

DISCUSSION

// NO.22-030 | 07/2022

DISCUSSION PAPER

// MATILDE CAPPELLETTI AND LEONARDO GIUFFRIDA

Targeted Bidders in Government Tenders

TARGETED BIDDERS IN GOVERNMENT TENDERS*

Matilde Cappelletti^{1,2} and Leonardo M. Giuffrida^{2,3}

¹University of Mannheim

²ZEW Mannheim

³CESifo and MaCCI

July 13, 2022

A set-aside restricts participation in procurement contests to targeted firms. Despite being widely used, its effects on actual competition and contract outcomes are ambiguous. We pool a decade of US federal procurement data to shed light on this empirical question using a two-stage approach. To circumvent the lack of exogenous variation in our data, as a first step we draw on random forest techniques to calculate the likelihood of a tender being set aside. We then estimate the effect of restricted tenders on pre- and post-award outcomes using an inverse probability weighting regression adjustment. Set-asides prompt more firms to bid—that is, the increase in targeted bidders more than offsets the loss of untargeted. During the execution phase, set-aside contracts incur higher cost overruns and delays. The more restrictive the set-aside, the stronger these effects. In a subset of our data we leverage an expected spike in set-aside spending and we find no evidence of better performance by winners over a ten-year period.

JEL: D22; H32, H57; L25.

Keywords: small businesses, set-aside, competition, procurement, public contracts, random forest, firm dynamics.

*Cappelletti: matilde.cappelletti@zew.de; Giuffrida: leonardo.giuffrida@zew.de. We would like to thank Rodrigo Carril, Decio Coviello, 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 and the ECOBUS Seminar Series where earlier versions of this study were presented. Financial support from the Leibniz SAW project “Market Design by Public Authorities” is gratefully acknowledged. Cheng Chen, Julia Schepp, and Giovanni D’Ambrosio provided excellent research assistance.

I Introduction

Policymakers traditionally view high competition in public procurement as an effective way to increase the value of taxpayers' money and, consequently, view a lack of adequate competition as problematic (e.g., EC, 2017). Regulations thus tend to encourage open procedures in order to boost firm participation. Nonetheless, unrestricted tenders can backfire if the playing field is not levelled among potential suppliers, as large firms tend to have lower operating costs—due to greater experience and economies of scale—and are consequently more price competitive. Combined with the red tape and entry costs embedded in public contracting, these dynamics may deter small firms from winning contracts and, due to the massive weight of public procurement spending, have cascading effects for the structure of the economy. Governments recognize this trade-off inherent to open procedures and explicitly intervene with ad hoc programs. This is because small businesses are highly relevant to the economy, accounting for 99% of all businesses and two-thirds of employment in OECD countries (OECD, 2019).

One means of ensuring the adequate participation of small businesses in the government procurement market and achieving distributional goals is to exclude tout-court large firms from participating in specific tenders, i.e., the so-called “small business set-aside”. Although set-aside is a policy pursued by many governments, there is no conclusive evidence on its effectiveness or overarching impact on the government budget and targeted firms (Athey et al., 2013; Denes, 1997; Marion, 2009, 2007; Nakabayashi, 2013; Tkachenko et al., 2019). This paper aims to fill these gaps by examining the impact of set-asides across an entire procurement market as well as the long-term effects on the performance of the recipient firm. Specifically, we aim to jointly analyze two sets of responses to set-asides at different stages of the procurement process. First, we examine the implications at the award stage in terms of the actual number of bids. Answering this question is not as straightforward as it initially appears. On the one hand, restricting competition prevents larger and more efficient firms from entering the market and reduces competition. On the other hand, targeted firms that would have not competed for the same contract in an open tender may perceive higher probability of winning with a set-aside and place a bid. The net response depends on potential competition and is therefore an empirical question. Second, we investigate whether the program causes a deterioration in the execution of awarded contracts. Regardless of which competition channels dominate, the implications for the execution stage of the contract are ambiguous due to supplier selection and bidding behavior response. We follow the related literature (e.g., Decarolis, 2014; Hart and Moore, 1988; Herweg and Schwarz, 2018) and study the effect of set-asides on delays and extra costs—relative to the contract amount and expected duration, respectively—as the main metrics for assessing the final contract outcomes. The rationale is that a contract amendment captures Williamson (1971)'s transaction costs which are at play irrespective of the reason behind the renegotiation and which have been shown to be economically sizeable for public contracts (Bajari et al., 2014; Spiller, 2008). Consistently with the negative association between competition and contract enforcement depicted by Spulber (1990), our results show that set-asides jointly increase the actual number of bids and induce higher costs overruns and more delays. We refer to the first and second sets of results as a short-term analysis because they focus on single procurement processes.

Third, in an additional analysis, we provide suggestive evidence on the long-run implications of a specific category of set-aside on the behavior of the exposed firms in terms of their future contracting performance.

The setting for this study is US federal procurement, which has the largest and longest-standing set-aside program worldwide, as the government aims to allocate 23% of annual procurement budget—i.e., \$100 to 150 billion—to small businesses. For this study, we rely on the universe of service and construction contract-level data from 2008 through 2018. Our goal is to estimate, in a reduced-form fashion, the effect of limiting participation to two different groups of targeted firms, i.e., small businesses (under the small business set-aside or SBSA) or small *and* disadvantaged businesses (under the socio-economic set-aside or SESA), on the above-mentioned outcomes with respect to a counterfactual scenario without a set-aside. To infer causality, we address concerns related to a potential selection bias arising from the contracting officer’s discretion on whether to make use of set-asides. Thus, if contracting officers who award more set-asides are also dealing with less complex contracts, normally associated with less renegotiations, they could systematically report better ex-post outcomes.

To circumvent the lack of quasi-experimental variation, our approach consists of two stages. In the first stage, we employ a machine learning approach to predict the propensity scores of set-aside tenders. Specifically, we improve the propensity score calculation using random forest techniques, whose greatest merit is the possibility to better exploit the wealth of our data for prediction, with reduced overfitting concerns. This is confirmed by the out-of-sample prediction accuracy of the treatment. Accuracy of the treatment prediction is 76% for the random forest and only 52% for the logit, which is the most common method to estimate the propensity score.¹ Moreover, being the random forest a non-parametric estimator, the risk of model misspecification is reduced. In the second stage, we use an inverse probability weighting regression adjustment (IPWRA) to estimate the impact of set-aside tenders on the aforementioned procurement outcomes in a high-dimensional fixed effect setup. The inverse probability weighting employs the propensity scores as weights and removes the differences in observable characteristics while giving more weight to those observations with a low probability of receiving the treatment—i.e., the tender being set aside—and vice versa. It also allows us to include in the second stage a battery of fixed effects that are key for identification. Leaving them out of the analysis could lead to additional bias from omitted variables. We believe that our chosen method is the best and most accurate for estimating treatment effects in our setting.

