2 ABOUT HERE ***

All in all, these results show that restricting participation to small or disadvantaged firms increases competition in government tenders, confirming the intended equity effect of the policy. This points to a positive competition effect as put forward in Section III. Second, we find an increase in spending through additional budgetary costs and delays in contract execution. Over both cost and time dimensions, set-asides fare worse in the execution phase, thus pointing toward evidence in favor of H.b3 and H.b4. In the next subsection, we discuss how to empirically disentangle these mutually non-excluding channels. Finally, in V.2.3, we discuss the validity and robustness of the results.

V.2.2 Channels for the inefficiency effect

The goal of this subsection is to unveil potential mechanisms through which set-asides worsen contract outcomes at the execution stage. We caution that the estimates of the exercises in this section cannot be interpreted causally, but rather they provide suggestive evidence on two channels rationalizing the estimated inefficiency from set-asides, i.e., the adverse selection effect (H.b3) and the moral hazard effect (H.b4).

Do firms behave differently with set-aside? We start with enriching the baseline IPW model with winner fixed effects. In case of a prevailing adverse selection, we would expect the inefficiency effect to cancel out and the inefficiency to be rationalized by a higher frequency of low-quality winners with set-asides. This would mean that the same firm is performing similarly with and without set-aside. Instead, in the case of prevailing moral hazard effect, we expect the result to hold when adding a firm fixed effect, meaning that the firms perform worse because set-asides lessen supplier incentives to exert effort.

We present the results in Table 3 for both the execution-stage outcomes—*Extra Cost* in columns 1-4 and *Delay* in columns 5-8. We assess the impact on the same firm awarding contracts with or without SBSA or DBSA procedures (columns 2 and 6). We also interact the winner fixed effects with the award year (columns 3 and 7) or with the purchasing sector code (columns 4 and 8) to additionally capture unobserved firm-specific time trends (e.g., management practices) or firm-product-specific features (e.g., specialization), respectively. Columns 1 and 5 report the baseline estimates from Table 2 to facilitate comparison. For SBSA, estimates turn insignificant across the different specifications of firm fixed effects, suggesting an adverse selection effect. For DBSA, instead, we find statistically indistinguishable results from the baseline. Although we cannot exclude adverse selection, a moral hazard effect seems at play for DBSA.

*** INSERT TABLE 3 ABOUT HERE ***

Are set-aside winners different? To further rationalize these findings—namely to corroborate adverse selection and exclude moral hazard for SBSAs and vice versa for DBSAs—we investigate whether SBSA and DBSA winners have different characteristics than open procedure winners. To account for winner's characteristics, we construct three different outcome variables. First, *Entry* is an indicator equal to one if the firm has not previously been awarded a contract in our data. Second, *Incumbent* is a binary variable equal to one if the winner is awarded at least one contract in the previous year. Third, we construct *Backlog* by counting the number of contracts won in the previous quarter. All these variables capture the concept of incumbency with different specifications. Incumbency is relevant as it informs about firms' experience in the procurement market.[28]

[28]Note that these three variables refer to previous procurement activity of the firm using the entire FPDS population of service and construction contracts. Hence, we avoid missing values by executing such

We find that firms winning SBSA contracts are indeed more likely to win for the first time, are less likely to be incumbents, and have a lower backlog, confirming prevailing adverse selection. For DBSA, we corroborate the exclusion of an adverse selection. Compared to non-set-asides winners, DBSA-winning firms are less likely to be winning for the first time, have higher incumbency, and do not show differences in terms of backlog. As these firms fare worse under DBSA, this evidence highlight a possible underlying moral hazard.

*** INSERT TABLE 4 ABOUT HERE ***

Does monitoring set-asides mitigate their inefficiency? To draw further intuitions on the channels at play, we leverage the role of performance bonds. Performance bonds, issued by surety companies, unequivocally distinguish construction from service contracts in the US federal procurement. The bonds are shown to enhance contract execution outcomes by enforcing effective screening and monitoring on suppliers, both theoretically (Calveras et al., 2004) and empirically (Giuffrida and Rovigatti, 2022), mitigating supply risks. Surety companies consider firms' past performance when issuing bonds, motivating suppliers to maintain high standards in their operations, even when bonds are not contractually required.[29]

This specific attribute of performance bonds provides a context to discern the underlying mechanisms of inefficiencies from set-aside contracts. Table 5 presents split-sample regressions for construction and service contracts. Service contracts exhibit heightened inefficiencies stemming from set-asides, whereas construction contracts are largely unaffected. Our findings indicate that through the embedded more screening (that mitigates adverse selection) and monitoring (that mitigates moral hazard during execution), performance bonds effectively redress the inefficiency introduced by both set-aside by fixing both sources of asymmetric information. In contrast, the lack of performance bonds in service contracts exposes them to such inefficiencies triggered by set-asides.

*** INSERT TABLE 5 ABOUT HERE ***

These results combined are suggestive evidence of a selection effect for SBSA. In other words, we find the SBSA selects worse firms. This can be rationalized by the fact that the barrier for entry is quite low for small firms with no further designation. As the small-business status can be self-certified, the chances of a bad winner are higher due to the absence of large firms. Instead, for participating in DBSA, firms have to be certified, which results in a stronger ex-ante screening.

V.2.3 Validity of the identification assumptions

In this subsection, we first report the tests of our identification assumptions that corroborate the internal validity of our causal claims. Then, we justify the chose random forest approach and

regressions on all years except for FY2008, the first year in the data. We replicate baseline analysis without FY2008 and results (unreported) are qualitatively and quantitatively unaffected to such exclusion.

[29]The rationale of a performance bond is consistent with Butler et al. (2020), who show that considering past performance can improve quality provision.

provide evidence on the better performance of the random forest compared to the logit. We refer the reader to Appendix D for additional robustness checks on methodology, alternative outcomes, and sample selection. Altogether, such empirical exercises highlight the robustness of our findings.

Testing the identification assumptions As discussed in Section V.1.2, our chosen method relies on two assumptions on the propensity score, i.e., the overlap and unconfoundedness assumptions. We test validity of the first assumption in two different ways. First, to check the satisfaction of the overlap assumption, we verify that the propensity score distributions for both the treated and control overlap. The results of this check are shown in Figure E1 in Appendix E. We find a strong overlap, and we exclude those units from the relevant population whose probability of receiving the treatment can be perfectly (or almost perfectly) predicted (Wooldridge, 2010). Therefore, we restrict our sample to units for which the propensity score is strictly between 0.01 and 0.99, implying removing 11.6% of the observations for the SBSA sample and 23.7% for the DBSA sample without loss of internal validity.

Second, in Figure 2, we look at the covariate balance between treated and control contracts. These results allow us to check the validity of both assumptions. Indeed, when covariate balance is achieved, the propensity score has been adequately specified, implying that treatment effect estimates is valid and unbiased (Ho et al., 2007; Zhao et al., 2016). For this purpose, we compute the standardized differences after applying the IPW (Imbens and Rubin, 2015). This procedure allows us to assess the comparability of treated and control units in the weighted sample (Austin, 2009). Moreover, the standardized differences are more robust than simply calculating the t-statistic or testing the difference in means because they do not depend directly on the sample size (Wooldridge, 2010). As long as the standardized differences do not exceed 0.25 (in absolute value), covariate balancing between groups should be satisfied (Imbens and Rubin, 2015; McCaffrey et al., 2004; Stuart, 2010). We obtain excellent results: for SBSA all of the variables used for propensity score prediction are below 0.1. For DBSA, 5.6% are above 0.1, and none are above 0.15.

*** INSERT FIGURE 2 ABOUT HERE ***

We test the validity of the unconfoundedness assumption following Imbens and Rubin (2015). To do so, we assess the sensitivity of our estimates to different choices of predictor variables. In these tests, we remove some variables from the propensity score prediction covariates vector. In other words, we rely on the concept of subset unconfoundedness, which is a more restrictive condition than the original assumption on which the propensity score relies, the unconfoundedness assumption (Caliendo and Kopeinig, 2008). As mentioned before, this assumption is not directly testable, and consequently neither is subset unconfoundedness. However, the two assumptions combined have some testable implications: We can test whether adjusting for differences in a subset of covariates give similar point estimates as with the full set of covariates. When we do so, we find that the coefficients are very stable across different specifications and are not statistically different. We report the results in Figure 3. The first line in each figure reports the baseline estimates from Table 2, in the specification with fixed effects and additional controls. We report the coefficients for SBSA with red dots, while we use blue triangles for DBSA. In lines two and three, we report the coefficients excluding variables that we created to account for, respectively, (i) the performance in terms of extra cost and delay in the previous period, at the agency level and (ii) the level of competition in the geographical region and sector. Therefore, we can conclude that we find no evidence of (i) a lack of the overlap assumption and (ii) the violation of the unconfoundedness assumption.

*** INSERT FIGURE 3 ABOUT HERE ***

Testing the random forest's performance As discussed in Section V.1.1, we choose the random forest for three reasons. For each dimension, we now provide empirical evidence to strengthen our arguments. First, we show that the random forest is robust to the specification of its parameter, by twisting the main parameter of the random forest, i.e. the number of variables selected at each node.[30] In Figure 3 line four and five, we impose the number of variables instead of using the number of variables based on prediction accuracy.[31] We report the results for alternative numbers of variables selected at each node.[32]

Second, we test the prediction performance of random forest vis-à-vis logit. We train both models on the same sub-sample of the data and predict the treatment on the unused part of the sample in the second step.[33] The random forest has a much higher accuracy of 77%, while the logit correctly classifies the treatment in only 48% of the cases. This difference in prediction performance suggests that the random forest does a better job of using the observable information provided in the data. We note that, however, the logit performs better in terms of the overlap assumption. This can be observed in E2 in Appendix E, which replicates Figure E1 using the logit.

Third, we have shown in this subsection that the random forest achieves great covariate balance. We stress here that the random forest, by allowing the inclusion a rich set of covariates and controlling for nonlinearities and interaction, achieves better covariate balance than a logit. We provide evidence of this in Figure E3 in Appendix E, replicating Figure 2. We observe that, for some of the buyer characteristics, the covariate balance is not achieved with the logit, as some standardized differences are above the absolute value of 0.25. Therefore, the validity of the coefficient is not met for the logit approach, while it is for the random forest.

For completeness, in a robustness check, we also replicate the main analyses using the logit instead of the random forest. In Appendix D, we report the results with the logit and show that they are quantitatively and qualitatively similar to our baseline. Moreover, the coefficients with the logit are always larger than with the random forest. If anything, this could suggest that our chosen methodology underestimates the effect. We stress again that the logit estimates cannot be causally interpreted, given that with this method covariate balance is not achieved.

[30]Note that the second important parameter for the random forest is the number of trees (Liaw et al., 2002). However, this is not relevant in our case as we are already using 1,000 trees.

[31]We refer to Appendix C for more details about accuracy and its definition in this context.

[32]The random forest with the highest accuracy selected is the one with 16 variables for SBSA, and 8 variables for DBSA. As we look at the highest accuracy between 4, 8 and, 16 variables, we report the results for the two other options.

[33]We split the dataset and use 80% of it as training sample and the remaining 20% as testing sample.

VI Firm-level analysis

In the previous section, we show that set-asides boost the participation of targeted firms, promoting equity in the process, and induce poorer contract outcomes, both in terms of cost and time dimensions. In this way, excluding untargeted competitors creates inefficiencies in the government procurement market. Can this additional cost to the taxpayer be viewed as an "investment" from the policymaker's perspective? In other words, do firms that benefit from set-aside contracts perform better in the future, thus potentially leading to increased efficiency in the long run? To provide evidence for a preliminary answer to these questions, in this section we empirically investigate the underlying trade-off for the government between the market-enhancement effect (i.e., *H.c1*) and the deadweight-loss effect (i.e., *H.c2*) as posited in Section III.

Identifying a firm-level effect of set-asides poses empirical challenges. A primary concern is that winning *specific* contracts can be anticipated by a firm, making winning a set-aside contract endogenous to a firm's unobserved characteristics. For instance, we are agnostic about the political connectedness of targeted firms. Recent research suggests that connected companies are more likely not only to win contracts (Ağca and Igan, 2023), but also to thrive Akcigit et al. (2023). If a future winner can anticipate a contract—resulting from participation in a particular bidding process—then public demand remains endogenous in our models. This and other concerns could lead to biased estimates of the coefficients of interest.

To address such challenges, given that demand variation is a crucial driver of firm dynamics (see Section I), we propose an event-study exercise that builds on a large, permanent, and unexpected construction spending increase by the Department of Veteran Affairs (DVA) on a given category of DBSA (i.e., the service-disabled veteran-owned small business set-aside, henceforth VBSA for convenience).[34] In addition to the policy-induced variation, we leverage the variation brought about by the differential pre-period exposure of recipients eligible to the spending spike.

VI.1 The VBSA spending surge

The Recovery Act and veteran policies in 2009 We leverage the two government policies that were independently introduced during the same period. First, newly elected President Obama signed the Recovery Act into law in February 2009. This stimulus package contained provisions for the construction industry beginning in FY2009 to increase investment in the nation's physical infrastructure and cope with the ongoing economic downturn. Notably, such resources were mostly channeled through government procurement contracts.[35] Second, the years of the fiscal stimulus also saw policies in line with the stated mission of President Obama to increase subsidies for veterans. As a result, the President's budget request in FY2009 included \$140 billion for the DVA—a 40% increase in funding from FY2008.[36] Figure 4 shows how the combination of these

[34]As this section focuses on VBSA, we replicate our contract-level analysis of DBSA versus NSA using VBSA as the treatment group. This auxiliary exercise documents that VBSA is also disruptive for execution-stage outcomes. In particular, extra costs increase, but we find no meaningful effect in delays. We report the results in Table E1 in Appendix E.

[35]For more information, see here.

[36]In his first term, the President Obama made veterans' care one of his top priorities. See more here.

policies increased DVA's total construction spending by approximately 50% from 2008 to 2010.[37]

*** INSERT FIGURE 4 ABOUT HERE ***

Service-disabled veterans were a core target of the new veteran policy. As a result, the combined implicit effect of the Recovery Act and the new veteran policy was to increase VBSA spending in constructions. As Figure 5 shows, the absolute increase in construction spending in FY2009-2010 (compared to FY2008) for VBSAs was threefold, rising from approximately $0.5B in FY2008 to $1.5B in FY2010. Such spending amount holds in the years after, thus highlighting a persistent spending surge. Moreover, the same figure emphasizes that a similar increase in construction spending from the DVA for other types of set-asides is not observed in our data.

*** INSERT FIGURE 5 ABOUT HERE ***

Contamination concerns and firm sample selection This discussion raises the following question: Which firms benefited from the VBSA spending surge? In theory, any *small* construction firms classified as *service-disabled veteran-owned* were eligible to compete for the increased VBSA contract opportunities. As shown in Figure 6, the cumulative yearly number of bidders in VBSA auctions increases strongly from 2009 onward. The increasing opportunities for service-disabled veteran-owned small businesses likely drive this statistic as incentivizing incumbent businesses to participate in VBSA tenders. Also, small firms (with no further designation) could have been incentivized to switch to the small service-disabled veteran-owned category to take advantage of the increased government demand. For example, an existing small firm could have invited a disabled veteran to become a major shareholder in the company. Moreover, at the beginning of FY2009—i.e., after the election victory—President Obama also announced the withdrawal of (most) US troops from Iraq. Accordingly, many veterans returned home between 2009 to the end of 2011, suggesting a growing number of service-disabled veterans establishing startups and entering the construction industry.[38] The latter two aspects jointly highlight a possible concurrent supply shock of service-disabled veteran-owned small bidders that might contaminate the exogeneity of contract recipience and introduce self-selection in our sample.