We find that SBSAs increase cost overruns by 2.7 p.p. (+13%) and delays by 5.1 p.p. (+5.8%), while SESAs are more disruptive: cost overruns and delays increase by 7.3 p.p. (+35%) and 5.6 p.p. (+6.4%), respectively. SBSAs increase the number of bids by 14.1% (i.e., 0.5 more bidders) while SESAs have a slightly higher impact on competition of 19.2%—i.e., 0.7 more bidders. The results on the post-award performance hold with a binary variable for renegotiation and the number of renegotiations as well. All in all, the impact of set-asides on government service and construction budgets is negative when targeting small (and disadvantaged) firms. The results are also stable when the winning firm is fixed: We argue that it is not competition *per se* that affects the differential performance of firms, but the quality of competitors and their bidding behavior, which forces winners to behave differently vis-à-vis in open procedures. This

¹We explain this test in more details in Section VI.1.

pattern can motivate the fiercer disruption induced by SESAs, which is more restrictive than SBSAs for participation criteria. Indeed, the increase in the number of bidders is most likely driven by those firms that do not normally participate in open procurement auctions. Although we cannot observe in the data whether such bidders refrain from participating in unrestricted contests, it is reasonable to hypothesize that they are different from firms that normally participate in open tenders. To explore further the winner's characteristics as drivers of the results, we check whether the winner in set-asides is i) more likely to be located out of the performance state of the contract execution and ii) a first-time winner. The latter proxies for the lack of experience, which could explain their bidding behavior. However, we find no evidence of this channel. Instead, we find that the set-aside winners tend to be different in that they are based in another state. This is suggesting that, as hypothesized, the pool of participating firm is different under set-asides.

Then, we explore whether this spending inefficiency might reflect into improved performance in the procurement market by targeted firms in the long-run—an effect unexplored in the literature. In fact, the stated overarching goal of the set-aside program is to secure small (and disadvantaged) businesses with a share of procurement spending so that they can scale up, thrive, and then compete on a leveled playing field with other businesses down the road. Essentially, we want to provide a preliminary evidence on whether the US Set-Aside program might be considered an industrial policy or a subsidy policy. We restrict our analysis to construction projects awarded by the Department of Veteran Affairs to exploit a massive, unexpected increase in spending occurred in 2009-2010 in a particular set-aside program—i.e., the Veteran-Disabled Owned Business set-aside—for targeted firms. We find that incumbent firms are exposed to the increase in demand demand—compared to those who do not—perform far worse in future procurement activities. They win fewer contracts, rely increasingly on set-aside bidding, work increasingly locally, and sell fewer varieties of products and services. Although none of these results is a perfect test in isolation, we believe that in combination they provide suggestive evidence that the program generates short-term, contract-level inefficiency associated with long-term, firm-level inefficiency—at least in the businesses' activity in the government market.

This paper is organized as follows. Section II presents the relevant literature and highlights the contribution of this paper. Section III provides key institutional information on US government's set-aside programs. In section IV, we outline a theoretical background to our research. Section V presents and describes the data. The identification concerns and strategy are explained in Section VI. The following section reports our findings and discusses their robustness and drivers. Section VIII displays our empirical exercise for assessing the long-run implications of set-asides for target firms. Finally, section IX concludes.

II Literature Review

This paper builds on a large empirical literature examining the relationship between competition and procurement outcomes. Coviello et al. (2017) provide contrasting evidence to the common belief that open competition is the best tool for boosting value for taxpayers. They find that increasing buyers' discretion on winner selection has a neutral and sometimes even positive effect on procurement outcomes, although it

increases the chances of the same firms winning again. Kang and Miller (2022) observe little competition in US federal procurement auctions and, in line with Coviello et al. (2017), find an increase in costs when public officials' discretion is reduced. Calzolari and Spagnolo (2009) provide a theoretical background to explain these empirical results. They find that buyers optimally choose constrained auctions that threaten exclusion of suppliers with poor past 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. Recently, Tukiainen et al. (2021) proposed a choice experiment with real bureaucrats to isolate their preferences for desired levels of competition from its instrumental role of identifying of non-performing suppliers, as highlighted in the papers above. Public procurers do actually value some competition but have non-linear preferences that dissipate with additional bid submissions. A paper close to ours is Carril et al. (2022), which studies the relationship between tender publicity, actual competition, and post-award performance across a variety of heterogeneous goods and services. Carril et al. (2022) show shows that publicity increases competition, but its benefits and costs depend on the complexity of the purchased good or service.² Our work examines the trade-off between competition and outcome from a different perspective. Our results show that set-aside policies, which impose a mechanical constraint on participants in competitive bidding, are associated with an actual increase in participants—most likely from disadvantaged potential entrants, who offset the loss of excluded advantaged firms—and a deterioration in ex-post outcomes.

Our paper also relates closely to studies that examine the effects of policies that mechanically alter competition in public tenders. We can categorize such policies into two categories: Bidding preferences (Ayres and Cramton, 1995; Krasnokutskaya and Seim, 2011; Marion, 2007, 2009) and set-aside programs (Athey et al., 2013; Denes, 1997; Nakabayashi, 2013; Tkachenko et al., 2019).³ Most of the existing studies focus only on a particular sector or geographic area. For instance, Athey et al. (2013) use timber auction data and find that small business set-asides produce a decline in revenue and efficiency for the Forest Service. They show that bid subsidies for small bidders would be more effective than set-asides to achieve distributional goals. Denes (1997), using US Army Corps data from 1990-1991, finds no evidence that set-asides for small businesses increase the cost of government contracts. Among the papers that focus on small business set-aside programs, our paper is closely related to the counterfactual analysis in Nakabayashi (2013). The author provides evidence for a positive impact of small business set-aside programs on the Japanese public construction market. His results suggest that eliminating the program would lead to a 40% decline in small business participation, and he estimates the detrimental impacts from lack of competition would be higher than the efficiency loss caused by small businesses winning contracts. Our results are consistent with the participation effect but contrast with the efficiency argument, for they indicate that set-asides have a negative impact on contract-level outcomes. Tkachenko et al. (2019) recently found

²Other existing work on the trade-off between competition and ex post performance includes Spulber (1990), Bajari et al. (2014), and Decarolis (2014).

³“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 favored firms are not excluded from competition.

that set-aside auctions have more efficient outcomes for procurement in terms of lower award prices for homogeneous goods. Krasnokutskaya and Seim (2011) find that preference programs increase procurement costs. Nonetheless, tailoring the program to different project categories increases costs marginally because the effect is very heterogeneous. Marion (2007) and Marion (2009) also find a cost increase from such programs.