The post-FY2009 service-disabled veteran-owned small firm pool is thus most likely composed of incumbents and entrants. Specifically, we define incumbents as firms that were awarded at least one VBSA construction contract before FY2009 *or* labeled service-disabled veteran-owned despite being awarded construction contracts outside the VBSA program. We define entrants as

[37]Note that we are using FPDS data before FY 2008 for this analysis, which we do not employ in the rest of the paper. The choice is dictated by the improvements in data quality of FPDS from 2008 onward—e.g., several pieces of information needed for the propensity score estimation are either not present before 2008 or are coded differently. For the firm-level analysis, we leverage only a few pieces of procurement information from FPDS available prior from 2004, namely the winner identity (i.e., its DUNS number), the set-aside category, and contract value, the ingredient we need to build the outcomes of our firm-level analysis.

[38]This trend has been observed and reported by the media. Most veterans who start businesses do so in the construction industry because construction jobs best match veterans' skills. For example, see `https://www.nvti.org/2023/08/31/employment-in-the-construction-industry-for-veterans/`. Recently, Coile et al. (2021) has shown that 30 percent of service-disabled veterans of the Vietnam War have transitioned into self-employment.

newly re-labeled or newly established service-disabled veteran-owned firms in the data after 2009. Our analysis restricts the attention to incumbents to reduce self-selection issues and increase firm comparability. In doing so, we spotlight 999 firms potentially exposed to the persistent demand surge equally unexpected. Indeed, they could not perfectly anticipate in FY2008 the recession—the first GDP decline was registered at the end of summer 2008, that is, shortly before the start of FY2009—and the resulting introduction of the Recovery Act. The same applies to the launch of Obama's veteran policy, as presidential elections came in November 2008 (second month of FY 2009).

*** INSERT FIGURE 6 ABOUT HERE ***

Stylized facts We briefly report key descriptives on our sample of incumbents and VBSA construction contracts. 324 incumbents receive 57.2% of the increase in cumulative VBSA construction spending in FY2009–2010 (the first two years of the shock), while 425 entrants receive the remaining 42.8%. This allows variability in contract recipience that we could exploit for identification. Incumbents in 2008, our base year in the analysis, win on average 6 construction contracts, amounting to about $1.8 million, 69% of which are awarded through VBSA. Moreover, they are usually awarded 80% of their contracts with some set-aside. Finally, on average, they sell in 1.6 US states and sell 3.2 different categories of works and services to the government.

VI.2 Long-run implications of VBSAs

Firm-level outcome variables Due to the lack of information in the data on firms' activity in the private market or their balance sheets, we study incumbent performance dynamics in the procurement market. We rely on four different metrics built at the level of firm i in a given FY t:

1. *Log(VBSA Total Sales)*: the log total sales to the government through VBSA contracts;

2. *Set-Aside Share*: the share of set-aside sales over total procurement sales;

3. *# of States*: the number of US states in which the firm performs its contracting activity, weighted by the yearly number of contracts won by a firm;

4. *# of Categories*: the number of different procurement categories associated with the firm sales, weighted by the yearly number of contracts won by a firm.

Studying the first outcome represents as a "first stage" for this analysis as it verifies whether higher incumbency in the VBSA construction market truly mirrors higher total VBSA construction awards after the spending shock. The dynamics of *Set-Aside Share* are particularly useful for testing the market-enhancement versus deadweight-loss effect hypothesis. On the one hand, a negative effect could be due to an increase in firm size (i.e., exceeding the revenues or employment thresholds for small firms definition) or an increase in firm competitiveness. In the second case, the increased share of procurement revenues results from competition with large firms outside the set-aside program, regardless of an increase in size. On the other hand, a positive effect hints toward the lack of incentive to grow and lose eligibility for set-aside participation, especially if a size effect is at stake. The last two metrics measure the expansion into new markets and the introduction of new category variants, respectively; furthermore, they have the merit of being less affected by a potential crowd-out issue of public revenues over private revenues and mirror with different nuances true firm activity scale.

The event study To identify the effect of permanent variation in VBSA demand on incumbents' outcomes, we employ an event study framework comparing incumbents *differentially exposed* to the demand shock. The idea is that the ex-post dynamics of the less exposed incumbent firms trace out the dynamics that we would have observed in the more exposed treated observations absent the treatment.[39] We define the exposure as $Share_i^{pre}$, measured as the annual average non-VBSA sales over total procurement sales for the five years before the treatment (i.e., between FY2004 and FY2008). Formally, this is equivalent to

$$Share_i^{pre} = \frac{1}{5}\sum_{t=2004}^{2008} \frac{\text{Non-VBSA Revenues}_t}{\text{Total Revenues}_t.}$$

Our hypothesis is that incumbents that rely less heavily on VBSAs before the demand surge have more scope to be affected than incumbents that rely less on it. The logic is that, even if both incumbents are service-disabled veteran-owned firms, those with a high value of $Share_i^{pre}$ are competitive enough to win a high share of their revenues in open tenders (or other set-asides). When the VBSA demand increases, such firms are therefore more likely to win more VBSA contracts (where they do not compete openly with untargeted firms). In other words, a service-disabled veteran-owned small firm that typically received a high share of revenues from non-VBSA contracts before the demand surge has higher chance to win the increased VBSA opportunities in tenders contested with targeted firms than a firm that typically receives a lower share of non-VBSA. This hypothesis is a necessary condition for the causal claim that the different ex-post dynamics are actually caused by the spending shock. We provide evidence on the relevance of this assumption through the "first stage" estimates jointly with the lack of pre-trends in outcomes.

Figure E4 shows the frequency of $Share_i^{pre}$. About 26.7% of the firms exhibit a share equal to zero, meaning that the entirety of their public revenues stems from VBSA awards. Instead, 40.9% obtain all their revenues without winning any VBSA contracts. The median firm's share is 75%. This descriptive evidence further highlights variation in the exposure useful for identification.

Formalizing this idea, we estimate the following event-study regression,

$$W_{i,t} = \theta_i + \iota_t + \sum_{T \neq 2008} b_t \cdot 1(t = T) \times Share_i^{pre} + \omega_{i,t}, \tag{3}$$

where W_{it} stands for the four outcomes of interest for the firm i in period t. Firm and time fixed effects are represented by θ_i and ι_t, respectively. $Share_i^{pre}$ is the continuous treatment (i.e., exposure) variable and $\omega_{i,t}$ is the error term. The primary coefficients of interest are the b_t's—the

[39] A similar approach to study policies with no variation in the timing of treatment and a measure of agents' exposure to the policy has been recently adopted by Beheshti (2022) and Coviello et al. (2022).

event-study coefficients of the interaction of $Share_i^{pre}$ with time fixed effects.

The long-run effects Panels A through D of Figure 7 report a visual representation of the $\hat{b_t}$'s for each outcome, where FY2008 is chosen as the base year $t-1$. By observing the 95-percent confidence interval in the figures, we see that all but two coefficients before the treatment across the four outcomes are statistically insignificant. Furthermore, we test the absence of pre-trends non-parametrically by checking the joint statistical significance of the point estimates. On the bottom-right of each figure, we report P_b, the p-value of the joint test on the pre-treatment coefficients, which does not reject the parallel trend assumption for any outcome.

Panel A of Figure 7 confirms our exposure assumption and provides "first stage" evidence. Indeed, we find that the higher the exposure, the more the incumbents increase their VBSA procurement sales after the demand surge. They do so permanently as the shock. We report the results for *Set-Aside Share* in Panel B. We find evidence that exposed firms increase their share of set-aside sales. This suggests that the intensity of unexpected VBSA demand exposure does not lead to a size or efficiency effects for recipient firms. We consider the impact on our proxies for scale, i.e., *# of States* and *# of Categories*, in Panels C and D, respectively. We find (weak) evidence that exposed firms become less localized. Recipients sell additional product categories, as the coefficients are jointly statistically significant in the post-period. Jointly considered, the latter results suggest a positive scale effect.

Robustness checks We propose two exercises for the robustness of these results. First, the results remain virtually unchanged when we focus on post-period contracts with more than one bid, since they are more likely to have been won unexpectedly. The graphical evidence can be found in Figure E6 in Appendix E. Second, we replicate Panel C separately for each US region as defined by the US Census. Indeed, it might be easier for a firm to enter new states if located in regions consisting of several small states (e.g., the Northeast) than in regions with a few large states (e.g., the West). We report our results in Figure E5 in Appendix E. We find similar effects in all regions, i.e., a (weak) increase in the number of states.

Overall, we find no clear-cut evidence of a market-enhancement effect of VBSA exposure through the analyzed dimensions. While there is some evidence that firm size increases over time—at least as measured by greater product variety and (weakly) more states of activity—there is no evidence that their performance increases more broadly, as firms tend to sell more within the set-aside program. On the contrary, this last finding points to a deadweight-loss effect at play as winners become increasingly dependent on set-aside tenders for their operations. Firms restrict their growth, if any, to the maximum eligible size for set-aside.

VII Conclusions

This paper investigates the contract- and firm-level implications of the US federal set-aside procurement program. We show that set-asides prompt more firms to bid—that is, the increase in targeted bidders more than offsets the loss in untargeted bidders. During the execution phase, set-aside contracts incur higher cost overruns and delays. The more restrictive the set-aside, the stronger these effects, although not significantly so. Our contract-level results originate from a market with over 140,000 service and construction projects performed throughout the US over eleven years. Our estimates are robust to alternative methodologies, outcomes, and sample selections. When focusing on firms unexpectedly exposed to a surge in service-disabled veteran-owned set-aside spending, we find mixed evidence of performance improvement over the long run, at least in terms of future procurement outcomes. This evidence suggests that the service-disabled veteran-owned set-aside program is *de facto* more of a subsidy than an industrial policy.

Our microeconomic analysis provides evidence that set-asides introduce inefficiencies into the US federal procurement process. However, we do not advocate for discontinuing programs that support targeted businesses. Instead, in light of existing literature, we suggest exploring alternatives to plain set-asides or modifying them with tailor-made designs to improve welfare. On the one hand, bid subsidies could replace set-asides and preserve the benefits of set-aside auctions while limiting their distortions. Studies of preference programs (e.g., Athey et al., 2013) show that if smaller firms are encouraged to participate with a bid subsidy and larger firms are not excluded, competition increases at no efficiency cost. On the other hand, set-asides could promote equity in the procurement market at a lower efficiency cost with an alternative tailored design. Jehiel and Lamy (2020) advocate increasing revenues in set-aside auctions by excluding the targeted incumbent—i.e., bidders whose entry costs are zero or already sunk. The authors show that, for a given set of participants, the indirect benefit of excluding the targeted incumbent—obtained by increasing the participation of the targeted entrants—always dominates the direct cost of not having the incumbent.

Future research could explore how combining set-asides with other policies could be beneficial as this line of research is underexplored. Our heterogeneity analysis of procurement categories shows that set-asides do not induce meaningful inefficiencies in the case of construction contracts. We attribute this finding in our data to performance bonds, which add an extra layer of screening and monitoring to the procurement of public works and prevent the inefficiencies that arise from adverse selection and moral hazard in set-aside tenders.

Bibliography

Abadie, A., Athey, S., Imbens, G. W., and Wooldridge, J. M. (2022). When Should You Adjust Standard Errors for Clustering? *The Quarterly Journal of Economics*, 138(1):1–35.

Akcigit, U., Baslandze, S., and Lotti, F. (2023). Connecting to power: Political connections, innovation, and firm dynamics. *Econometrica*, 91(2):529–564.

Alcalde, J. and Dahm, M. (2024). On the trade-off between supplier diversity and cost-effective procurement. *Journal of Economic Behavior & Organization*, 217:63–90.

Andini, M., Boldrini, M., Ciani, E., De Blasio, G., D'Ignazio, A., and Paladini, A. (2022). Machine learning in the service of policy targeting: the case of public credit guarantees. *Journal of Economic Behavior & Organization*, 198:434–475.

Andini, M., Ciani, E., de Blasio, G., D'Ignazio, A., and Salvestrini, V. (2018). Targeting with machine learning: An application to a tax rebate program in italy. *Journal of Economic Behavior & Organization*, 156:86–102.

Angrist, D. (1998). Estimating the Labor Market Impact of Voluntary Military Service Using Social Security Data on Military Applicants. *Econometrica 66: 2*.

Antulov-Fantulin, N., Lagravinese, R., and Resce, G. (2021). Predicting bankruptcy of local government: A machine learning approach. *Journal of Economic Behavior & Organization*, 183:681–699.

Athey, S., Coey, D., and Levin, J. (2013). Set-asides and subsidies in auctions. *American Economic Journal: Microeconomics*, 5(1):1–27.

Ağca, c. and Igan, D. (2023). The lion's share: Evidence from federal contracts on the value of political connections. *The Journal of Law and Economics*, 66(3):609–638.

Austin, P. C. (2009). Balance diagnostics for comparing the distribution of baseline covariates between treatment groups in propensity-score matched samples. *Statistics in medicine*, 28(25):3083–3107.

Austin, P. C. (2011). An introduction to propensity score methods for reducing the effects of confounding in observational studies. *Multivariate Behavioral Research*, 46(3):399–424.

Austin, P. C. and Stuart, E. A. (2015). Moving towards best practice when using inverse probability of treatment weighting (IPTW) using the propensity score to estimate causal treatment effects in observational studies. *Statistics in medicine*, 34(28):3661–3679.

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

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

Beheshti, D. (2022). The impact of opioids on the labor market: Evidence from drug rescheduling. *Journal of Human Resources*.

Best, M. C., Hjort, J., and Szakonyi, D. (2023). Individuals and organizations as sources of state effectiveness. *American Economic Review*, 113(8):2121–67.

Beuve, J., Moszoro, M. W., and Saussier, S. (2019). Political contestability and public contract rigidity: An analysis of procurement contracts. *Journal of Economics & Management Strategy*, 28(2):316–335.

Breiman, L. (2001). Random forests. *Machine Learning*, 45(1):5–32.

Breiman, L. and Cutler, A. (2011). Manual–setting up, using, and understanding random forests V4. 0. 2003. *URL https://www. stat. berkeley. edu/˜ breiman/Using_random_forests_v4. 0. pdf.*

Bruce, J. R., de Figueiredo, J. M., and Silverman, B. S. (2019). Public contracting for private innovation: Government capabilities, decision rights, and performance outcomes. *Strategic Management Journal*, 40(4):533–555.

Butler, J. V., Carbone, E., Conzo, P., and Spagnolo, G. (2020). Past performance and entry in procurement: An experimental investigation. *Journal of Economic Behavior & Organization*, 173:179–195.

Caliendo, M. and Kopeinig, S. (2008). Some practical guidance for the implementation of propensity score matching. *Journal of Economic Surveys*, 22(1):31–72.

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

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

Calzolari, G. and Spagnolo, G. (2009). Relational Contracts and Competitive Screening. *CEPR Discussion Papers*.

Cantillon, E. (2008). The effect of bidders' asymmetries on expected revenue in auctions. *Games and Economic Behavior*, 62(1):1–25.

Cappelletti, M., Giuffrida, L. M., and Rovigatti, G. (2022). Procuring survival. *CESifo Working Paper*, (10124).

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

Carril, R. et al. (2021). Rules versus discretion in public procurement. *Barcelon GSE Working Paper Series*, (1232).

Carril, R., Gonzalez-Lira, A., and Walker, M. S. (2022). Competition under incomplete contracts and the design of procurement policies. *Economics Working Papers. Department of Economics and Business, Universitat Pompeu Fabra*.

Carril, R. and Guo, A. (2023). The impact of preference programs in public procurement: Evidence from veteran set-asides. *BSE Working Papers*, (1417).

Coile, C., Duggan, M., and Guo, A. (2021). To work for yourself, for others, or not at all? how disability benefits affect the employment decisions of older veterans. *Journal of Policy Analysis and Management*, 40(3):686–714.

Coviello, D., Guglielmo, A., and Spagnolo, G. (2017). The effect of discretion on procurement performance. *Management Science*, 64(2):715–738.

Coviello, D., Marino, I., Nannicini, T., and Persico, N. (2022). Demand shocks and firm investment: Micro-evidence from fiscal retrenchment in Italy. *The Economic Journal*, 132(642):582–617.

Czarnitzki, D., Hünermund, P., and Moshgbar, N. (2020). Public procurement of innovation: Evidence from a German legislative reform. *International Journal of Industrial Organization*, 71:102620.

De Silva, D. G., Dunne, T., Kosmopoulou, G., and Lamarche, C. (2012). Disadvantaged business enterprise goals in government procurement contracting: An analysis of bidding behavior and costs. *International Journal of Industrial Organization*, 30(4):377–388.

De Silva, D. G., Kosmopoulou, G., and Lamarche, C. (2017). Subcontracting and the survival of plants in the road construction industry: A panel quantile regression analysis. *Journal of Economic Behavior & Organization*, 137:113–131.

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. M., Iossa, E., Mollisi, V., Raiteri, E., and Spagnolo, G. (2021). Buyers' role in innovation procurement: Evidence from US military R&D contracts. *Journal of Economics & Management Strategy*, 30(4):697–720.

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

Denes, T. A. (1997). Do small business set-asides increase the cost of government contracting? *Public Administration Review*, pages 441–444.

Desrieux, C., Chong, E., and Saussier, S. (2013). Putting all one's eggs in one basket: Relational contracts and the management of local public services. *Journal of Economic Behavior & Organization*, 89:167–186.

di Giovanni, J., García-Santana, M., Jeenas, P., Moral-Benito, E., and Pijoan-Mas, J. (2022). Government Procurement and Access to Credit: Firm Dynamics and Aggregate Implications. Staff Reports 1006, Federal Reserve Bank of New York.

Dilger, R. J. (2024). An overview of small business contracting. Report R45576, Congressional Research Service.

Dupraz, Y. (2013). Using weights in Stata. Technical report, Paris School of Economics.

EC (2017). European semester: Thematic factsheet–public procurement.

Fadic, M. (2020). Letting luck decide: Government procurement and the growth of small firms. *The Journal of Development Studies*, 56(7):1263–1276.

Ferraz, C., Finan, F., and Szerman, D. (2015). Procuring firm growth: The effects of government purchases on firm dynamics. *National Bureau of Economic Research*, (21219).

Fischbacher, U., Fong, C. M., and Fehr, E. (2009). Fairness, errors and the power of competition. *Journal of Economic Behavior & Organization*, 72(1):527–545.

Foster, L., Haltiwanger, J., and Syverson, C. (2016). The slow growth of new plants: Learning about demand? *Economica*, 83(329):91–129.

Gerardino, M. P., Litschig, S., and Pomeranz, D. (2017). Can audits backfire? evidence from public procurement in chile. *National Bureau of Economic Research*, (23978).

Gibson, J. and McKenzie, D. (2014). The development impact of a best practice seasonal worker policy. *Review of Economics and Statistics*, 96(2):229–243.

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

Goldman, J. (2019). Government as Customer of Last Resort: The Stabilizing Effects of Government Purchases on Firms. *The Review of Financial Studies*, 33(2):610–643.

Goller, D., Lechner, M., Moczall, A., and Wolff, J. (2020). Does the estimation of the propensity score by machine learning improve matching estimation? The case of Germany's programmes for long term unemployed. *Labour Economics*, 65:101855.

Gugler, K., Weichselbaumer, M., and Zulehner, C. (2020). Employment behavior and the economic crisis: Evidence from winners and runners-up in procurement auctions. *Journal of Public Economics*, 182:104112.

Hart, O. and Moore, J. (1988). Incomplete contracts and renegotiation. *Econometrica*, 56(4):755–785.

Hart, O. D. and Holmstrm, B. (1986). The theory of contracts.

Hebous, S. and Zimmermann, T. (2021). Can government demand stimulate private investment? evidence from U.S. federal procurement. *Journal of Monetary Economics*, 118:178–194.

Heckman, J. J. and Robb Jr, R. (1985). Alternative methods for evaluating the impact of interventions: An overview. *Journal of Econometrics*, 30(1-2):239–267.

Hersh, J., Lang, B. J., and Lang, M. (2022). Car accidents, smartphone adoption and 3G coverage. *Journal of Economic Behavior & Organization*, 196:278–293.

Herweg, F. and Schwarz, M. A. (2018). Optimal cost overruns: Procurement auctions with renegotiation. *International Economic Review*, 59(4):1995–2021.

Ho, D. E., Imai, K., King, G., and Stuart, E. A. (2007). Matching as nonparametric preprocessing for reducing model dependence in parametric causal inference. *Political analysis*, 15(3):199–236.

Imbens, G. W. and Rubin, D. B. (2015). *Causal inference in statistics, social, and biomedical sciences*. Cambridge University Press.

Jehiel, P. and Lamy, L. (2015). On discrimination in auctions with endogenous entry. *American Economic Review*, 105(8):2595–2643.

Jehiel, P. and Lamy, L. (2020). On the Benefits of Set-Asides. *Journal of the European Economic Association*, 18(4):1655–1696.

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

Krasnokutskaya, E. and Seim, K. (2011). Bid preference programs and participation in highway procurement auctions. *American Economic Review*, 101(6):2653–86.

Laffont, J.-J. and Tirole, J. (1993). *A theory of incentives in procurement and regulation.* MIT press.

Lee, B. K., Lessler, J., and Stuart, E. A. (2010). Improving propensity score weighting using machine learning. *Statistics in Medicine*, 29(3):337–346.

Lee, M. (2021). Government purchases and firm growth. *Available at SSRN 3823255.*

Li, T. and Zheng, X. (2009). Entry and Competition Effects in First-Price Auctions: Theory and Evidence from Procurement Auctions. *Review of Economic Studies*, 76(4):1397–1429.

Liaw, A., Wiener, M., et al. (2002). Classification and regression by randomForest. *R news*, 2(3):18–22.

Liebman, J. B. and Mahoney, N. (2017). Do expiring budgets lead to wasteful year-end spending? Evidence from federal procurement. *American Economic Review*, 107(11):3510–49.

Lippmann, Q. (2021). Are gender quotas on candidates bound to be ineffective? *Journal of Economic Behavior & Organization*, 191:661–678.

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

Marion, J. (2009). How costly is affirmative action? Government contracting and California's Proposition 209. *The Review of Economics and Statistics*, 91(3):503–522.

McCaffrey, D. F., Ridgeway, G., and Morral, A. R. (2004). Propensity score estimation with boosted regression for evaluating causal effects in observational studies. *Psychological methods*, 9(4):403.

Nakabayashi, J. (2013). Small business set-asides in procurement auctions: An empirical analysis. *Journal of Public Economics*, 100:28–44.

OECD (2019). *OECD SME and Entrepreneurship Outlook 2019.* OECD Publishing, Paris.

Pérez-Castrillo, D. and Riedinger, N. (2004). Auditing cost overrun claims. *Journal of Economic Behavior & Organization*, 54(2):267–285.

Pozzi, A. and Schivardi, F. (2016). Demand or productivity: What determines firm growth? *The RAND Journal of Economics*, 47(3):608–630.

Rosa, B. V. (2019). Resident bid preference, affiliation, and procurement competition: Evidence from new mexico. *The Journal of Industrial Economics*, 67(2):161–208.

Rosa, B. V. (2020). Subcontracting requirements and the cost of government procurement. *The RAND Journal of Economics.*

Spiller, P. T. (2008). An Institutional Theory of Public Contracts: Regulatory Implications. *National Bureau of Economic Research*, (14152).

Spulber, D. F. (1990). Auctions and Contract Enforcement. *The Journal of Law, Economics, and Organization*, 6(2):325–344.

Strobl, C., Boulesteix, A.-L., Zeileis, A., and Hothorn, T. (2007). Bias in random forest variable importance measures: Illustrations, sources and a solution. *BMC bioinformatics*, 8(1):25.

Stuart, E. A. (2010). Matching methods for causal inference: A review and a look forward. *Statistical science: a review journal of the Institute of Mathematical Statistics*, 25(1):1.

Tkachenko, A., Valbonesi, P., Shadrina, E., and Shagbazian, G. (2019). Efficient design of set-aside auctions for small businesses: an empirical analysis. "Marco Fanno" Working Papers 0240, Dipartimento di Scienze Economiche "Marco Fanno".

Williamson, O. E. (1971). The vertical integration of production: market failure considerations. *The American Economic Review*, 61(2):112–123.

Wooldridge, J. M. (2010). *Econometric analysis of cross section and panel data.* MIT press.

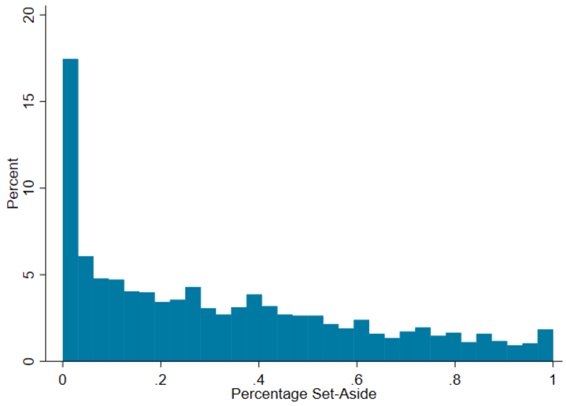
Zhao, P., Su, X., Ge, T., and Fan, J. (2016). Propensity score and proximity matching using random forest. *Contemporary Clinical Trials*, 47:85–92.

Zhao, Z. (2004). Using matching to estimate treatment effects: Data requirements, matching metrics, and Monte Carlo evidence. *Review of Economics and Statistics*, 86(1):91–107.

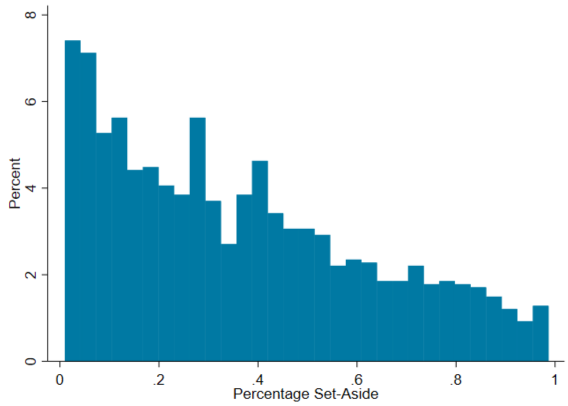
A Appendix: Main figures and tables

Figure 1: Set-aside Spending Shares per Sector

Panel A: All sectors

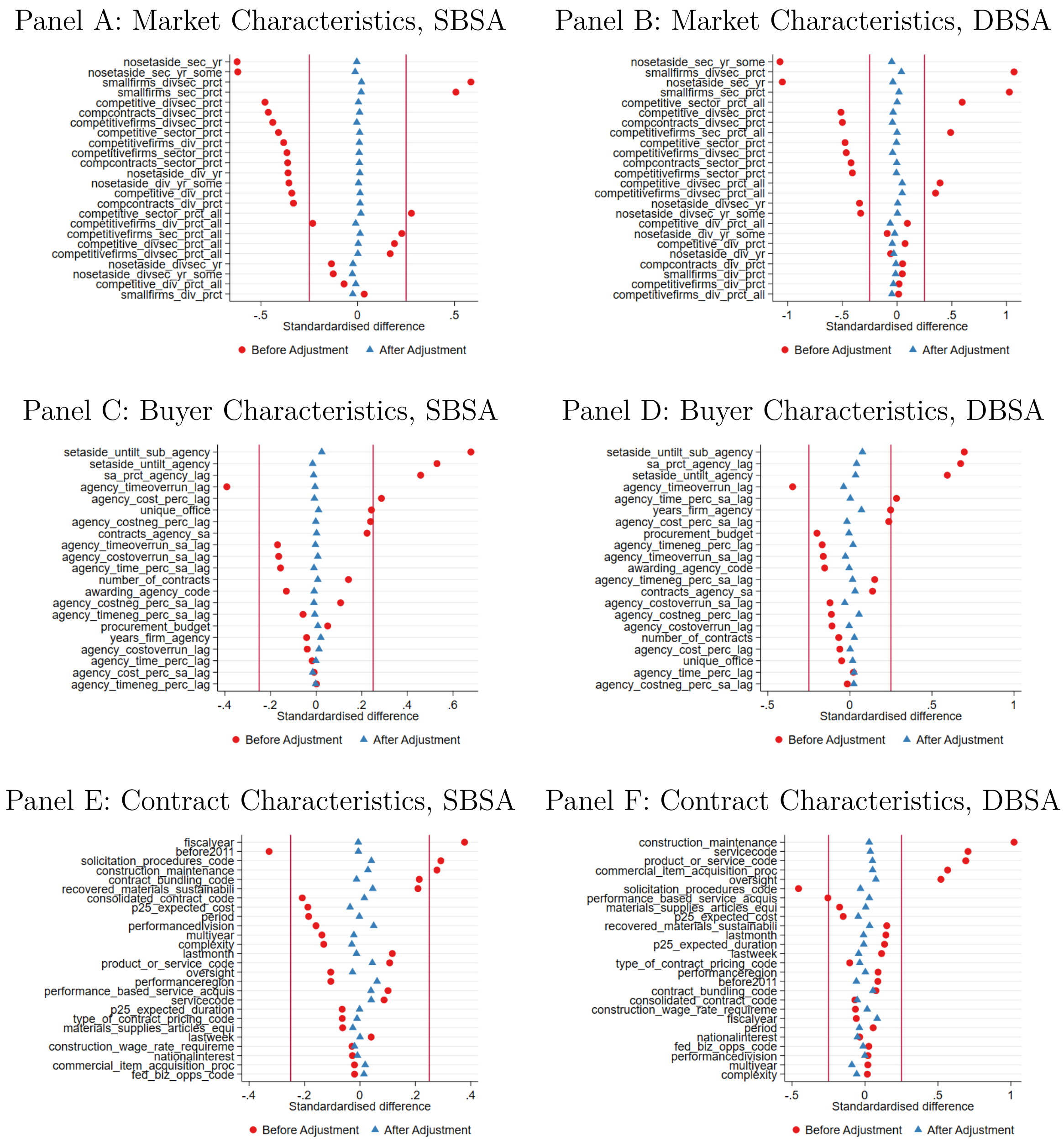


Panel B: Sectors with spending $\in [1\%, 99\%]$



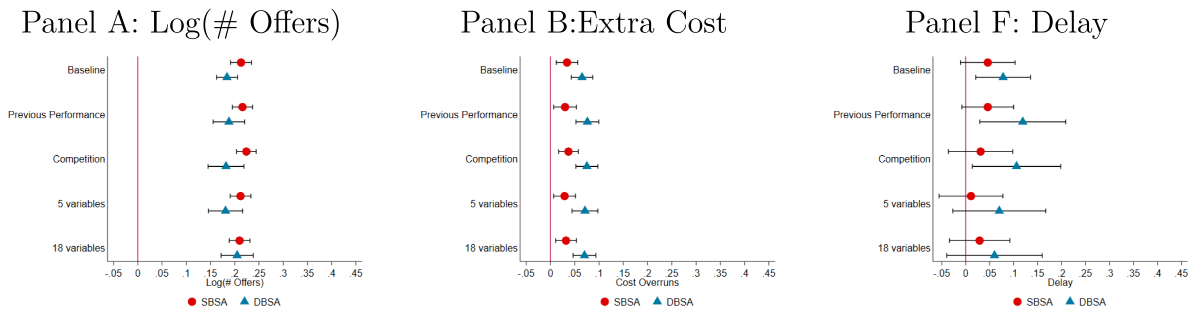
Notes: The figure shows the percentage of set-aside expenditure for each sector. In this context, we define the sector using the most granular definition (all 4 digits) of the FPDS variable "Product or Service Code". Panel A reports all sectors, while Panel B provides a closer look at the percentages for sectors with a share between 1% and 99%.