Despite this long list of contributions, the evidence in the literature on how altering competition affects the procurement process is inconclusive. The following four dimensions remain unresolved to date. First, the question remains whether the estimated effects can be generalized to a broader range of procurement categories, contracting mechanisms, and other procurement outcomes. Second, all related work analyzes only the effects of the program on ex-ante procurement outcomes. Third, there is no evidence regarding the effect of more restrictive set-asides, i.e., the SESAs. Fourth, the focus so far has been on the short-term effects of preference programs, while the long-term effects of such programs have been neglected. We contribute along all the four issues. First, we use a large data set on service and construction contracts from 11 years of government contracting across the US, based on different award mechanisms. Second, we evaluate the impact in terms of contract outcomes and ex-ante outcomes. Third, we are the first to estimate the effect of other types of set-aside on procurement outcomes. Fourth, we exploit a huge and unexpected increase in spending in a particular set-aside program to inform a demand boost and assess the long-term implication of this policy for winning firms.

Our work is thus also related to the literature that studies the determinants of public procurement efficiency. Examples include awarding design (Decarolis, 2018), wasteful end-of-year spending (Liebman and Mahoney, 2017), the role of buyers (Best et al., 2017; Decarolis et al., 2020, 2021), external audits (Gerardino et al., 2017), industry consolidation (Carril and Duggan, 2020), and the impact of centralized purchasing (Bandiera et al., 2009). To our knowledge, our paper is also the first in the economic literature to employ the random forest technique to estimate the propensity score and combine it with an inverse-weighted probability regression adjustment. Goller et al. (2019) is an exception. However, their work focuses on improving the methodology rather than on applying the technique. Some previous work has already used random forest for predictions. Examples are Fang et al. (2010), Maubert et al. (2019), and Narla and Rehkopf (2019).

III 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.

III.1 Set-Aside Programs in US Government Procurement

The set-aside program has a long history in federal acquisitions, dating back to the 1950s when the 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 is by far the largest set-aside program in terms of dedicated budget. Currently, the federal government aims to allocate 23% of its contracting budget to SBSAs.⁴ Over 2008 to 2018, the aggregate percentage of actual annual spending in SBSA increased, with levels below the goal until 2012 and above it from 2013 onward.⁵ In federal government contracting, revenues (or employment) below a threshold defines a company as small for a specific tender category.⁶ To participate in a specific SBSA solicitation, firms self-certify that they meet the size requirements for “small business.”

In addition to SBSA, the government seeks to award at least 3 to 5% of the budget to each of the SESA. Similar in spirit and implementation to SBSAs, SESAs target specific subsets of small businesses. These programs are more restrictive than SBSA 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 (VDSA) program, a category of SESA, a service-disabled veteran with a service-connected disability must own a majority stake in the small business and be in a management position.⁷

III.2 Set-Asides in Practice

During the procurement planning stage, according to Federal Acquisition Regulation (FAR), the contracting officer may decide whether to set aside a contract for small business concerns by considering two factors, namely (i) the number and type of small businesses in the marketplace deemed capable of performing the work, and (ii) the estimated contract value (FAR 19.502). The first factor is commonly known as the “rule of two,” which states that the contracting officer must have a reasonable expectation of receiving a competitive bid from at least *two* small businesses to set aside a contract. 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.⁸ Nonetheless, such

⁴Each agency has a different goal that changes annually and is negotiated with the Small Business Administration.

⁵More statistics and information are available at <https://www.sba.gov/document/support--small-business-procurement-scorecard-overview>.

⁶The 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 fifth 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 is not 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. 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.

⁷For more details on the other types of set-asides, see Appendix B. We will discuss again the VDSA in Section VIII.

⁸For 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 contracting officers. Contracting offi-

policies characterize the decision as highly discretionary because the decision not to set aside is possible as long as it can be shown that the rule of two is not met. If the rule is satisfied and the expected value of the contract is below the Simplified Acquisition Threshold (i.e., \$100,000 and \$150,000 during the period under analysis), contracting officers are obligated to set aside an acquisition for small businesses concerns (FAR 19.502-2). If the original value is above the threshold, contracting officers *can* set aside contracts with justification, as long as the rule of two is met. The same threshold guides contracting officers in choosing how to set aside contracts (FAR 19.203). Below the Simplified Acquisition Threshold, the requirement to set aside acquisitions for small businesses allows the contract to be awarded under one of the SESAs. Above the threshold, however, the contracting officer must first consider all other socioeconomic programs.⁹ Among the SESA programs, the contracting officer is free to choose which order to use, as there is no predetermined order of preference.

IV Theoretical Considerations

The impact of an ex-ante restriction on participation for procurement outcomes depends on the interplay of various factors. In our context, we consider the selection of potential bidders based on their size (or an additional dimension) as a restriction to competition—which makes the value distributions across the set-aside and not-set-aside contests likely asymmetric and the implications unclear.

To better represent possible drivers of the causal effects of set-aside programs on procurement outcomes, we first postulate the following opposing hypotheses. On the one hand, setting aside contracts to target firms diminishes the number of bids: the increased competition from small entrant firms does not compensate for the missed competition from large firms. On the other hand, setting aside contracts to target firms boosts the number of bids: the increased competition from small entrant firms more than compensates for the missed competition from large firms. The latter hypothesis seems plausible and worth investigating, as we observe a statistically significant increase in the number of bids for set-aside tenders in the summary statistics (see Section V). This effect is consistent with Cantillon (2008)’s argument that lower asymmetry in production costs between bidders leads to more intense competition. Conversely, Li and Zheng (2009) postulate an “entry effect” when the number of *potential* bidders makes entry less profitable due to the expected more intense competition, lower chance of winning, and lower incentive to pay the entry costs. In a sense, as the empirical results will confirm, set-asides may be conceptualized as a policy to encourage the entry of more bidders—keeping agnostic about the dynamics and size of potential competition for set-asides.

Yet, how does excluding large firms from bidding and facilitating the participation of small bidders affect the efficiency of the procurement process? If small bidders deliver better quality, set-aside programs lead to cost savings and performance improvements. However, if this relationship works the other way,

cers may use this database for consultation but must carefully consider a small business bidder’s self-certification before awarding a contract.

⁹If the rule of two is not met for all SESAs, the contracting officer may use SBSA.

excluding large bidders could backfire and lead to the selection of suppliers that perform poorly ex-post. Overall, the direction and the size of a set-aside effect hinges on entry costs and the number and relative strength of potential bidders. To examine the actual impact of increasing levels of competition on contract outcomes, we might turn to the theoretical auction literature as the award format also plays a role. However, given our heterogeneous sample of procurements awarded through different award mechanisms (i.e., not only first-price sealed-bid auctions), we need to go beyond private-value and common-value setups (Goeree and Offerman, 1999). With different award mechanisms, the relationship between the set-asides, actual competition, and contract outcomes is ambiguous and remains an empirical question.

Ideally, we would want to detect bidding behavior of the same firm with and without the programs. As an alternative, as we do not observe the bid distribution, we propose a reduced-form approach to see whether, in our large sample of heterogeneous procurement contracts, the competition effect (i.e., decreasing winning bid) overshadows possibly countervailing effects, such as the affiliated effect (Pinkse and Tan, 2005)—occurring both in private and common-value models— or the winner’s curse—occurring only in common-value models (Bulow and Klemperer, 2002). If the former effect dominates, policies that encourage more potential bidders could improve award amounts by reducing the equilibrium bid and the government’s ex-ante cost of procurement. Nevertheless, such measures may be undesirable if contract terms can be renegotiated ex-post by vendors. Social welfare may decline if potential competition becomes more intense and additional costs offset ex-ante discounts. Due to the nature of our procurement contracts, which allow for contract modifications, we will consider ex-post outcomes of set-aside contracts vis-à-vis tender outcomes in order to perform an overarching assessment of impacts to the government budget.

V Data

In this section, we first present the data source and sample selection, then describe the outcomes of interest and give 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.¹⁰ The FAR prescribes contracting agencies to report all awarded contracts with an estimated value above the *Micro-Purchase Threshold* (i.e., above \$3,500 in the period under study) to the FPDS. Any subsequent contract modification, regardless of amount, must also be reported, explained, and categorized in the FPDS, which allows us to observe all transactions associated with the same contract. The dataset provides a wealth of information, including over 200 variables on contract, seller, and buyer characteristics. Examples of contract-level information include the type and 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

¹⁰USAspending.gov provides daily updated data. In addition, agencies are required to report their awards within 30 days. Classified contracts are exempt from this requirement. The Department of Defense may report an award within 90 days to protect its operations. Carril et al. (2022), Decarolis et al. (2020), Decarolis et al. (2021), Giuffrida and Rovigatti (2022), Kang and Miller (2022), and Liebman and Mahoney (2017), among others, source their data from the FPDS.

the seller's ID and headquarters' location, and what small business standards it meets, if any. Examples of buyer-level information include the awarding agency and the contracting office.

Sample Selection We consider federal service and construction contracts solicited and awarded by all federal departments and executed within US borders during fiscal years (FY) 2008 through 2018. Restricting the sample to service and construction contracts allows us to compute metrics for post-award procurement outcomes. Service and construction contracts imply an execution phase that needs to be monitored as well as variation in cost overruns and delays. Therefore, we exclude all research and development, physical goods, and lease or rental contracts.¹¹ We also limit our sample to contracts open to competition, awarded through simplified acquisitions, negotiated procedures, or sealed bidding via a fixed-price contract pricing.¹² Similarly, we disregard indefinite delivery contracts.¹³ Finally, in order to exclude the smallest contracts from the sample, we restrict our attention to contracts that are expected to last longer than 2 weeks *and* have an expected cost larger than \$25,000.¹⁴

Procurement Outcomes We consider two *ex-ante* tender outcomes: the log of *Award Amount* and 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. We are also interested in the effect of set-asides on award values. However, the latter need to be considered together with ex-post contract renegotiation, which may impose additional costs. As the main post-award contract outcome, we focus on *Extra Cost* and *Delay*, which are the most common post-award metrics in both project management (Herweg and Schwarz, 2018) and empirical economics (Decarolis, 2014). *Extra Cost* is the sum of all cost renegotiations related to a project that exceed the expected budget. We define it as (in \$): $Extra\ Cost = Final\ Cost - Award\ Amount$, 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 $Delay = Final\ Duration - Expected\ Duration$, which is measured in days and corresponds to the difference between the actual and estimated completion date. We then compute 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 $Extra\ Cost\ (Ratio) = \frac{Extra\ Cost}{Award\ Amount}$ and $Delay\ (Ratio) = \frac{Delay}{Expected\ Duration}$. These metrics aim to capture the quality of contracts, based on the idea that renegotiations lead to adjustment costs that are suboptimal for all parties

¹¹For R&D contracts, we would need to obtain data on patent generation and follow an approach similar to Decarolis et al. (2020) and Giuffrida and Raiteri (2021). The outcome of interest would be the probability of R&D contracts obtaining a patent. Relevant outcomes are unit prices, as in Kastanioti et al. (2013). However, the FPDS withholds this type of unit price information. Finally, for leases or rentals, contract renegotiations are not an indicator of poor outcomes.

¹²The 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.

¹³Indefinite Deliveries represent a streamlined method used by the government to facilitate the purchase of goods and services. It is based on agreements with a supplier for an unspecified amount of goods and services for a specified time.

¹⁴In Section VII, we test for the sensitivity of our results to a change in sample. We find that the estimates are robust to the time dimension selection. 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.

involved.¹⁵ To test the baseline results on our preferred post-award metrics, 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. The details on these two variables are presented in the Appendix E.

Stylized Facts Table A1 provides detailed statistics on various outcomes of interest for the different control and treated groups. On average, treated tenders seem to have worse ex-post outcomes (i.e., more delays, higher additional costs) than the control (NSA) group. We observe this pattern for SBSA, and the difference is even higher for SESA. We find a similar spread in terms of *Delay (Ratio)*, as we observe a larger *Delay (Ratio)* for SBSA and SESA than for NSA. Note that treated contracts receive more bids on average than non-treated contracts. This observation suggests that a fraction of small firms that do not participate in open bidding participate and bid in SBSAs and SESAs, more than offsetting the non-participation of untargeted bidders. In addition, we document that 64% of NSA and SBSA contracts attract firms from the same state as the place of performance. The ratio is slightly lower for SESA, i.e., 58%, which seem to attract more firms out of state. The fraction of firm that win a federal contract for the first time is highest for SBSA, 25%, followed by NSA, 20%, and SESA, 16%. Moreover, Table A1 shows that the average award amount of SESAs is twice as large as that of SBSA and the former lasts longer. Thus, from a descriptive overview, the effect of competition restriction on award amount appears to be ambiguous. At the same time, the different categories of contracts differ substantially in observable characteristics (e.g., awarding mechanisms), a selection dimension we need to consider for identification.

VI Empirical Analysis

In this section, we first introduce the identification challenges and how we intend to circumvent them. Then, we explain our methodology, in particular how (i) we use random forest techniques to predict the propensity score of the set-aside treatments, (ii) propensity scores effectively force balance between the treatment and control groups, and (iii) the IPWRA estimates the average treatment effect on the treated.