Figure 2: Standardized Differences



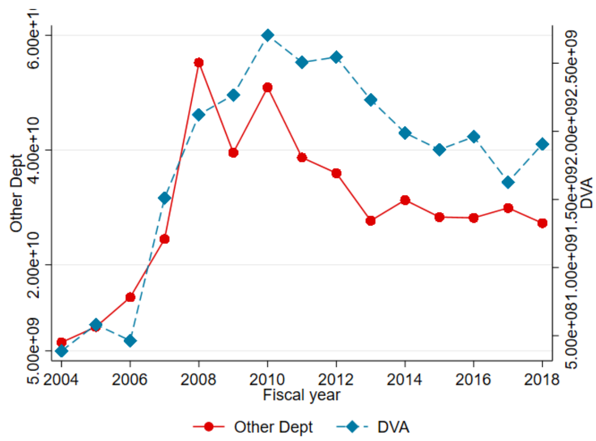
Notes: This figure reports the standardized differences before and after the IPW. The treatment for this sample is specified in the panel title; the control is non set-aside contract. We predict the propensity score using 71 variables. We report variables in the three relevant groups: market, buyer and contract characteristics. The red vertical lines represent the -0.25 and 0.25 thresholds. Below this threshold, in absolute value, covariate balance is achieved (Imbens and Rubin, 2015). Standardized differences for all the covariates use, after adjustment, are below the absolute value of 0.25.

Figure 3: Robustness Checks, Functional



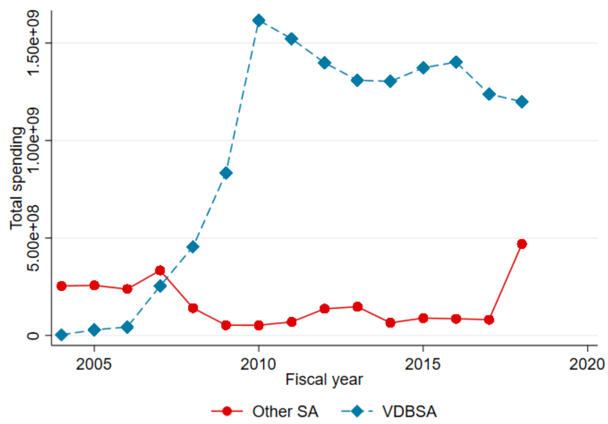
Notes: This figure reports the different robustness checks performed on our analysis. The panel title report the outcome variable of interest. The red dot reports the coefficient for small business set-asides (SBSA) as treatment group, the blue triangle for disadvantaged business set-asides (DBSA). The black confidence interval are at the 95% confidence level. We report the different functional robustness checks performed on the propensity score prediction with the random forest. The first line reports the baseline regression, as in Table 2, with 16 (8) variables selected at each node of the random forest, seed equal to 1994 and all variables. In the second and third line, we remove some variables related to previous performance and competition. In lines four and five, we change the number of variables selected by the random forest.

Figure 4: Total Spending by the DVA vs all Other Departments, Construction



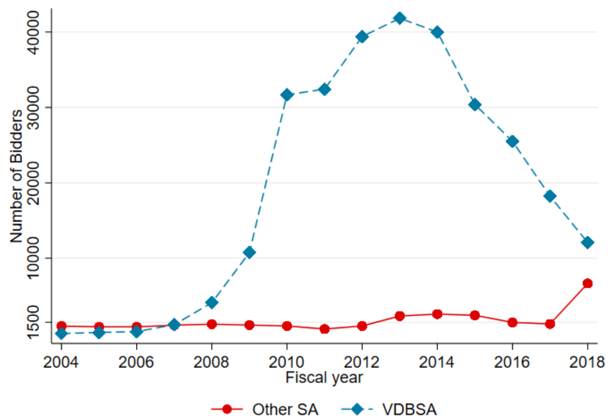
Notes: This figure reports the total yearly spending on construction. We plot separately the amount spent by the DVA and the one by all other departments. Refer to the y-axis on the right and on the left respectively.

Figure 5: Total Spending in VBSA vs Other Set-Asides by the DVA, Construction



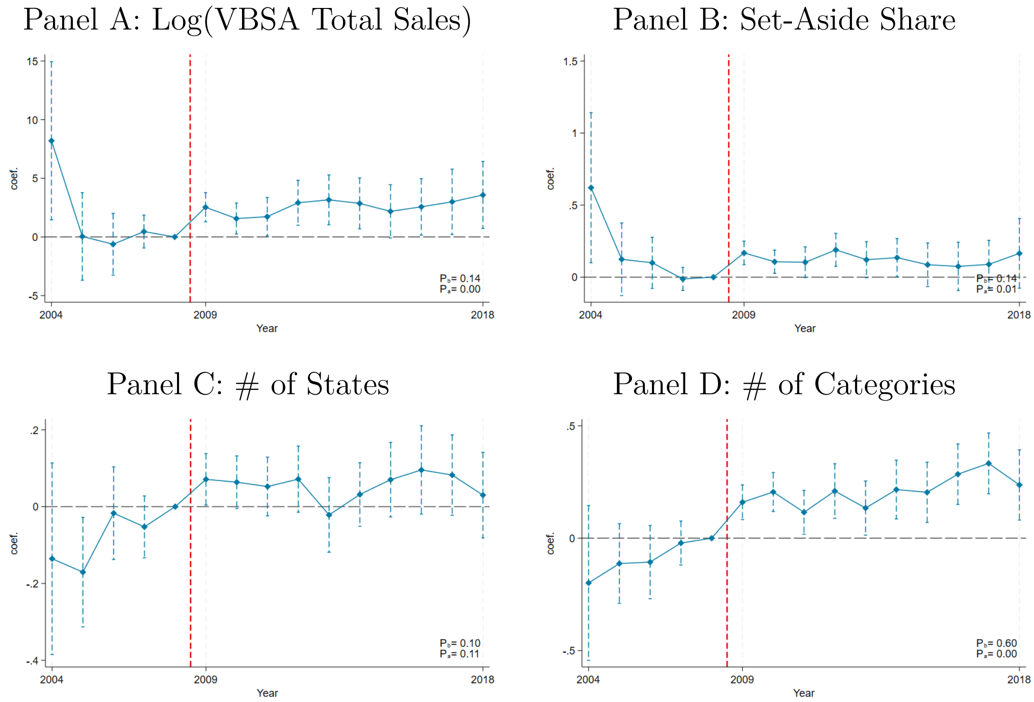
Notes: This figure reports the total yearly spending by the DVA on construction. We plot separately the amount spent on VBSA and the spending and on "Other SA", i.e., SBSA and other types of DBSA.

Figure 6: Number of Bids in VBSA vs Other Set-Asides for Contracts Awarded by the DVA, Construction



Notes: This figure reports the total number of bids received by the DVA's construction tenders. We plot separately the bids on VBSA and on "Other SA", i.e., SBSA and other types of DBSA.

Figure 7: Firm-Level Event Study



Notes: Results for the event study analysis, reporting the coefficients of the interaction of $Share_{it}^{pre}$ with FY FEs. The policy was enacted in FY2009 (at time *t*), i.e., to the right of the red vertical line. FY2008, time *t-1*, is chosen as the base year and all reported coefficients are relative to it. The dependent variables are constructed at the firm-year level. Panel A reports the log total sales to the government awarded through VBSA. Panel B shows the share of set-aside awards over total sales. Panel C and D report, respectively, the number of different procurement categories associated with its sales and the number of states in which it performs its activity. We report the 95 percent confidence intervals using standard errors clustered at the firm level. P_b reports the joint-significance of the coefficients in the before period, i.e. before FY2009, while P_a in the after period.

Table 1: Summary Statistics for Treatment and Control Groups

	NSA (N= 66,290)		SBSA (N = 54,158)		DBSA (N = 20,751)	
	Mean	Median	Mean	Median	Mean	Median
# Offers	3.65	2.0	4.30	3.0	4.88	4.0
# Renegotiations	1.43	0.0	1.41	0.0	2.42	1.0
Award Amount ($)	1189.35	73.2	462.10	67.3	999.93	214.8
Extra Cost	0.21	0.0	0.24	0.0	0.24	0.0
Expected Duration (days)	287.48	222.0	229.83	152.0	280.53	212.0
Delay	0.88	0.0	0.92	0.0	1.11	0.2

Notes: This table reports summary statistic at the contract level for non set-aside contracts (NSA), small business set-aside (SBSA) and other small business set-asides, i.e. disadvantaged business set-aside (DBSA). Dollar values are in thousands. *# Offers* counts the number of offers received for a given contract. *Extra Cost* indicates the share of the cost overrun relative to the expected cost. *# of Renegotiations* variables report the number of in-scope contract modifications. *Expected Duration* is the duration of the contracts expressed in days. *Delay* measures the ratio of delay relative to the expected duration.

Table 2: Baseline Outcomes, IPW

	Log(# Offers)		Extra Cost		Delay	
	(1)	(2)	(3)	(4)	(5)	(6)
SBSA	**0.293**	**0.213**	**0.019**	**0.034**	**0.060**	**0.046**
	(0.011)	(0.011)	(0.012)	(0.011)	(0.027)	(0.029)
	0.754	0.837	0.234	0.253	0.921	0.952
	106,432	*74,934*	*105,730*	*74,425*	*106,418*	*74,928*
DBSA	**0.274**	**0.184**	**0.049**	**0.065**	**0.090**	**0.078**
	(0.021)	(0.019)	(0.012)	(0.014)	(0.041)	(0.046)
	0.761	0.856	0.247	0.268	0.960	0.995
	66,337	*46,533*	*65,956*	*46,242*	*66,331*	*46,529*
Buyer*FY FE		✓		✓		✓
Market Controls		✓		✓		✓
Contract Controls		✓		✓		✓

Notes: Results for the ATT of the IPW on the three outcomes: (1). *Log(# Offers)* is the log of the number of offers received for a given contract, (2) *Extra Cost* indicates the share of the cost overrun relative to the expected cost, and (3) *Delay* measures the ratio of delay relative to the expected duration. The treated are contracts awarded with restricted solicitations: In the top panel contracts are set aside for small businesses (SBSA), in the bottom they are disadvantaged business set-asides (DBSA). First-stage predictors accounting for contract and market characteristics, and a buyer/fiscal-year fixed effect are included, when indicated, in the regressions. The propensity score of the probability of being treated is separately and previously predicted with the "randomForest" package in R. We report in bold the coefficient estimates, standard errors are in parentheses. The third row refers to the mean outcome for the control group and the fourth row (in italics) reports the number of observations.

Table 3: Execution-stage Outcomes with Seller Fixed Effects

	Extra Cost				Delay			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
SBSA	**0.034**	**0.015**	**0.020**	**0.017**	**0.046**	**-0.009**	**-0.047**	**-0.020**
	(0.011)	(0.010)	(0.014)	(0.011)	(0.029)	(0.032)	(0.046)	(0.037)
	0.253	0.237	0.225	0.237	0.952	0.911	0.899	0.895
	74,425	*52,124*	*29,584*	*43,856*	*74,928*	*52,550*	*29,896*	*44,215*
DBSA	**0.065**	**0.049**	**0.093**	**0.034**	**0.078**	**0.120**	**0.138**	**0.105**
	(0.014)	(0.023)	(0.031)	(0.030)	(0.046)	(0.067)	(0.108)	(0.079)
	0.268	0.255	0.249	0.256	0.995	0.946	0.941	0.942
	46,242	*31,458*	*17,915*	*25,645*	*46,529*	*31,690*	*18,073*	*25,835*
Buyer*FY FE	✓	✓	✓	✓	✓	✓	✓	✓
Market Controls	✓	✓	✓	✓	✓	✓	✓	✓
Contract Controls	✓	✓	✓	✓	✓	✓	✓	✓
Seller FE		✓	✓	✓		✓	✓	✓
Seller*FY FE			✓				✓	
Seller*Sector Code FE				✓				✓

Notes: In columns 1 and 5, we replicate Table 2 columns 4 and 6, to which we add seller fixed effects (columns 2 and 6), firm-year fixed effects (columns 3 and 7) as well as seller-one-digit-service-code fixed effects (columns 4 and 8).

Table 4: Entry, Incumbency, and Backlog

	Entry [N,Y]	Incumbent [N,Y]	Backlog
SBSA	**0.036**	**-0.075**	**-0.257**
	(0.006)	(0.008)	(0.012)
	0.159	0.696	0.828
	70,894	*70,894*	*70,894*
DBSA	**-0.056**	**0.078**	**-0.014**
	(0.009)	(0.013)	(0.018)
	0.172	0.675	0.740
	42,844	*42,844*	*42,844*
Buyer*FY FE	✓	✓	✓
Market Controls	✓	✓	✓
Contract Controls	✓	✓	✓

Notes: The baseline model specification from Table 2 (columns 4 or 6) is replicate on firm-level outcomes: *Entry [N,Y]* is an indicator equal to one if the firm is winning a contract for the first time. *Incumbent [N,Y]* is a binary variable equal to one if the firm is an incumbent, i.e., won a contract in the previous year. *Backlog* counts the number of contracts won in the previous quarter.

Table 5: ATET – Heterogeneity by Works and Services

	Log(# Offers)			Extra Cost			Delay		
	Baseline	Works	Services	Baseline	Works	Services	Baseline	Works	Services
SBSA	**0.213**	**0.076**	**0.266**	**0.034**	**0.003**	**0.039**	**0.046**	**-0.005**	**0.070**
	(0.011)	(0.020)	(0.013)	(0.011)	(0.008)	(0.014)	(0.029)	(0.050)	(0.034)
	0.837	1.305	0.705	0.253	0.133	0.287	0.952	0.898	0.972
	74,934	*18,382*	*56,149*	*74,425*	*18,233*	*55,794*	*74,928*	*18,379*	*56,146*
DBSA	**0.184**	**0.085**	**0.330**	**0.065**	**-0.008**	**0.159**	**0.078**	**-0.199**	**0.356**
	(0.019)	(0.022)	(0.023)	(0.014)	(0.007)	(0.028)	(0.046)	(0.077)	(0.063)
	0.856	1.305	0.700	0.268	0.128	0.318	0.995	0.914	1.023
	46,533	*15,818*	*30,491*	*46,242*	*15,712*	*30,315*	*46,529*	*15,815*	*30,490*
Buyer*FY FE	✓	✓	✓	✓	✓	✓	✓	✓	✓
Market Controls	✓	✓	✓	✓	✓	✓	✓	✓	✓
Contract Controls	✓	✓	✓	✓	✓	✓	✓	✓	✓

Notes: Results for the ATT of the IPW on the three main outcomes, *Log(# Offers)*, *Extra Cost* and *Delay.* We report, for each outcome, the baseline results as in Table 2. We then report results separately by procurement category, either services or constructions.

B Appendix: Additional information on set-asides

Definition of a small business The Small Business Authority defines a vendor as small based on the particular service category to which the contract belongs and its characteristics, whether revenues or revenues and number of employees. The list is revised every five year. The NAICS code defines the industry. For example, consider a company with 700 employees that is involved in "Wind Electric Power Generation" (NAICS 221115) or "Tire Manufacturing except retreading" (NAICS 326211). The firm would not be considered small in the former case because it has more than 250 employees (i.e., the threshold for this NAICS category). However, in the latter category, it would be considered small because it has fewer than 1,500 employees.[40]

The different types of DBSA Other prominent DBSA programs besides the VBSA include the Historically Underutilized Business Zones (HUBZone) program, the 8(a) Business Development Program, and the Women-Owned Small Business program. The first helps small businesses in urban and rural communities gain preferential access to government procurement opportunities. To be eligible for the HUBZone program, a US citizen must own and control the business, and it must be a Community Development Corporation, an agricultural cooperative, a Native Hawaiian organization, or an Native American Indian tribe. The second targets small businesses owned by socially and economically disadvantaged individuals or organizations. The third supports women-owned businesses. To qualify for the Women-Owned Small Business Program, a company must be small and at least 51% owned and controlled by US citizens, and women must manage day-to-day operations and long-term decision-making. It is also industry-specific and available to industries where women are underrepresented. Unlike SBSAs, firms can only bid on DBSAs tenders if they have gone through a certification process. A firm that is certified for DBSAs is automatically eligible to participate in SBSA solicitations.