VI.1 Identification Concerns and Strategy

Our main empirical goal is to identify and quantify how restricting competition to a set of targeted firms impacts procurement outcomes. Formally, we are interested in estimating the following model:

$$Y_{i,b,t} = \alpha + \beta^{SA} SetAside_i^{SA} + \gamma \mathbf{X}_i + \eta \mathbf{K}_{z,t} + \delta_{b,t} + \epsilon_{i,b,t} \quad (1)$$

¹⁵Spiller (2008) argues that contract changes are suboptimal in the public procurement context because they 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 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.

where Y_i denotes our outcomes of interest— e.g., number of bidders, the award value, extra costs and delays—associated with a contract i awarded by a buyer b in the FY t . *SetAside* is the treatment indicator that takes the value 1 if the contract is awarded through a set-aside procedure, i.e., SBSA or SESA, and 0 if no set-aside program is implemented (i.e., $SA = \{SBSA; SESA\}$). The parameter of interest in equation (1) is β^{SA} as we are interested in the causal effect of SBSA and SESA programs on the selected outcomes. We proceed with the rest of the analysis using two different subsamples of the dataset, each one keeping one of the two treatment groups and holding the NSA as the counterfactual. Let X_i be a vector of control variables that account for contract characteristics.¹⁶ Then, $\delta_{b,t}$ stands for the buyer-year fixed effect controlling for the within-agency budget cycle. Fixing the annual budget is important for this analysis to account for the possible increasing (decreasing) pressure on the officer to award a contract through set-aside due to the distance from the set-aside spending target. Finally, $K_{z,t}$ includes a set of dynamic market characteristics. Specifying the market-time variable feature is critical for considering the procurement official’s evaluation of the satisfaction of the rule of two over time. Because a market is defined loosely in the set-aside guidelines, we abstain from imposing a specific market definition. Therefore, to account for market characteristics, we construct discrete measures of potential competition within 9 US geographical areas (e.g., Pacific, Mountain, South Atlantic) and at the sector level. The sector is defined by the most granular product or service code in the small business classification.

Our fixed effects specification allows us to control for several omitted variables in our analysis. However, as shown in the previous section, treated and untreated contracts differ significantly across several observable characteristics. Such differences suggest that selection into treatment is a legitimate concern for identification. Therefore, estimating β^{SA} with a naive OLS—even when controlling for several layers of fixed effects—would lead to biased estimates. Moreover, we are interested in the counterfactual situation, i.e., how the treated contracts would have been fulfilled had they not been treated. The chosen identification method should take into account both concerns, i.e., selection into treatment and the valuation problem (Caliendo and Kopeinig, 2008).

In an ideal experiment, we would randomly assign contracts to a treatment or control group. In our setting, we cannot rely on a natural experiment to control for possible sources of endogeneity. Indeed, the threshold for implementing the set-aside is not enforced, and the officer has great discretion over it.¹⁷ To approximate a randomized experiment, we adopt a propensity score approach, where the score is estimated as the probability of set-aside treatment assignment conditional on observed covariates. In this manner, we may perform an unbiased counterfactual analysis by eliminating a large number of observable differences between the treatment groups and the control group. In particular, we are interested in the treatment effect on targeted tenders. In the impact evaluation literature, this particular treatment effect is referred to as the average treatment effect on the treated (ATET) and is defined in the potential outcome framework as: $\theta_{atet} = E(Y(1) - Y(0)|SA = 1)$, where $Y(1)$ is the outcome for the restricted solicitation, and $Y(0)$ is

¹⁶For a complete list of treatment predictors used in the analysis, see the Online Appendix.

¹⁷A fuzzy regression discontinuity at the simplified acquisition threshold design cannot be used because the probability that a contract is set aside does not change at the regulatory thresholds in our data.

the counterfactual outcome had the contract had not received the treatment (Angrist, 1998; Heckman and Robb Jr, 1985).¹⁸

In recent years, the propensity score literature has shown that several machine learning techniques can overcome some of the well-known limits of logistic or probabilistic regression in predicting the propensity score. Among the different machine learning techniques, we choose the random forest for six reasons that are particularly relevant for our context. First, it is easy to implement as it requires only two parameters and it is not particularly sensitive to the specification of their values (Liaw et al., 2002). Second, being a non-parametric estimator implies that the risk of model misspecification is minimized. Third, it allows us to control for a large number of relevant observable characteristics in a data-driven manner, so as to fully exploit the richness of our available dataset.¹⁹ Fourth, as explained in Section III, the decision to set aside a contract, i.e., to assign treatment, is highly discretionary given the loose guidelines of the FAR. The random forest is beneficial in this context because it determines the assignment rule in a data-driven manner, and it has been shown to perform exceptionally well when the exact treatment assignment rule is unknown (Nichols and McBride, 2017).²⁰ In this vein, the random forest can help shed light on the crucial factors that influence the decision to set aside a contract. Fifth, it addresses several other problems we expect to encounter in our setup, such as missing data, categorical variables, nonlinear relationships and interactions between variables (Zhao et al., 2016). Finally, the random forest does better than the logit—which is usually the traditional method for estimating the propensity score—in predicting the out-of-sample treatment. To test this, we train on the same sample both the logit and the random forest. In a second step, we predict the treatment on the unused part of the sample, i.e., the test sample, and we predict the treatment for each unit. Whereas the logit correctly classifies the treatment in 52% of the cases, the random forest has a much higher accuracy, namely 76%.

VI.2 First Stage

In this subsection, we first explain the approach used to predict the probability that a contract will be treated with set-asides using random forest, following in particular Lee et al. (2010).²¹ We then show that inverse probability weighting successfully balances the features between the two groups (Ho et al., 2007). We refer to Appendix D for further details on our two-step identification strategy.

Random Forest and Propensity Scores Random forests are 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 addresses possible overfitting concerns by growing trees,

¹⁸Recently, Decarolis et al. (2021) has employed IPWRA in a procurement context.

¹⁹In high-dimensional cases, logistic or probabilistic regression may perform worse than machine learning methods, and may not achieve covariate balance (Lee et al., 2010; Zhao et al., 2016).

²⁰Blundell and Dias (2009) emphasize how crucial it is to understand what determines the treatment assignment rule when considering a matching approach.

²¹Zhao et al. (2016) predicts the propensity score also using random forest, but uses a slightly different approach. Here we follow Lee et al. (2010) in more detail for practical application of the algorithm.

each time using a different bootstrap sample of the data. Then, the outcome of each observation is predicted by a majority vote, when the outcome is binary. The random forest deals with overfitting by selecting a random subset of features at each split (or node), i.e., each yes/no question. In a decision tree, the variable used at each node is the best *among all variables* in the data. In a random forest, instead, only a subset of variables is randomly selected at each node, and the selected variable is the best *among a subset of variables*.

In this work, we are interested in the probability of obtaining the treatment for each contract. Indeed, we use it to predict the propensity score $p(W_i)$ for each tender i of receiving a binary treatment, where W_i is the vector of covariates. Given the 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 SESA. For each sample, 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) to compute the propensity score, we obtain a continuous propensity score for each contract by taking the average of the predicted outcomes of each tree.²⁴

Covariates According to Caliendo and Kopeinig (2008), predictors for the treatment are key to provide an unbiased estimate of the propensity score. We select all covariates from FPDS that (i) intuitively correlate with both treatment and outcomes simultaneously, (ii) are measured prior to treatment, and (iii) are orthogonal to the anticipation of treatment.²⁵ Note that the use of too many variables can increase the variance of the estimators even though the coefficients remain unbiased and consistent (Bryson et al., 2002). However, Rubin and Thomas (1996) argue that variables should not be removed just for the sake of parsimony; instead, they should be included as long as the econometrician believes they relate to the correct covariates and outcome. This argument corroborates the random forest approach adopted. After selecting several variables that meet the unbiasedness criteria, we allow the random forest to use the most relevant variables for a propensity score prediction, rather than imposing variable selection ourselves.²⁶ We select 76 variables to predict the propensity score. We provide the full list of variables in the Online Appendix.