Table B1: Types of Set-Asides

	All Departments		Dept. Veteran Affairs (DVA)		All Dept. (Except DVA)	
Set-aside Type	N	Percentage	N	Percentage	N	Percentage
NSA	66,290	46.9	7,963	40.4	58,327	48.0
SBSA	54,158	38.3	2,514	12.8	51,644	42.5
SESA	20,751	14.7	9,216	46.8	11,535	9.5
Total	141,234	100	19,693	100	121,541	100

Notes: This table reports summary statistic for type of set-aside in our sample of contracts across all departments, only for DVA, and excluding DVA.

Table B1 shows that 53.1% of contracts in the working sample are awarded through some type of set-aside procurement, 38.3% through SBSA, and 14.7% through DBSA. These percentages change for the DVA, mostly due to the VBSA program—a major DBSA category: only 12.8% are awarded through SBSA, and 46.8% are awarded through DBSA.

[40]For more detailed information on the Small Business Standards, see the Small Business Size Standards Matched to the North American Industry Classification System Code table available at `www.sba.gov/content/small-business-size-standards`.

C Appendix: Additional details on the contract-level empirical strategy

In this Appendix, we provide four additional pieces of information regarding the empirical strategy for the contract-level analysis. First, we report detailed information about the set-aside treatment predictors. Second, we explain in more detail the random forest approach and how we set the different parameters. Third, we provide some first stage results in which we identify the most important treatment predictors using the random forest. Fourth, we formally explain the IPW estimator used in the second stage, as well as the relevant assumptions for this methods.

Treatment predictors by groups We can divide the control variables used in the propensity score into three main categories: Market, buyer, and contract characteristics. We cannot include information about the seller in this stage because the winning firm is determined after the treatment is assigned. Such information is processed in the second stage only. We start by describing the variables included and the rationale behind this selection. In the Online Appendix, we report the full dictionary of predictors.

- *Market characteristics.* With different market characteristics, we aim to replicate the information set underlying the agency's implementation of the Rule of Two. To approximate the application of this rule as closely as possible we do not impose a specific "competitive small firm" and market definitions as they are not clearly defined by regulation; rather, we set up variables for different constellations of "competitive small firms" and market definitions, which we all include as predictors. For the different measures that we construct and that we present below, we consider three possible and relevant definitions of market: (i) the *sector*, i.e. the product (four-digits) code of the contract as defined by FPDS, to capture a nation-wide product-wise boundaries of market definition; (ii) the Census *division*—which splits the US in nine different areas, e.g, Pacific, Mountain, South Atlantic—of contract performance, to capture the geographic boundaries of market definition; iii) a combination of the two.

  We define a contract as *competitive* if (i) it is awarded without set-aside and (ii) at least two bids were submitted. A targeted firm is competitive if it secures competitive contracts. Our rational is that in a market with many (few) competitive targeted firms, the Rule of Two will be more (less) likely satisfied. Consequently, agencies might be more likely to set aside contract in such markets. Therefore, we first construct proxies for the competitiveness of a market, such as the share of contracts that are competitive in a given market. In addition, we calculate different proxies looking at small firms. For instance, we calculate percentage of competitive small businesses relative to all small businesses. We also control for the percentage of small firms winning without set-aside, as well as the percentage of small firms in a given market.

- *Buyer characteristics.* With buyer characteristics, we are approximating agency quality, experience, as well as set aside culture and goals. It is often impractical to specify fully complete procurement contracts (Hart and Moore, 1988). However, some buyers may be better at drafting more complete contracts, resulting in less renegotiations and more bids. Competence in drafting contracts could be correlated with a better knowledge in setting contracts aside, with consequent effects on the contract outcomes. Therefore, agency's past experience should be considered for the treatment prediction. We construct a proxy that takes into account the agency's experience with set-asides to control for the contracting agency's performance on past delays and additional costs on awarded contracts. We include the proportion of contracts where cost or schedule overruns were reported prior to each award in *all* set-aside contracts.

We also include three different measures of agency size. First, we define *Contracting Offices* as the number of distinct contracting units within an agency. With this variable, we are also able to control for the complexity of the agency structure as well as for the intensity procurement activity. Second, we include *Procurement Budget*, the agency's total yearly procurement spending for all purchase categories. Finally, *Contracts* is the agency's total yearly number of contracts awarded for all purchase categories. The latter is complementary to *Procurement Budget* since an agency might have a limited budget but award several small contracts. Together, these measures proxy well agency's contracting activity.

Finally, most importantly, we account for the fact that if the agency falls short of its annual set-aside target, it might set aside more contracts. Conversely, if the agency exceeds its set-aside target, it might set aside fewer contracts. We control for this with *Set Aside Until t*, the cumulative sum of contracts set-aside value divided by the cumulative sum of all contracts value per agency (and per subagency) until the start of each contract i. Note that the given contract i is not included. Because of the possibility of time-invariant agency characteristics, i.e., a tendency to award or not award set-aside contracts, we include a lagged variable of the proportion of the previous FY, i.e., the proportion of set-aside contracts to all contracts awarded by a given agency (and subagency).

- *Contract characteristics.* Contract characteristics allow us to control for different levels of procurement heterogeneity possibly correlated with both the treatment and the outcome and that explains the agency's (conditionally discretionary) set-aside decision. For instance, we include time-related predictors, such as *LastWeek* and *LastMonth*, which indicate whether the contract is awarded in the last week or month of FY, respectively, and capture both the contract-level spending rush highlighted by the literature and worse performance (Liebman and Mahoney, 2017). *Period* is a dummy variable that equals 1 if the contract is awarded during the Great Recession, which in the US officially lasted from December 2007 to June 2009. This might affect procurement budgets considerations across agencies and outcomes. In addition, we include sector-related information, such as *Construction_maintenance* is a dummy variable indicating that the contract is for the performance of a construction project. With the variables *performancedivision* and *perfomanceregion*, we also control for the geographical location, i.e. in which region and division the contract is performed.

  With the same logic, we include for different proxies for the contract complexity. Such proxies are critical if more complex contracts are less (more) likely to be set aside, receive less (more) and incur more (less) often in renegotiations. For instance, we consider *Quartile Award Amount* and *Quartile Expected Duration* as proxy variables for contract complexity. The variables rank each contract in a quartile of the sector-specific distribution according to its expected cost and expected duration, respectively. Based on these two variables, we construct a comprehensive measure of complexity accounting for both the time and cost dimensions. *Complexity* is a categorical variable taking the value of one for contracts in the first and second quartile for *both* amount and duration. It is equal to three for the fourth quartile in *both* categories and equals to two in all other cases. Moreover, *MultiYear*, an indicator that equals 1 if the contract is expected to last longer than one year.

Finally, we control for a large number of contract characteristics thare are reported in FDPS, for instance, *Nationalinterest* indicates whether the contract was created for the national interest. This might be relevant, as such contracts might be less likely to be set-aside to, e.g., quickly limit damages provoked by hurricanes. We also observe whether for the contract some kind of environmental-related clause is applied. *Oversight* is a dummy variable that has a value of 1 if the contract is subject to oversight by the agency (in addition to the surety company on performance bonds if the contract is construction-related).[41]

The random forest The random forest is a machine learning method developed by Breiman (2001). Its goal is to predict outcomes based on the available covariates. For instance, it can be used to predict the selling price of a house given its characteristics, such as the number of rooms and the location. In this paper, we use the random forest to predict whether a given contract will be set aside based on the buyer, contract, and market characteristics. Random forests build on decision trees, which are composed of a series of yes/no questions leading to a class prediction for each observation—in this case, whether the contract is in the treatment or control group. We talk about classification trees if the outcome variable is binary, whereas regression trees are used to predict continuous variables. Since we use the random forest to predict the propensity score $p(X_i)$, where X_i is the vector of covariates for each tender for receiving a binary treatment, the random forest will aggregate several classification trees.

Important features of the trees are *nodes* and *branches*, whose meaning can best be illustrated with the help of Figure C1, where the treatment is SBSA. Starting from the green ellipse on top of the figure, "Before2011 = 1" is the first *node* or *split*. At this point, 100% of the sample is at this node, and 53% of the sample has received the treatment. If the statement is true, namely, if the contract is awarded before 2011, i.e. *Before2011* equals 1, then we move to the left *branch*. We reach the second node, where we find 61% of the sample, and 35% of it has received the treatment. At this node, we can split the sample by asking another question, namely, whether the given variable takes a value below 0.59. We continue until we have reached a final node at the bottom of the figure. In case the variable takes a value below 0.59 for a given contract, we predict that the contract will not be set aside.

Although decision trees can be a helpful tool for prediction, they can often lead to overfitting.

[41]See Carril et al. (2021), Calvo et al. (2019) and Giuffrida and Rovigatti (2022), who empirically investigate the effect of public and private supervision on contract outcomes.

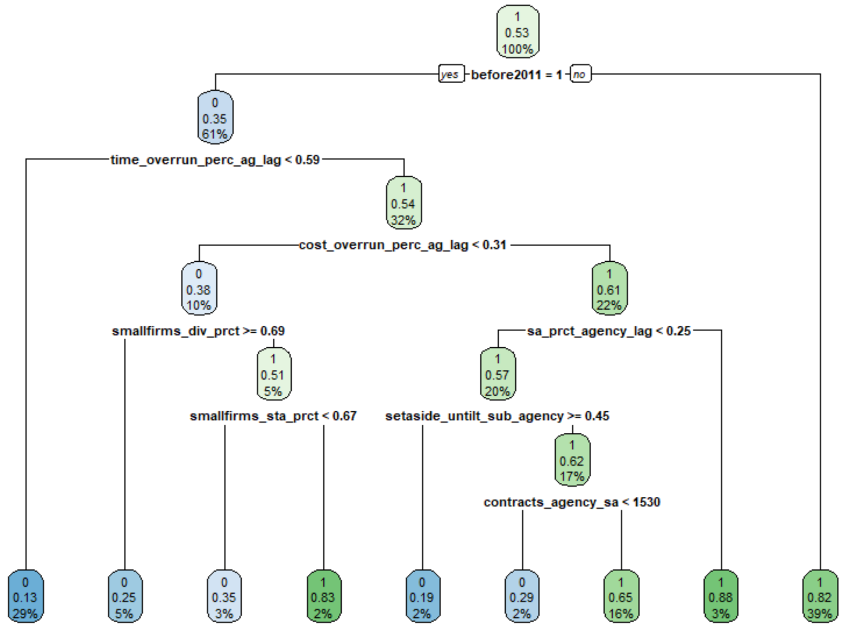


Figure C1: Example of a Classification Tree

Notes: This decision tree is obtained with the R package "rpart". The tree predicts the probability of obtaining the treatment for each observation (i.e. whether the contract is set aside). The Sample and variables are the same as for the random forest. However, this tree is only reported as an illustrative example to show how decision trees *might* be built. Random forests build on classification and regression trees but are more complex and less prone to overfitting. Hence, this tree is not necessarily representative of the trees obtained when running a random forest. For more precise information regarding the variables that are important for random forest, refer to Figure C2.

Random forests address precisely this concern since they work in the following way: several trees are grown using a different bootstrap sample of the data every time, and then, a majority vote is taken to predict the outcome of each observation. Random forests add an extra level of randomness, as they select a random subset of features at each split (or node). In a decision tree, the variable used at each node is the best *among all variables* in the data. Instead, in a random forest, only a subset of variables is randomly selected at each node, and the ones chosen are the best *among a subset of variables*.[42]

When growing each tree, we obtain an estimate of the error term, called *out-of-bag error*, by predicting the outcome for the observations not in the bootstrap sample (Breiman, 2001). This estimate is essential for the random forest, as it is used to determine the importance of each variable. Different ways exist to determine variable importance. In this paper, we use the minimal depth. The lower the number of minimal depth, the more important the variable. The number

[42]To illustrate these concepts with an example, imagine we grow 200 trees, and each of them will have a slightly different structure because (i) we use a different sample every time (since we bootstrap), and (ii) we randomly select different variables at every node, e.g., ten variables out of 100 available variables. Suppose 170 out of the 200 trees predict that contract i will be treated, whereas 30 trees predict that it will be untreated. The outcome would then predict that the contract is treated because the majority vote is used to predict the outcome.

indicates the depth of the node—the closer the node is to the root of the tree, the lower the minimal depth. Hence, a low minimal depth means that the variable splits many observations into two groups. In the example reported in Figure C1, *Before2011* would have depth 1.

Breiman and Cutler (2011) suggest that at least 1,000 up to 5,000 trees should be grown if there are many variables and if the researcher is interested in stable importance of variables. When growing more trees, the trade-off is incurring higher computational costs for achieving greater accuracy. Since the author does not provide more precise guidelines for setting the number of trees, we simulate propensity scores for growing 1,000 up to 5,000 trees, and we observe similar results in sign, magnitude, and significance levels in the baseline analysis. Finally, we decide to grow 1,000 trees, as we are interested in the determinants of the treatment but we also want to be parsimonious in terms of computational power.

For tuning the optimal number of randomly selected variables at each node, we follow again Breiman and Cutler (2011). The authors suggest trying different numbers of randomly selected variables combined with a relatively small number of trees. An exact number of trees is not provided. Thus, we use 200 trees. We train the data with a sample, which constitutes 20% of the original dataset, and we keep the remaining 80% aside for testing the error prediction. We choose three different numbers of variables to conduct three trials. First, we compute our number of variables as the square root of the total number of variables (i.e., 8 since we have 71 variables). Then, we use half as many variables, and finally, twice as many. For the baseline regressions, we obtain the lowest prediction error for 16 (8) variables for the SBSA (DBSA) treatment, so we use it throughout the analysis. Indeed, it has the highest accuracy, which is the percentage of correct out-of-sample prediction when using the fitted random forest. Note, however, that we estimate the propensity score for the three different numbers of randomly selected variables and obtain stable estimates of the ATT.[43] This shows that the results are not sensitive to the different specifications of the propensity score.[44]

First-stage results In this subsection, we identify key variables for the decision to restrict participation to targeted firms. Predicting the propensity score using random forests enables us to do so. In Figure C2, we report the distribution of the minimal depth for the ten variables that are selected most often in one of the first nodes of the tree. In other words, those are the most critical factors in predicting the treatment as a low number of minimal depth means that a lot of observations are divided into groups based on this variable.

[43]The results of this test are reported in Figure 3.

[44]We want to be extremely careful and make sure that the results are robust because they are sensitive to the different propensity score specifications when using IPW (Zhao, 2004).

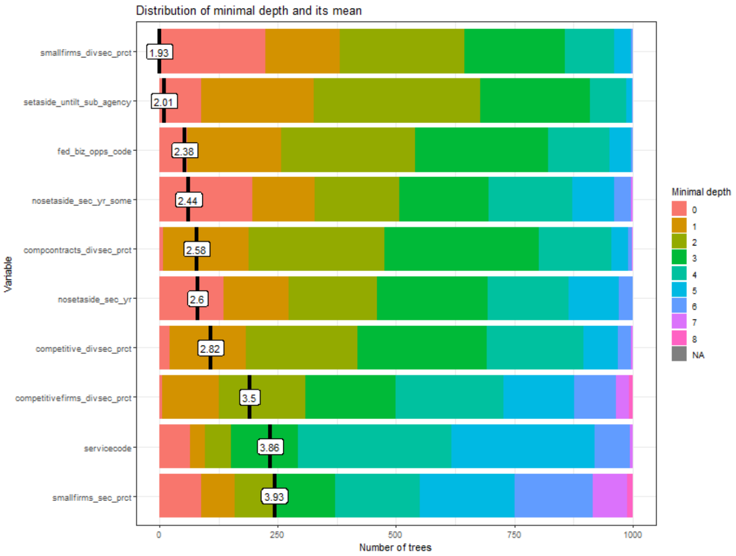


Figure C2: Variable Importance According to Minimal Depth

Notes: This figure reports the minimal depth for the most important variables used in the random forest.