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

²³See Appendix D for the exact steps in the random forest.

²⁴Each 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, since each tree is grown differently as the variables at each node are chosen randomly, and this classifies each unit differently.

²⁵Economic theory, prior research, and institutional settings should be the guideposts for selecting meaningful variables that are correlated with both treatment and outcomes (Sianesi, 2004 and Smith and Todd, 2005).

²⁶See Section VI and Appendix D, for more details.

Covariates Balance To check the satisfaction of the overlap (or common-support) assumption, we verify that the propensity score distributions for both the treated and untreated overlap.²⁷ We exclude from the relevant population those units 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 9.5% of the observations for the SBSA sample and 34.4% for the SESA sample.

Second, it is crucial to check the covariate balance after predicting the propensity score. Ho et al. (2007) claim that when covariate balance is achieved, the propensity score has been adequately specified, implying that treatment effect estimates can be valid and unbiased (Zhao et al., 2016). We check whether the overlap assumption holds by looking at the covariate balance between treated and control contracts.²⁸ For this purpose, we can compute the normalized difference statistic (also called standardized differences) after applying the inverse probability weighting (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 normalized difference statistic is more robust than simply calculating the t-statistic or testing the difference in means because it does not depend directly on the sample size (Wooldridge, 2010). As long as the normalized 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 as 98.5% of the variables used for propensity score prediction are below 0.1, the remaining 1.5% are above 0.1, and none are above 0.15.²⁹ Figure A1 shows the standardized differences before and after weighting for the most critical variables selected by the random forest. For instance, we report the standardized difference for the share of small firms that also win contracts without any set-aside, in a given sector and year. For this variable, the standardized difference moves from -0.796 before adjusting for the IPW, to -0.083 after the adjustment. This results is particularly promising because, as Garrido et al. (2014) suggests, the balance of important covariates is even more relevant than the balance of other variables.

VI.3 Second Stage

Using probability weighting, an incorrect propensity score estimation model leads to biased estimates of the average treatment effect on the treated (Drake, 1993). IPWRA is a doubly robust method that can guard against this problem.³⁰ It allows us to combine two methods with attractive properties. The regression

²⁷The assumption states that each unit has a positive probability of being assigned to each treatment condition, i.e., the treatment group and the control group. The results of this check are shown in Figure A2 in Appendix D.

²⁸Note that the overlap assumption is common assumption for observational studies. It states that each individual in the sample could receive any treatment level and that we cannot *perfectly* predict the probability of receiving treatment. We state it explicitly in Equation (5) in the Appendix D. We refer the reader to the same Appendix for additional details.

²⁹At this point, we only check that the variables used for propensity score are balanced, as this is the goal of the propensity score, i.e., to balance the variables that affect the probability of receiving the treatment. Therefore, we do not include the covariates that are later used to estimate the outcome.

³⁰Hebous and Zimmermann (2019) use a propensity score matching approach and show that the results are robust to both specifications. Unlike our approach, their level of observation is the firm, and they estimate the probability that a firm wins a given contract.

adjustment accounts for different treated and untreated trends by allowing for different coefficients. The inverse probability weighting 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. Inverse probability weighting does this by using the reciprocal probability of being in the treatment group. We believe that the chosen method, i.e., using IPWRA after predicting the propensity score with random forest, is the best and most accurate way among all available in this setting to estimate our ATETs. In Section VII, we test that the results are not qualitatively sensitive to the chosen approach.

In our regressions, when indicated, we include a battery of fixed effects to capture unobserved exogenous reasons for outcome variation, as discussed in Section VI.1. For the standard errors of the estimates, we compute the variance-covariance matrix with a sandwich estimator when weights are used (Dupraz, 2013). However, since we estimate all the different steps manually, we adjust standard errors for Equation 10 with 1,000 bootstrap repetitions (Wooldridge, 2010). Following the clustering recommendations of Abadie et al. (2017), we do not cluster the standard errors as the treatment decision is at the individual (i.e., contract) level.

VII Results

In this section, we first document the effect of both set-aside programs on tender and contract outcomes. We then discuss the robustness of these findings and isolate possible effect channels. In summary, these exercises show that more bidders participate in set-asides and that this competitive pressure drives the award amount down for SBSAs and up for SESAs. Post-award performance suggest additional extra costs and longer delays. The more restrictive the set-aside the stronger these effects. Firm fixed effects suggest that selection does not seem to be a channel at play.

VII.1 Baseline Results

in Table A1, we report the results of our IPWRA method with the propensity score estimated via random forest. The top rows report the estimated coefficient β^{SA} from the equation for $SA = SBSA$, and the bottom row reports the results for $SA = SESA$. First, columns (1) and (2) report the effect on the number of bids received. we report the estimates on the two contract outcomes, i.e. *Extra Cost (Ratio)* (columns 1 and 2) and *Delay (Ratio)* (columns 3 and 4). The odd columns show the estimates on contract outcomes with a plain model specification, while the even columns include the fixed controls and the controls from Equation 1, that is, buyer \times FY fixed effects, plus contract- and market-level covariates. All regressions are run on the sample of observations of the even columns to ensure comparability between estimates. We set the richest model as our preferred specification.

We find that SBSA increases the number of bids by $\approx 14.1\%$ —i.e., $(\exp(0.132) - 1) \times 100$. SESAs have a stronger effect on competition (i.e., 19.2%). The magnitude of these effects is approximately 0.5 and 0.7 more bidders, respectively, thus signaling a slightly higher impact of SESA on entry. Both set-aside

programs cause a deterioration in contract outcomes, namely an increase in *Extra Costs (ratio)* and *Delays (ratio)*. The impact of the SESA on both contract outcomes is also stronger than that of the SBSA. In both cases, fixed effects and other outcome predictors do not alter much the results compared to the most parsimonious models. Specifically, the SBSA program increases *Extra Cost* relative to the award amount by 2.7 p.p.—+13%—and it increases *Delays* relative to the expected duration by 5.1 p.p.—+5.8%. The SESA program is more disruptive: *Extra Cost* and *Delays* increase by 7.3 p.p. and 5.6 p.p.—+35% and +6.4% relative to their benchmarks.