We had previously argued that, following the FAR regulations, the the agency's decision to set-aside a contract is driven by two factors: (i) specific rules based on an assessment of potential competition between targeted firms and (ii) agency–year spending targets for set-aside contracts. Consistently with the regulation, we find that two variables that matter the most for a contract to be set aside are the share of small firms in a division/sector and divisions and the share of contracts awarded with setaside out of all contracts per subagency. Among the other variables, important factors are small-firm related: (a) Share of competitive contracts out of all contracts awarded to small firms, per division and year, (b) Share of competitive small firms out of all small firms, per division, sector and year, (c) the share of small firms out of all firms that sometimes win without set-aside, per sector and year, (d) the share of small firms out of all firms that always win without set-aside, per sector and year, and (e) the share of small firms per sector per year. The agency also considers the share of competitive contracts out of all contracts, by division and sector, per fiscal year and. Finally, some contract-characteristics are relevant, for instance whether the contract is publicized on FedBizOpps, an online government-wide platform where all contracts above $25,000 should be publicized and the one-letter code that best identifies the product or service procured, i.e., construction, quality control, information technology, and so on.

We must be careful when interpreting these results since it has been shown that random forests favor continuous variables over dichotomous ones. Hence, random forests may not be sensitive measures of variable importance (Strobl et al., 2007). This might be particularly relevant in the case of variable selection, which was, however, not the aim of the random forest employed here. To compute the propensity score, we are rather interested in the prediction of the treatment. In general, although other measures might be better, minimal depth is informative to understand how the propensity score is calculated and what factors are most decisive for the algorithm predictions.

Second stage: IPW estimator The second stage of our empirical strategy relies on two main assumptions. First, the *conditional independence assumption* restricts the dependence between the treatment model and the potential outcomes. In other words, it assumes the nonexistence of observables or unobservables that might influence selection into treatment and that are omitted from the model.

$$Y(0), Y(1) \perp\!\!\!\perp D \mid X, \forall X \tag{4}$$

Second, the *common support or overlap assumption* states that each individual in the sample could receive any treatment level and that we cannot *perfectly* predict the probability of receiving treatment.

$$0 < P(D = 1|X) < 1 \tag{5}$$

In Section V.1, we have shown that the common support assumption is satisfied given that covariates are balanced. Therefore, we are able to obtain unbiased estimators of the ATT. Note that for checking whether covariates are balanced, we use the standardized difference instead of computing t-statistics or testing the difference in means. The standardized difference "compares the difference in means in units of the pooled standard deviation" (Austin (2011), p. 412). According to Imbens and Rubin (2015), it is formally defined as follows:

$$\hat{\Delta}_{ct} = \frac{\bar{X}_t - \bar{X}_c}{\sqrt{\frac{\sigma_c + \sigma_t}{2}}} \tag{6}$$

For the weighted mean, we can substitute in Equation (6): $\bar{W}_{weight} = \frac{\sum \omega_i X_i}{\sum \omega_i}$, where ω_i is the weight for each unit (Austin and Stuart, 2015). The weighted sample variance is defined as:

$$s^2_{weight} = \frac{\sum \omega_i}{(\sum \omega_i)^2 - \sum \omega_i^2} \sum \omega_i (X_i - \bar{X}_{weight})^2 \tag{7}$$

We apply a weighted least squares regression using the inverse probability of the propensity score $p_i(X_i)$ as the weight in our high-dimensional fixed effect setting. For the treated contracts, i.e., those awarded by set-aside, we have $ipw_i = 1$, and for the untreated, i.e., those not awarded by set-aside, we have $ipw_i = \frac{p_i(X_i)}{1-p_i(X_i)}$. Assuming a linear form of the conditional mean, we apply a weighted least squares regression using the *ipw* weights. Thus, we can define $\hat{\tau}_{ATT}$—the IPW estimator for the ATT—as follows:

$$\hat{\tau}_{ATT} = N^{-1} \sum_{i=1}^{N} SetAside_i Y_i - \frac{(1 - SetAside_i)\, Y_i\, p_i(X_i)}{1 - p_i(X_i)}, \tag{8}$$

where X_i denote the vectors of covariates for each observation and Y_i is the outcome of interest.

D Appendix: Additional robustness checks

D.1 Robustness to methodology

In Table D1, we check whether our takeaways from the baseline analysis are sensitive to changes in methodology. Columns 1-3 report estimates on *Log(# Offers)*, columns 4-6 on *Extra Cost* and columns 7-9 on *Delay.* The top panel reports again the estimated coefficient for the SBSA, while the bottom one reports the estimated coefficients for the DBSA. Columns 1, 4 and 7 report baseline estimates from, respectively, columns 2, 4, and 6 in Table 2.

Table D1: Alternative Second-stage Methods

	Log(# Offers)			Extra Cost			Delay		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	IPW (RF)	PSM (RF)	IPW (logit)	IPW (RF)	PSM (RF)	IPW (logit)	IPW (RF)	PSM (RF)	IPW (logit)
SBSA	**0.213**	**0.298**	**0.261**	**0.035**	**0.021**	**0.078**	**0.047**	**0.058**	**0.142**
	(0.0112)	(0.0084)	(0.0068)	(0.0113)	(0.0081)	(0.0061)	(0.0292)	(0.0219)	(0.0171)
	0.837	0.754	0.875	0.253	0.234	0.231	0.952	0.921	0.927
	74,934	*106,432*	*81,716*	*74,425*	*105,730*	*81,194*	*74,928*	*106,418*	*81,708*
DBSA	**0.184**	**0.252**	**0.257**	**0.075**	**0.0540**	**0.098**	**0.099**	**0.083**	**0.221**
	(0.0189)	(0.0192)	(0.0104)	(0.0138)	(0.0191)	(0.0092)	(0.0467)	(0.0515)	(0.0255)
	0.856	0.761	0.837	0.268	0.247	0.245	0.995	0.960	0.947
	46,533	*66,337*	*56,804*	*46,242*	*65,956*	*56,483*	*46,529*	*66,331*	*56,799*
Buyer*FY FE	✓	✓	✓	✓	✓	✓	✓	✓	✓
Market Controls	✓	✓	✓	✓	✓	✓	✓	✓	✓
Contract Controls	✓	✓	✓	✓	✓	✓	✓	✓	✓

Notes: Results for the ATT of the IPW on the three main outcomes, *Log(# Offers)*, *Extra Cost* and *Delay.* We report, for each outcome, the baseline results as in Table 2 (i.e., IPW (RF)). Column 2, 5, and 8 reports the traditional propensity score matching (PSM) approach with the previously estimated propensity score. Note that Stata 17 user-written program `psmatch2` does not allow to specify both a varlist and a propensity score, so only the propensity score is specified. Column 3, 6, and 9 reports results from IPW estimated with the propensity score predicted with a logistic regression.

Using the same propensity score obtained estimated with random forest, we first test the robustness of the chosen second-stage estimator—i.e., the IPW. Columns 2, 5, and 8 report the same coefficients using a propensity score matching approach for which we employ a more traditional matching approach: A kernel matching with a biweight distance, which defines a neighborhood for each treated observation and constructs the counterfactual using all control observations within the neighborhood. The method assigns a positive weight to all those observations within the previously defined neighborhood and a zero weight to the remaining observations (Caliendo and Kopeinig, 2008).[45] Finally, in columns 3, 6, and 9, we run the IPW by calculating the propensity

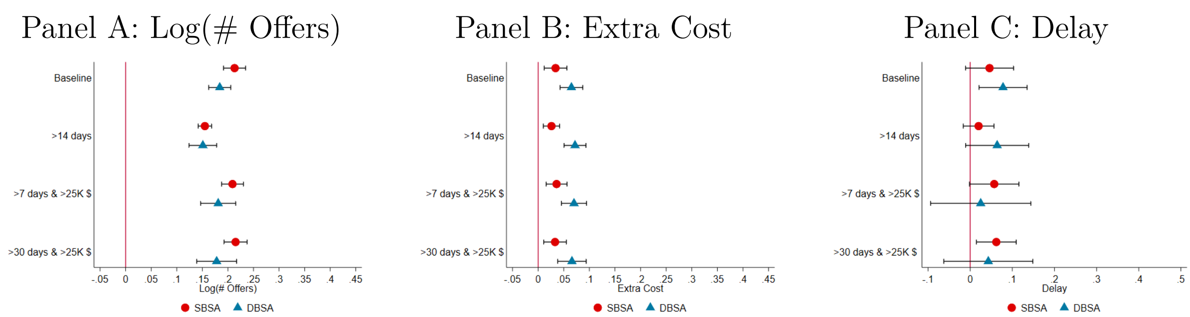
[45] To select the best method, we looked at kernel density graphs. We obtained similar results as after matching but kernel matching provided the best results. When looking at the standardized differences, kernel matching performed the best among our options. Nevertheless, it did not perform as well as IPW, since not all covariates are balanced. However, compared to IPW, PSM does a better job at achieving the same post-matching density for the propensity score. We still prefer IPW because it achieves a better covariance balance in terms of standardized differences, which is the key concept underlying matching approaches.

score with a classic logistic regression using the same set of variables employed for the random forest prediction. To sum up, on the one hand, the estimates confirm the sensibility of the choice of IPW since it does a better job of balancing the covariates without losing many observations. On the other hand, the results are qualitatively in line with different methods and specifications.

D.2 Robustness to sample selection

We then test the robustness of our results by using a different sample selection in terms of contract amount and duration, as reported in the right-side panels of Figure D1. We replicate the results for our preferred specification, with fixed effects and additional controls. We recall that our baseline sample described in Section IV includes contracts expected to be worth more than \$25,000 *and* last more than 14 days.

Figure D1: Robustness Checks, Sample



Notes: This figure reports the different robustness checks performed on our analysis. The panel title report the outcome variable of interest. The red dot reports the coefficient for small business set-asides (SBSAs) as treatment group, the blue triangle for disadvantaged business set-asides (DBSAs). The black confidence interval are at the 95% confidence level. The first line report the baseline coefficients, for contracts expected to last longer than 14 days and expected to cost over \$25,000. In the second line, we keep all contracts above 14 days, in the third line our sample is restricted to contract longer than 7 days and over \$25,000.

The first line reports the baseline estimates from Table 2; the second line includes all contracts above 14 days, irrespetive of their size; the third adds contracts between 7 and 14 days to the baseline sample; the fourth removes contracts between 14 and 30 days from the baseline sample. Note that we make the sample selection before predicting propensity scores, that is before running the random forest, which is therefore estimated again for each subsample. The results for *Log(# Offers)* are very stable and differ statistically only slightly when including all contracts above 14 days. For the outcome *Extra Cost* we find that the results both qualitatively and quantitatively almost identical. The same is also true for *Delay*, though we find a higher variability in the estimates, which is probably due to the presence of huge outliers in our data. In general, we can conclude that for all the three variables the results for the different samples are not statistically different and hence do not depend on the chosen specification. This is true for both treatments that we consider.

D.3 Robustness to alternative outcome specifications

As mentioned in Section IV, we also build two secondary outcomes, *Renegotiation* and *# Renegotiations*, which measure the extensive and intensive margins of contract amendment, respectively. *Renegotiation* is an indicator that takes the value 1 if we observe at least one *in-scope* modification that increases the final cost of a given contract.[46] Note that we consider a renegotiation as *in-scope* if the reason for the modification is consistent with the original contract terms.[47] *Out-of-scope* would be adding a new task to the contract in disregard of the original plan as to type, scope of work, period of performance, and method of performance.

For example, we consider a modification to be *out-of-scope* if it is an administrative or financing action, i.e., a modification to report a cash-only transaction. We calculate *Extra Cost* and *Delay* only from modifications within scope. Although the majority of contracts report *Extra Cost* either equal to zero or greater than zero, the variable can also take negative values. The same is true for the variable *Delay*, since we observe both positive and negative obligations (i.e., de-obligations). On the one hand, de-obligation has a positive effect as it implies a reduction in procurement costs. In turn, a reduction in procurement costs implies that more funds are available for other projects. On the other hand, de-obligations entail transaction and adjustment costs.

To account for transaction costs, we adjust *Renegotiation* when we construct the variables *Renegotiation* and *# Renegotiations*. We set them equal to 0 if we observe *only* de-obligations, and all of them are in-scope. In this way, we emphasize that the final cost has not increased. We set *Renegotiation* equal to 1 if we observe an obligation and de-obligation that are both in-scope. Given our goal of accounting for the intensive margin of renegotiation, the binary indicator should equal 1 because at least one negative modification has occurred.

To account for adaptation costs, we build the outcome variable *# Renegotiations*. This method is introduced by Bajari et al. (2014, p. 1289), who argue that renegotiation "generates adaptation costs in the form of haggling, dispute resolution, and opportunistic behavior." Therefore, we capture all adjustments and disruptions by counting all changes for a given contract, regardless of their monetary value, i.e., whether positive or negative. For an overall performance analysis, extra costs or total costs seem more relevant when renegotiations are allowed, while award amount may be preferred when renegotiations are rare events (as in the case of supplies). Moreover, one important caveat of this outcome for our set of service and construction contracts is that we do not consider homogeneous goods and lack a proper set of contracts for comparative purposes.

[46] Given the available data, we cannot proceed in the same way for an increase in final duration due to the construction of the date variable. Indeed, the dataset does not report the reasons of modification for an increase in final duration, it only reports expected end date and actual end date. The construction of these variables is similar to Calvo et al. (2019), Decarolis et al. (2020), Giuffrida and Rovigatti (2022), and Kang and Miller (2022).

[47] According to the FPDS data dictionary, this is the case when the reason for the modification is one of the following: "Supplemental Agreement for work within scope," "Change Order," "Terminate for Convenience," "Exercise an option," "Definitize letter order," or "Definitize change order." All other modifications are considered *out-of-scope*. Note that the previously cited papers use slightly different definitions of "in-scope modification." Nevertheless, the results are robust to such changes, as Decarolis et al. (2021) finds no different results when using the definition from Kang and Miller (2022).

Table D2: Additional Contract Outcomes

	Renegotiation [N,Y]		# Renegotiation	
	(1)	(2)	(3)	(4)
SBSA	**0.047**	**0.053**	**0.006**	**0.010**
	(0.007)	(0.007)	(0.009)	(0.009)
	0.333	0.359	0.309	0.357
	106,440	*74,940*	*106,440*	*74,940*
SESA	**0.057**	**0.048**	**-0.048**	**-0.063**
	(0.012)	(0.012)	(0.035)	(0.029)
	0.350	0.381	0.333	0.390
	66,342	*46,536*	*66,342*	*46,536*
Buyer*FY FE		✓		✓
Market Controls		✓		✓
Contract Controls		✓		✓

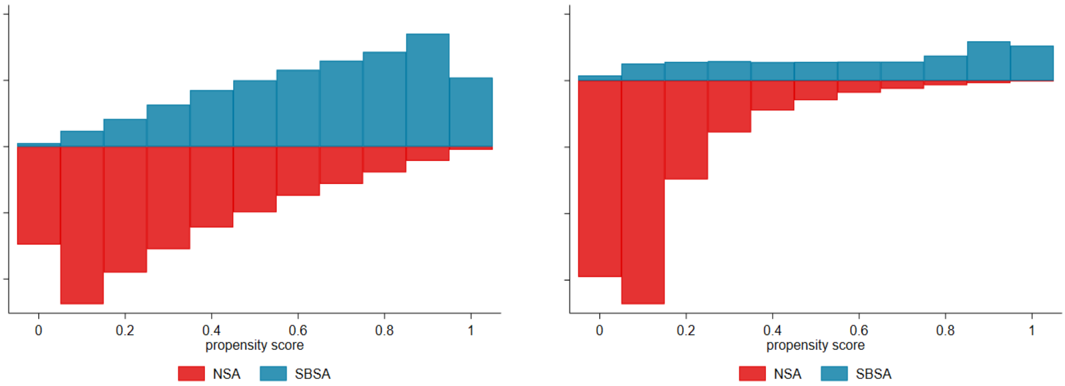
Notes: Results for the ATT of the IPW on two additional outcomes, namely *Renegotiation* and *# Renegotiation*. *Renegotiation* indicates extensively in-scope modifications of contract terms. *# of Renegotiations* variables report the number of in-scope contract modifications.

As we report in Table D2, we observe that set-asides cause the probability for a contract to undergo an amendment to increase by 5.3 and 4.8 p.p.. This is the most reliable point estimate as IPW best performs with dichotomous treatment and outcome variables. The coefficient on *# Renegotiations* is also positive, though not statistically significant for SBSA, while it is negative and significant for DBSA. This finding implies that contracts awarded under DBSA are more likely to be renegotiated, but such renegotiations tend to be larger in size, as we observe an increase in *Extra Cost* and *Delay* and a concomitant decrease in *# Renegotiations*. The results hold as compared to the baseline and tend to speak in favor of a deterioration in contract performance due to the set-aside's restriction to entry—in particular for SBSA—consistent with our baseline execution-stage findings.

E Appendix: Additional figures and tables

Figure E1: Propensity Score Distribution for the Treated and the Untreated

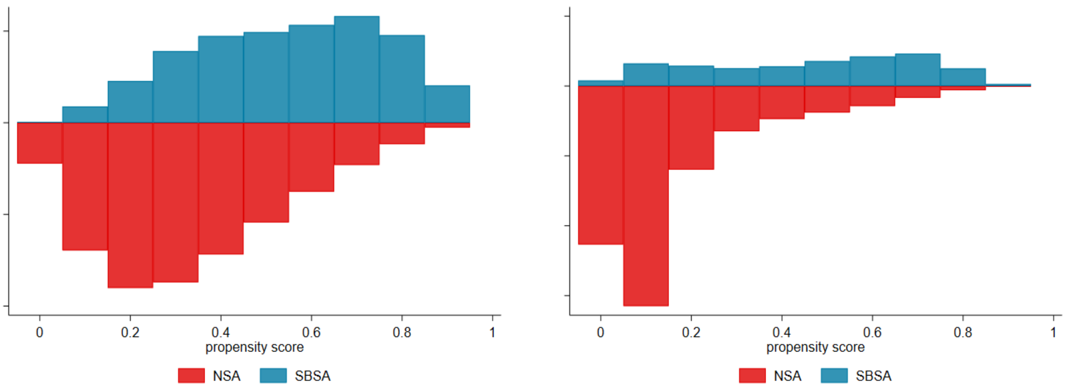
Panel A: SBSA Panel B: DBSA



Notes: This figure reports the distribution of treated and untreated observations according to the value of the propensity score. This figure provides some evidence that the overlap assumption is satisfied, since we do not observe values of the propensity score that are taken by one of the two groups.

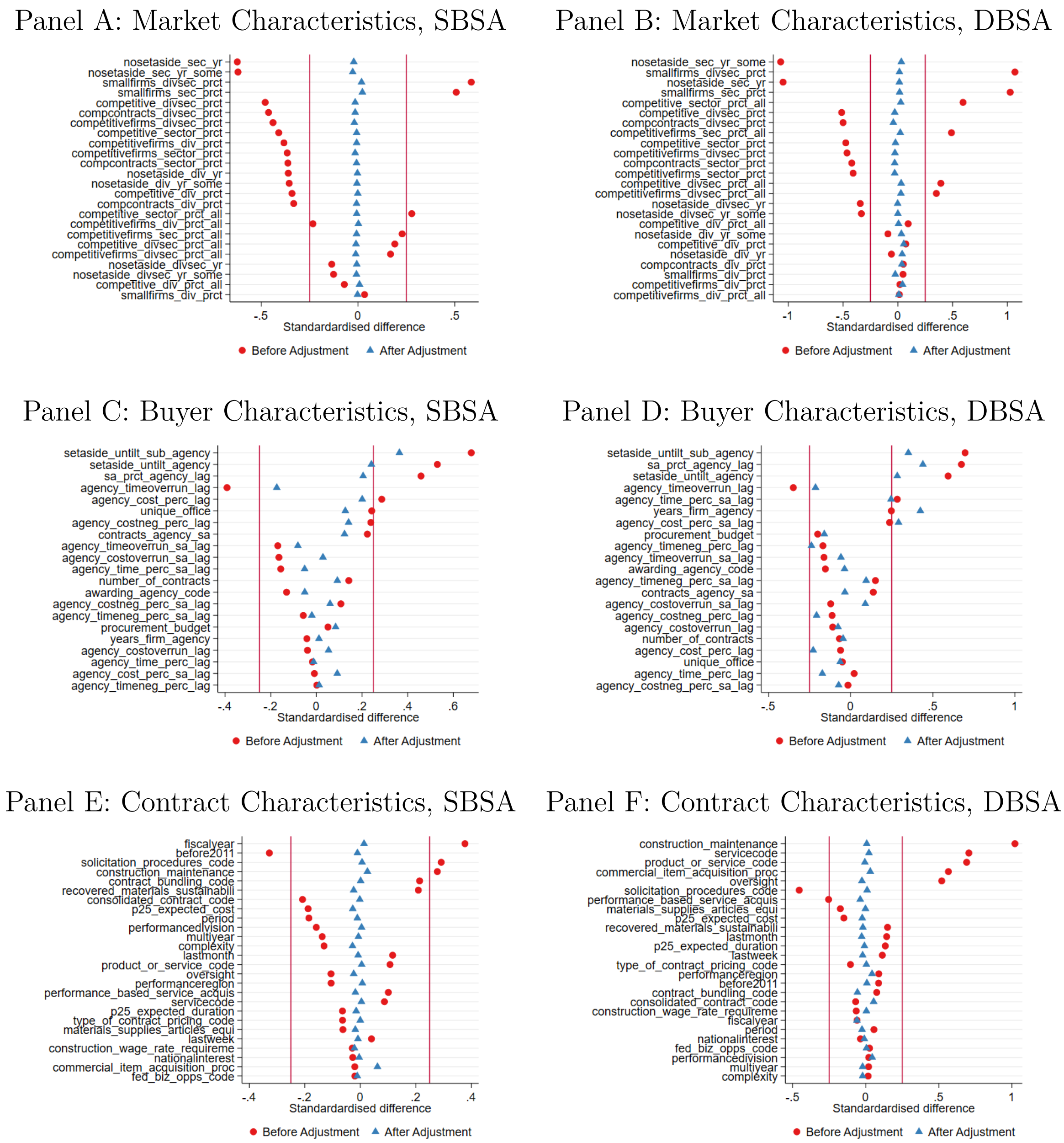
Figure E2: Propensity Score Distribution for the Treated and the Untreated, with Logit

Panel A: SBSA Panel B: DBSA



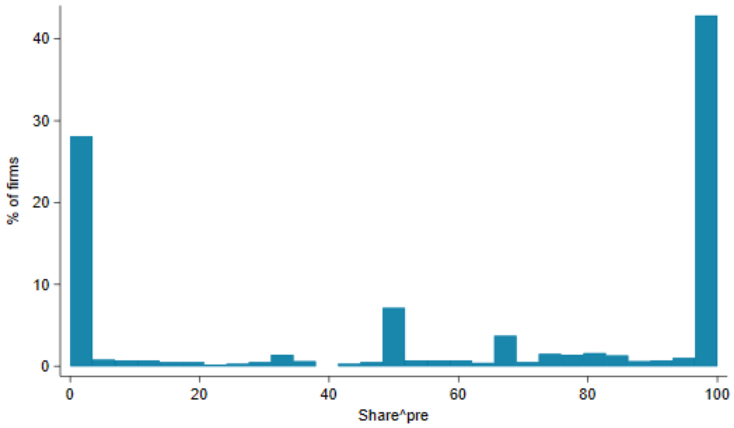
Notes: This figure reports the distribution of treated and untreated observations according to the value of the propensity score. The latter is estimated using logit. This figure provides some evidence that the overlap assumption is satisfied, since we do not observe values of the propensity score that are taken by one of the two groups.

Figure E3: Standardized Differences, with Logit



Notes: This figure reports the standardized differences before and after the IPW. The propensity score is estimated using logit. The treatment for this sample is specified in the panel title; the control is non set-aside contract. We predict the propensity score using 71 variables. We report variables in the three relevant groups: market, buyer and contract characteristics. The red vertical lines represent the -0.25 and 0.25 thresholds. Below this threshold, in absolute value, covariate balance is achieved (Imbens and Rubin, 2015). Standardized differences for all the covariates use, after adjustment, are below the absolute value of 0.25.

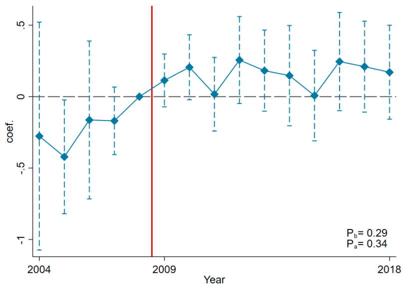
Figure E4: Firm Exposure to VBSA Spending Increase



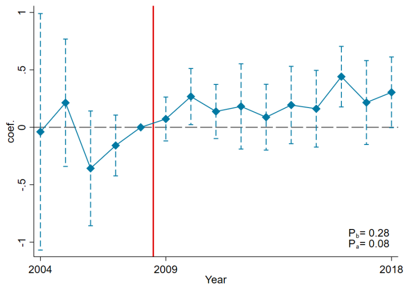
Notes: This figure reports the distribution of $Share_i^{pre}$, i.e., the annual average non-VBSA revenues out of total procurement revenues for VBSA incumbents between FY2004 and FY2008.

Figure E5: Event Study: # of States by US Region

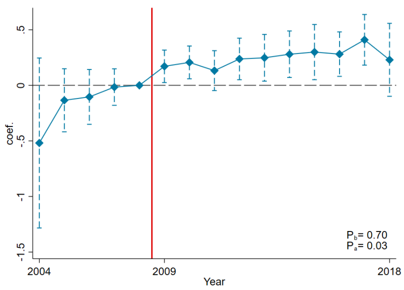
Panel A: Northeast



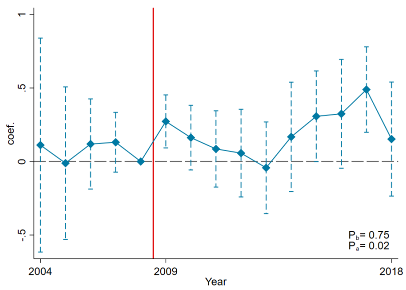
Panel B: Midwest



Panel C: South

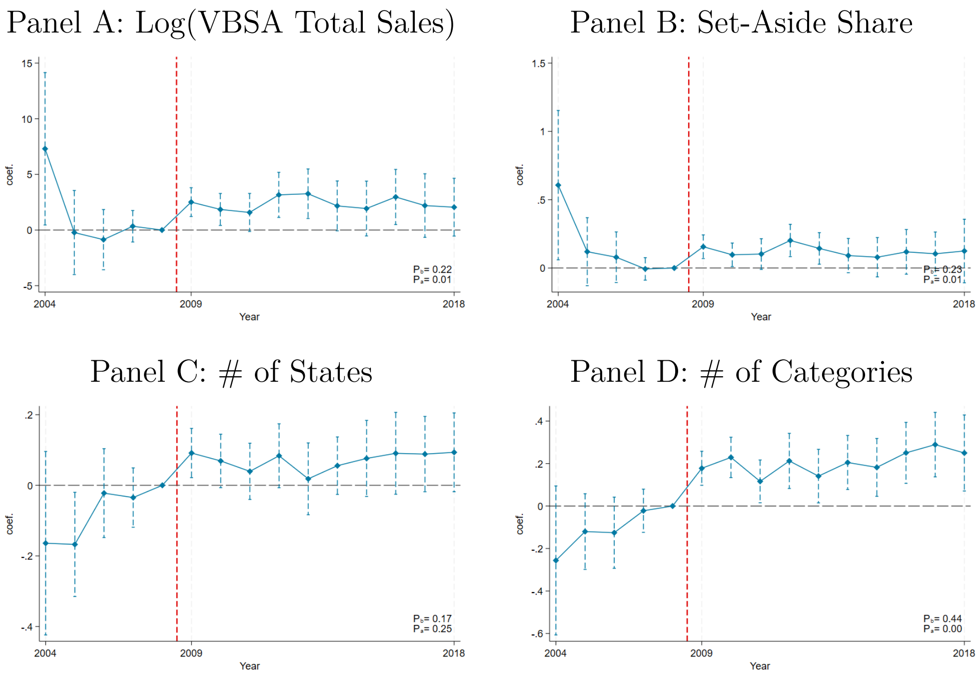


Panel D: West



Notes: Results for the event study analysis, reporting the coefficients of the interaction of $Share_{it}^{pre}$ with FY FEs. The policy was enacted in FY2009 (at time t), shown by the red vertical line. FY2008, time *t-1*, is chosen as the base year and all reported coefficients are relative to it. The dependent variables are constructed at the firm-year level. Panel A reports the log total sales to the government awarded through VBSA. Panel B shows the share of set-aside awards over total sales. Panel C and D report, respectively, the number of different procurement categories associated with its sales and the number of states in which it performs its activity. We report the 95 percent confidence intervals using standard errors clustered at the firm level. P_b reports the joint-significance of the coefficients in the before period, i.e. before FY2009, while P_a in the after period.

Figure E6: Event Study, Competitive Awards Only



Notes: Results for the event study analysis, reporting the coefficients of the interaction of $Share_{it}^{pre}$ with FY FEs. For this analysis, we include only competitive awards, as they are more unexpected. The policy was enacted in FY2009 (at time t), shown by the red vertical line. FY2008, time *t-1*, is chosen as the base year and all reported coefficients are relative to it. The dependent variables are constructed at the firm-year level. Panel A reports the log total sales to the government awarded through VBSA. Panel B shows the share of set-aside awards over total sales. Panel C and D report, respectively, the number of different procurement categories associated with its sales and the number of states in which it performs its activity. We report the 95 percent confidence intervals using standard errors clustered at the firm level. P_b reports the joint-significance of the coefficients in the before period, i.e. before FY2009, while P_a in the after period.

Table E1: VBSA vs NSA

	Log(# Offers)	Extra Cost	Delay
	(1)	(2)	(3)
SBSA	**0.296**	**0.081**	**-0.037**
	(0.0240)	(0.0296)	(0.0779)
	0.672	0.414	1.474
	5,523	*5,494*	*5,521*
Buyer*FY FE	✓	✓	✓
Market Controls	✓	✓	✓
Contract Controls	✓	✓	✓

Notes: Results for the ATT of the IPW on the three outcomes: (1) *Log(# Offers)* is the log of the number of offers received for a given contract, (2) *Extra Cost* indicates the share of the cost overrun relative to the expected cost, and (3) *Delay* measures the ratio of delay relative to the expected duration. The treated are contracts awarded with the service-disabled veteran-owned set-aside (VBSA). Several controls accounting for contract and market characteristics, and a buyer/fiscal-year fixed effect are included. The propensity score of the probability of being treated is separately predicted with the "randomForest" package in R. We report in bold the coefficient estimates, standard errors are in parentheses.The third row refers to the mean outcome for the control group and the fourth row (in italics) reports the number of observations.

Online appendix

DD (Blank) Form 2579 (DFARS §219.201(10)(B): Pages 1-2)

SMALL BUSINESS COORDINATION RECORD
(See DFARS PGI 253.219-70 for form completion instructions.)