The increase in competition intensity seems to confirm the hypotheses put forward in Section IV. The more stringent targeting seems therefore to affect the quality of competition, in addition to its intensity: small, disadvantaged firms respond stronger than small firms to the higher competition that results from the exclusion of untargeted firms when compared to a counterfactual scenario of open participation. Over the cost and time dimension, in the post-award phase, winners tend to fare worse in set-asides. Post-award performance is worse with SESA than with SBSA. This difference in point estimates between SESA and SBSA is significant as their confidence intervals do not overlap. One possibility is that competition is more restricted ex-ante (by excluding non-socioeconomically disadvantaged small firms from bidding in addition to large firms). Also, the number of bids is higher in SESA than in SBSA (see Table A1), which supports the hypothesis that restricting competition to disadvantaged firms by excluding more efficient contractors from the competition influences participants' bidding behavior while in case of SBSA, winners seem to more than offset the price discounts granted in competitive bidding by renegotiating contract terms. We will discuss this idea further below when exploring possible channels that drive the results. These results show that restricting competition to small firms increase government spending through additional budgetary costs and increase delays in contract execution. Set-asides that target small and simultaneously disadvantaged firms result in worse contract performance, both in terms of costs and time.

VII.2 The Role of Selection

We now discuss a possible channel that could affect ex-post outcomes. The increase in the number of bidders is most likely driven by those firms that do not normally participate in open procurement tenders. Although we cannot empirically observe in the data whether such bidders refrain from participating in unrestricted tenders, it is reasonable to hypothesize that they are different from firms that normally participate in openly competitive tenders. They are also likely to be less experienced in public procurement. This lack of experience could rationalize a different bidding behavior.

Using the information available in our dataset, we construct two firm-level outcomes, i.e. *Same State* and *First Win*, to understand whether the winning firms in set-asides are different in observable characteristics than firms that win in unrestricted tenders. In particular, we find that firms winning set-asides are less likely to have their primary place of performance in the same state where the contract is performed. As shown in Column 2 of Figure A3, when controlling for buyer and year fixed effects as well as including market and contract controls, we find a reduction in the probability of coming from the same state by 1.5

p.p. for SBSA and 2.2 p.p. for SESA. This evidence can have two different, non-competing explanations. First, this suggests that set-asides are appealing for small firms and do attract a different pool of participants, assuming that the winning firm identity reflect, at least partially, the composition of the bidders. Second, this could be evidence of set-asides lessening political connection between buyer and seller. Indeed, the literature has shown that firm in geographical proximity of the buyer are more likely to win contracts because of political connections (Jääskeläinen and Tukiainen, 2019).

In addition, to proxy for the experience of firms winning with set-aside, we construct a binary variable that is equal to one if a given firm is awarded a federal contract for the first time in the 2008-2018 time span. Column 4 of Table A3 shows that we do not find evidence for less-experienced winners in set-asides. Given that we can observe only a limited type and number of firm-level outcomes, to understand whether the lack of experience explain the worse outcomes in set-asides, we also run the baseline regression with firm fixed effects. The purpose of the fixed-effects specification presented in Table A4 is to assess for both ex-post outcomes—incremental cost in columns 1-4 and delay in columns 5-8—the impact on the same firm awarding contracts with or without SBSA or SESA procedures (columns 2 and 6), including within the same year (columns 3 and 7) or when the purchasing NAICS code is also hold fixed (columns 4 and 8). Columns 1 and 4 report the baseline estimates from Table A2 to facilitate comparison. Despite the different underlying samples of contracts required to satisfy the variation in contracts within the fixed effects (e.g., two contracts per year per firm), the point estimates are fairly stable for both set-aside contracts with respect to the benchmark. The findings suggest that the same firm performs worse if it is exposed to set-aside tendering in the bidding phase. Following the positive effect on the number of bidders at the tender stage, this result seems to corroborate the hypothesis that more intense competition (and/or its composition) leads the same firm to underbid—which we cannot test in the data due to the absence of reserve price—and then renegotiate the terms of the contract, which leads to worse performance in the execution stage.

VII.3 Robustness Checks

We test the robustness of the IPWRA approach over four different dimensions: methodology, alternative outcomes, institutions, and sample.

First, in Table A6, we test whether our estimates are sensitive to changes in methodology. Columns 1-3 report estimates on *Extra Cost* and columns 4-6 on *Delay*. The top row reports again the estimated coefficient for the SBSA, while the bottom row reports the estimated coefficient for the SESA. Columns 1 and 4 report baseline estimates from Table A2, columns 2 and 4. Columns 2 and 5 report the propensity score matching approach with the previously estimated propensity score obtained by random forest. We employ a more traditional matching approach than IPWRA. We use 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).³¹ In columns 3 and 6, we run the IPWRA by calculating the propensity score with a classic logistic regression using the same set of variables employed for estimating the random forest prediction.³² To sum up, on the one hand, the methodology robustness checks confirm the sensibility of the choice of IPWRA, as since it does a better job of balancing the covariates without losing many observations. On the other hand, the results are qualitatively in line across different methods and specifications.

Second, we test the robustness of our results by twisting some parameters and characteristics of the random forest. First, we test for using two different seeds from the one that we use throughout the entire paper. Second, we impose the number of variables selected at each node of the random forest. In the rest of the analysis, we decided the number of variables based on prediction accuracy and selected the one with the highest accuracy. We report the results for 5, 11 and 22 variables selected at each node, as we use these three options for the rest of the analysis. Finally, following Imbens and Rubin (2015), we assess the sensitivity of our estimates to different control choices. This exercise in our random forest framework is particularly useful to test the extent of omitted variables concerns and corroborate our claim of causality, in the spirit of Altonji et al. (2005). We report two different tests excluding variables that we created to account for (i) performance in terms of cost overruns and delay in the previous period, mainly at the agency level (ii) the level of competition in the geographical region and sector. For instance, in the latter case, we counted how many small firms win without a set-aside. We report the results in Table A7 in Appendix A. We find that the results are very stable across different specifications and are not sensitive to the chosen specification.

Third, we explore the effect of *Set Aside* on the secondary outcomes as well. Our aim is to corroborate the effect on the intensive margin (i.e., # *Renegotiations*) and explore the extensive margin response (i.e., *Renegotiation* dummy). As we report in Table A2, we observe that set-asides cause the probability for a contract to undergo an amendment to increase by 5.0 and 4.5 p.p. points. This is the most reliable estimate as IPWRA best performs with dichotomous treatment and outcome variables. The coefficient on # *Renegotiations* is also positive and significant for SBSA, but it is negative and significant for SESA. However, it is very small in magnitude for the latter. The results hold as compared to the baseline and strongly speak in favor of a deterioration in contract performance due to the set-aside's restriction to entry—in particular for SESA—consistent with our baseline post-award estimates.

Finally, we test the robustness of our results by using a different sample selection, as reported in Table A8. Note that we proceed with the sample selection before running the random forest, which is therefore run again for each subsample. We find that the results are very similar and stable, both qualitatively and

³¹To select the best method, we looked at kernel density graphs. We obtained similar results for the situation after matching but kernel matching provided the best results. When looking at the standardized differences, kernel matching performed best among our options. Nevertheless, it did not perform as well as IPWRA, since not all covariates are balanced. However, compared to IPWRA, PSM does a better job at achieving the same density for the propensity score after matching. We still prefer IPWRA because it is a doubly-robust approach and allows us to better shape the model. Moreover, IPWRA achieves a better covariance balance in terms of standardized differences, which is the key concept underlying matching approaches.