1. CONTROL NO. *(Optional)*	**2. PURCHASE REQUEST/ REQUISITION NO.**	**3. TOTAL ESTIMATED VALUE** *(With options)*	**4a. PIID**	b. IDV PIID *(If applicable)*	**5. MOD/AMDMT NO.**

6a. CONTRACTING OFFICER NAME *(Last, First, Middle Initial)*	b. DODAAC	c. OFFICE SYMBOL
d. E-MAIL ADDRESS	e. TELEPHONE NUMBER *(Include Area Code)*	

7a. ITEM AND/OR SERVICE DESCRIPTION

b. PRODUCT OR SERVICE CODE	c. NAICS CODE	d. SIZE STANDARD

8. PERIOD OF PERFORMANCE/ DELIVERY DATES *(Including options)*

9. PURPOSE OF COORDINATION *(X one)*
☐ Initial Coordination ☐ Withdrawal ☐ Change

10. RECOMMENDATION *(X all that apply)*

☐ a. SMALL BUSINESS SET-ASIDE *(X one)*
☐ 100% ☐ Partial ________%

☐ b. SECTION 8(a) *(X one)*
☐ Competitive ☐ Sole Source

☐ c. HISTORICALLY UNDERUTILIZED BUSINESS ZONE (HUBZone) SMALL BUSINESS *(X one)*
☐ Competitive ☐ Sole Source

☐ d. SERVICE-DISABLED VETERAN-OWNED SMALL BUSINESS (SDVOSB) *(X one)*
☐ Competitive ☐ Sole Source

☐ e. ECONOMICALLY DISADVANTAGED WOMEN-OWNED SMALL BUSINESS (EDWOSB) SET-ASIDE

☐ f. WOMEN-OWNED SMALL BUSINESS (WOSB) ELIGIBLE UNDER WOSB PROGRAM SET-ASIDE

☐ g. OTHER SET-ASIDE *(Cite authority, e.g., FAR 26.202-1 or 6.208; or DFARS 226.71)*

☐ h. OTHER THAN FULL AND OPEN COMPETITION NOT PREVIOUSLY ADDRESSED

☐ i. FULL AND OPEN COMPETITION *(Complete block 13)*
☐ HUBZONE PRICE EVALUATION PREFERENCE *(Ref. FAR 19.1307)*

☐ j. MULTIPLE AWARD ☐ Contract ☐ Delivery/Task Order
☐ Reserves *(FAR 19.5) (List type(s) of small business, e.g., WOSB, SDVOSB)*

11a. MARKET RESEARCH/ACQUISITION PLAN

b. SYNOPSIS REQUIRED *(X one)*
☐ YES ☐ NO *(Provide FAR 5.202 exception)* ____________________
(NOTE: Synopsis not required if <$25,000; see FAR 5.101(a)(1).)

c. SMALL BUSINESS PROGRESS PAYMENTS *(X one)*
☐ YES ☐ NO

12. CONSOLIDATED OR BUNDLED *(X as applicable)*

a. CONSOLIDATED REQUIREMENT *(Attach required documentation per DFARS 207.170.)* ☐ YES ☐ NO

b. BUNDLED REQUIREMENT *(Attach required documentation per FAR 7.107 including benefit analysis.)* ☐ YES ☐ NO

13. SUBCONTRACTING PLAN REQUIRED *(X one)* ☐ YES ☐ NO

14. ACQUISITION HISTORY	
a. IS THIS A NEW REQUIREMENT? *(X one)* ☐ Yes *(Proceed to Block 15)* ☐ No *(Continue to Blocks a(1) through (10), marking all that apply for the immediately preceding acquisition.)*	
☐ (1) SMALL BUSINESS SET-ASIDE *(X one)* ☐ 100% ☐ Partial _______%	☐ (2) SECTION 8(a) *(X one)* ☐ Competitive ☐ Sole Source
☐ (3) HISTORICALLY UNDERUTILIZED BUSINESS ZONE (HUBZone) SMALL BUSINESS *(X one)* ☐ Competitive ☐ Sole Source	☐ (4) SERVICE-DISABLED VETERAN-OWNED SMALL BUSINESS (SDVOSB) *(X one)* ☐ Competitive ☐ Sole Source
☐ (5) ECONOMICALLY DISADVANTAGED WOMEN-OWNED SMALL BUSINESS (EDWOSB) SET-ASIDE	☐ (6) WOMEN-OWNED SMALL BUSINESS (WOSB) ELIGIBLE UNDER WOSB PROGRAM SET-ASIDE
☐ (7) OTHER SET-ASIDE *(Cite authority, e.g., FAR 26.202-1 or 6.208; or DFARS 226.71)*	☐ (8) OTHER THAN FULL AND OPEN COMPETITION NOT PREVIOUSLY ADDRESSED
☐ (9) FULL AND OPEN COMPETITION *(Complete block 13)* ☐ HUBZONE PRICE EVALUATION PREFERENCE *(Ref. FAR 19.1307)*	☐ (10) MULTIPLE AWARD ☐ Contract ☐ Delivery/Task Order ☐ Reserves *(FAR 19.5) (List type(s) of small business, e.g., WOSB, SDVOSB)*
b. PREVIOUSLY CONSOLIDATED OR BUNDLED? *(X one)* (1) CONSOLIDATED ☐ YES ☐ NO (2) BUNDLED ☐ YES ☐ NO	
c. DETAILS OF PREVIOUS AWARD(S) *(List details requested in instructions. Attach additional page(s) if necessary.)*	

15. CONTRACTING OFFICER	
a. NAME *(Last, First, Middle Initial)*	b. E-MAIL ADDRESS
c. SIGNATURE	d. DATE SIGNED *(YYYYMMDD)*

16. SMALL BUSINESS PROFESSIONAL/SMALL BUSINESS DIRECTOR REVIEW		
☐ Concur ☐ Non-concur	a. NAME *(Last, First, Middle Initial)*	b. E-MAIL ADDRESS
c. SMALL BUSINESS PROFESSIONAL/SMALL BUSINESS DIRECTOR REMARKS		
d. SIGNATURE	e. DATE SIGNED *(YYYYMMDD)*	f. DATE ACQUISITION PACKAGE PROVIDED TO SBA *(FAR 19.202-1(e)) (YYYYMMDD)*

17. SBA PROCUREMENT CENTER REPRESENTATIVE REVIEW		
☐ Concur ☐ Non-concur	a. NAME *(Last, First, Middle Initial)*	b. E-MAIL ADDRESS
c. SBA PROCUREMENT CENTER REPRESENTATIVE REMARKS		
d. SIGNATURE		e. DATE SIGNED *(YYYYMMDD)*

18. CONTRACTING OFFICER REVIEW		
a. CONTRACTING OFFICER REMARKS		
☐ Concur with PCR recommendation ☐ Reject PCR recommendation	d. SIGNATURE	e. DATE SIGNED *(YYYYMMDD)*

Full list of treatment predictors

We report the full list of the variables used in the first stage. We report a star (*) next to the variable if the variable was present in FPDS. All other variables are constructed on own calculations using the information provided by FPDS.

Market characteristics

- What share of contracts is competitive?
  - *compcontracts_div_prct*: Share of competitive contracts out of all contracts, by division, per fiscal year.
  - *compcontracts_sector_prct*: Share of competitive contracts out of all contracts, by sector, per fiscal year.
  - *compcontracts_divsec_prct*: Share of competitive contracts out of all contracts, by division and sector, per fiscal year.
- What share of contracts won by small firms is competitive?
  - *competitive_div_prct*: Share of competitive contracts out of all contracts awarded to small firms, per division and year.
  - *competitive_sector_prct*: Share of competitive contracts out of all contracts awarded to small firms, per sector and year.
  - *competitive_divsec_prct*: Share of competitive contracts out of all contracts awarded to small firms, per division, sector and year.
- What share of competitive contract is won by small firms?
  - *competitive_div_prct_all*: Share of contracts won by small firms out of competitive contracts, per division and year.
  - *competitive_sector_prct_all*: Share of contracts won by small firms out of competitive contracts, per sector and year.
  - *competitive_divsec_prct_all*: Share of contracts won by small firms out of competitive contracts, per division, sector and year.
- What share of small firms is competitive?
  - *competitivefirms_div_prct*: Share of competitive small firms out of all small firms, per division and year.
  - *competitivefirms_sector_prct*: Share of small competitive firms out of all small firms, per sector and year.
  - *competitivefirms_div_prct*: Share of competitive small firms out of all small firms, per division, sector and year.
- What share of competitive firms is small?
  - *competitivefirms_div_prct_all*: Share of competitive small firms out of all competitive firms, per division and year.
  - *competitivefirms_sec_prct_all*: Share of small competitive firms out of all competitive firms, per sector and year.

- *competitivefirms_divsec_prct_all*: Share of competitive small firms out of all competitive firms, per division, sector and year.

- What share of firms is small and win without set-aside? (always or sometimes)
  - *nosetaside_div_yr*: Share of small firms out of all firms that always win without set-aside, per division and year.
  - *nosetaside_div_yr_some*: Share of small firms out of all firms that sometimes win without set-aside, per division and year.
  - *nosetaside_sec_yr*: Share of small firms out of all firms that always win without set-aside, per sector and year.
  - *nosetaside_sec_yr_some*: Share of small firms out of all firms that sometimes win without set-aside, per sector and year.
  - *nosetaside_divsec_yr*: Share of small firms out of all firms that always win without set-aside, per division, sector and year.
  - *nosetaside_divsec_yr_some*: Share of small firms out of all firms that sometimes win without set-aside, per division, sector and year.
- What is the share of small firms?
  - *smallfirms_div_prct*: Share of small firms per division per year.
  - *smallfirms_sec_prct*: Share of small firms per sector per year.
  - *smallfirms_divsec_prct*: Share of small firms per division and sector per year.

List of buyer characteristics

- The awarding agency:
  - *awarding_agency_code*: Code identifying the awarding agency.
- Experience with cost renegotiations and delay in the previous fiscal year for each agency. For each variable type, we distinguish between set-aside and all contracts:
  - *agency_cost_perc_lag*: Lag of ratio of contracts with a cost overrun out of all contracts by agency per year.
  - *agency_cost_perc_sa_lag*: Lag of ratio of set-aside contracts with a cost overrun out of all set-aside contracts by agency per year.
  - *agency_costneg_perc_lag*: Lag of ratio of contracts with a negative cost overrun out of all contracts by agency per year.
  - *agency_costneg_perc_sa_lag*: Lag of ratio of set-aside contracts with a negative cost overrun out of all set-aside contracts by agency per year.

- *agency_costoverrun_lag*: Lag of the absolute cost overrun of the agency divided by the absolute sum of the expected duration by agency per year.
- *agency_costoverrun_sa_lag*: Lag of the absolute cost overrun of the agency divided by the absolute sum of the expected duration by agency per year, calculated for set-aside contracts only.
- *agency_time_perc_lag*: Lag of ratio of contracts with a delay out of all contracts by agency per year.
- *agency_time_perc_sa_lag*: Lag of ratio of set-aside contracts with a delay out of all set-aside contracts by agency per year.
- *agency_timeneg_perc_lag*: Lag of ratio of contracts with a negative delay out of all contracts by agency per year.
- *agency_timeneg_perc_sa_lag*: Lag of ratio of set-aside contracts with a negative delay out of all set-aside contracts by agency per year.
- *agency_timeoverrun_lag*: Lag of the absolute delay of the agency divided by the absolute sum of the expected duration by agency per year.
- *agency_timeoverrun_sa_lag*: Lag of the absolute delay of the agency divided by the absolute sum of the expected duration by agency per year, calculated for set-aside contracts only.

- Agency size:
  - *Contracting Offices*: Number of distinct contracting units within an agency.
  - *Contracts*: agency's total yearly number of contracts awarded for all purchase categories.
  - *Procurement Budget*: agency's total yearly procurement spending for all purchase categories.
- Percentage of set-aside in the previous year:
  - *sa_prct_agency_lag*: Ratio of the previous fiscal year, i.e. set-aside contracts out of all contracts awarded by a given agency.
- Set-aside contracts until that award, a proxy for pressure to set-aside a contract to reach the agency goal:
  - *contracts_agency_sa*: Cumulative count of set-aside contracts grouped by awarding agency and year (and sorting by action date).
  - *setaside_untilt_agency*: Cumulative sum of contracts awarded with set-aside out of the cumulative sum of all contracts per agency (contracts are sorted by awarding agency and starting date).
  - *setaside_untilt_sub_agency*: Cumulative sum of contracts awarded with set-aside out of the cumulative sum of all contracts per subagency.

- Number of years in which the same firms wins with same agency, a proxy for relational contracts:
  - *years_firm_agency*: Number of years in which the same firm wins with the same agency.

Contract characteristics

- Time dimension:
  - *before2011*: Indicator equal to 1 if the contract was awarded before 2011.
  - *fiscalyear*: Year of the contract award. The fiscal year starts October 1st each calendar year.
  - *lastmonth*: Indicator variable equal to one if the contract was awarded during the last month of the fiscal year, i.e. September.
  - *lastweek*: Indicator variable equal to one if the contract was awarded during the last week of September.
  - *period*: This indicator equals to one if the contract was awarded after December 2007 and before June 2009. It accounts for contracts awarded during the Great Recession.
- Geographical dimension:
  - *performancedivision*: Indicates each of the nine divisions of the place of performance of the contract, i.e., New England, Middle Atlantic, East North Central, West North Central, South Atlantic, East South Central, West South Central, Mountain, Pacific.
  - *performanceregion*: Indicates each of the four regions (West, South, Midwest, and North East) of the place of performance of the contract.
- Sector-related variables:
  - *construction_maintenance*: An indicator variable equal to one if the contract is based on the execution of a construction or maintenance project.
  - *product_or_service_code*: The code that best identifies the product or service procured.
  - *servicecode*: The code that best identifies the product or service procured, but only considering the first letter of the code, such as construction, quality control, information technology, and so on.
- Proxies for complexity:
  - *complexity*: A categorical variable taking the value of one for contracts in the first and second quartile for *both* amount and duration. It is equal to three for the fourth quartile in *both* categories and equals to two in all other cases.

- *multiyear*: Indicator variable equal to one if the expected duration is over 365 days.
- *oversight*: An indicator variable equal to one if the contract is subject to oversight by the buyer (in addition to the surety company when the contract is construction-based). Expected costs of above $100,000 was the threshold to apply for oversight until 2011, and expected costs of above $150,000 was the threshold after 2011.
- *p25_expected_cost*: This variable is a proxy of the complexity of the contract. It places the contract into a quartile according to the expected cost grouped by sector.
- *p25_expected_duration*: This variable is a proxy of the complexity of the contract. It places the contract into a quartile according to the expected duration grouped by sector.

• Other controls for contract heterogeneity:

- *commercial_item_acquisition_pr*: Indicates whether the solicitation meets the special requirements for the acquisition of commercial items intended to more closely resemble those customarily used in the commercial marketplace as defined by FAR Part 12.
- *consolidated_contract_code*: Indicates whether the contract is a consolidated contract. This is only 'True' if the Funding Agency or the contracting agency is a Department of Defense Agency.
- *construction_wage_rate_requireme*: An indicator equal to one if the transaction is subject to the Construction Wage Rate Requirements. The latter states that "all laborers and mechanics employed or working upon the site of the work will be paid unconditionally and not less often than once a week, and without subsequent deduction or rebate on any account."
- *contract_bundling_code*: Designates that the value of the contract, including all options, is expected to exceed the threshold whose value is: (1) $5 million until 09/27/2006, (2) $5.5 million from 09/28/2006 to 09/30/2010, (3) $6 million from 10/01/2010. It indicates the reason why the agency bundled contract requirements. 'Bundling' refers to the consolidation of two or more requirements for goods or services previously provided or performed under separate smaller contracts into a solicitation for a single contract that is likely to be unsuitable for award to a small business.
- *fed_biz_opps_code*: Description tag (by way of the FPDS Atom Feed) that explains the meaning of the code provided in the FedBizOpps Field.
- *materials_supplies_articles_eq*: Description tag (by way of the FPDS Atom Feed) that explains the meaning of the code provided in the Contracts for Materials, Supplies, Articles, and Equipment Exceeding $15,000 Field.
- *nationalinterest*: Indicator variable equal to one if the contract is created for the national interest, e.g. projects to limit damages provoked by hurricanes.
- *performance_based_service_acqu*: This variable describes the requirements in terms of results required rather than the methods of performance of the work.

- *recovered_materials_sustainabili*: Designates whether Recovered Material Certification and/or Estimate of Percentage of Recovered Material Content for EPA-Designated Products clauses were included in the contract. For instance, if the contract considered energy efficiency in the award.
- *solicitation_procedures_code*: Report this code for the type of solicitation procedure used.
- *type_of_contract_pricing_code*: The type of contract as defined in FAR Part 16 that applies to this procurement. As we keep only fixed price contracts, these are different types of fixed price, i.e., award fee, incentive, etc.