³²We use the same variables so that we are able to compare the two outputs. Although, with a non-data driven approach, we could achieve better results by selecting the variables more carefully and by testing for linearity and for interactions between variables. This is, however, beyond the scope of this paper.

quantitatively for the outcome *Extra Cost*, especially when considering the specification with fixed effects and additional controls. The same is also true for *Delay*, though we find a higher variability in the estimates, which is probably due to the source of the data and possible outliers.

VIII Discussion of the Long-Run Implications

The declared overarching goal of the US set-aside program is to enable small (and disadvantaged) businesses to enter the procurement market so they can expand, thrive and eventually compete on a level playing field with other firms. In the previous sections, we showed how set-asides effectively boosted the participation of targeted firms, but also induced poorer contract outcomes, both in terms of cost and time. In this way, intervening in the government procurement market to exclude some competitors creates inefficiencies, at least over the short run. Can this additional cost to the taxpayer be viewed as an investment? In other words, do firms who benefit from set-aside contracts perform better in the future, in line with the government’s objectives, thus potentially leading to increased efficiency in the long run? To answer these questions, we need to consider the underlying trade-offs for government. Over the short run, government is burdened with a less efficient contracting process. Over the long run, we can imagine that set-asides have both positive and negative effects on procurement performance. On the one hand, the growth opportunities provided by set-aside contracts could empower small businesses to expand and become more competitive (“market-enhancement effect”). On the other hand, persistent reliance on the set-aside program could dissuade firms from expanding to a point where they lose eligibility for participation (“deadweight-loss effect”). This section attempts to conduct a first empirical analysis of such trade-offs. Due to the specific nature of our contract sample, as well as the limited number of firm-level variables we have taken into account, our results are not necessarily generalizable to the broader population of firms.³³

VIII.1 The Identification Problem

The main objective of this section is to provide suggestive evidence on whether the award of set-aside contracts affects the future firm’s performance in government procurement. Identifying such an effect poses two major empirical challenges. First, we need to define a satisfactory way to measure a firm’s performance in the procurement market. We rely on four different metrics built at the level of firm i in a given FY t : the share of non-set-aside awards over total awards (i.e., *Non-SA-Share*), the number of different procurement categories associated with its sales (i.e., *Multiple Products*), the number of states in which it performed its activity (i.e., *Multiple States*), and the total sales to the government (i.e., *Total Sales*). Each of these four metrics aim to proxy the firm’s performance in government procurement and its size. In particular, the Non-SA-Share is useful for testing the market-enhancement versus deadweight-loss effect. Thus, we would ideally want to estimate the following model,

³³For a fully-fledged welfare analysis, short- and long-term analyses should be compared quantitatively. This is beyond the scope of this paper, however.

$$Y_{k,t+n} = \alpha + \sum_{j=t}^N \theta_j^m \text{SetAside}_{k,j}^m * \delta_j + \delta_t + \delta_k + \epsilon_{k,t}, \quad (2)$$

where $Y_{k,t+n}$ represents the aforementioned firm-level outcome variables and SetAside_k^m is a dummy indicating at least one set-aside contract at time t (i.e., $m = \textit{Extensive}$) or the log total set-aside award value (i.e., $m = \textit{Intensive}$) by the firm k . The regressor is interacted with FY fixed effects δ_j to test the dynamics of the effects over the N years after t . Firm fixed effects δ_k account for time-invariant heterogeneity across firms and all unobserved sub-industry and geographic effects.

A number of potential factors may prevent a causal interpretation of θ_j^m , i.e., the coefficient for the m -th set-aside exposure on firm-level procurement outcomes. A primary concern is that winning *specific* contracts can be anticipated by a firm and this leads to a correlation between SetAside_k^m and the error term $\epsilon_{k,t}$ in the Equation 2. For instance, we cannot detect the political connectedness of firms. Connected firms might be more likely to win contracts. If a future winner can anticipate a contract—resulting from participation in a particular bidding process—then public demand remains endogenous in our models. Second, we do not observe private-sector exposure and, accordingly, private-sector sales of a firm k . In this situation, the lack of a control for private-sector demand introduces omitted variable bias, as private-sector sales both predict the firm’s survival rate and are correlated with the public counterpart. This bias is due to the mutual crowding out effect of one market on the other, mainly due to the firm’s capacity constraints in the short run (i.e., one year). Essentially, decisions on how much to sell to private and public customers are not orthogonal to each other. Third, a significant source of bias may arise because we omit firm-level productivity from the model of firm outcomes. Contracts awarded by the government might be more likely to be awarded to the most productive firms from the pool of small/disadvantaged firms. If this is the case, firms that receive positive productivity shocks have a greater chance of receiving a set-aside contract. Moreover, the most productive firm also tends to have better procurement outcomes regardless of winning the contract, as it is likely to be more successful in the private-sector market. Taken together, these mechanisms lead to biased estimates of the coefficients of interest θ_j^m .

Ideally, we would like to match similar government contracts across pools of firms grouped into tenders with randomly assigned set-aside status and let the target firms compete. Then, given firms’ observable data, we could assess how being affected—both at the extensive and intensive margin—by set-aside awards affects firms’ future contracting outcomes. Such a scenario does not occur in our setting, and we must rely on a non-experimental context. As a second-best, we focus on a subset of set-asides for which we can approximate an unexpected demand surge. Specifically, we propose an empirical exercise that builds on an unanticipated increase in spending by the Department of Veteran Affairs (DVA) on a given category of SESA (i.e., the VDSA) for construction projects. We exploit the implications of this spending spike for eligible incumbents and observe their behavior in the government procurement market following their exposure to the unexpected increased demand relative to non-exposed incumbents that were equally likely to be exposed. Note that we exploit two different policies that resulted from the collapse of the Lehman Brothers, in September 2008. As the policies were passed at the beginning of 2009, we argue that in the

FY2008, which terminates at the end of September, they are still likely unanticipated.

VIII.2 A Set-Aside Spending Surge

The 2009-2010 Recovery Act and Veteran Policies Figure C1 shows that DVA's total annual spending on construction through VDSA increased threefold from FY2008 to FY2010. This massive increase in annual spending observed in our data likely resulted from the combination of two major policies that jointly occurred during the same period. First, the Recovery Act, passed by the 111th US Congress, was signed into law in February 2009 by President Obama. This stimulus package contained provisions directed at the construction industry beginning in FY2009 to increase investment in the nation's physical infrastructure. The provisions of the new law provided for additional spending on transportation infrastructure (\$49 billion), water and environmental infrastructure (\$21 billion), building infrastructure (\$29 billion), and energy/technology infrastructure (\$30 billion), which were mostly channeled through government procurement contracts.³⁴ 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 2009 included \$140 billion in funding for the Department of Veterans Affairs—a 40% increase in funding.³⁵ Figure C2 shows how the combination of policies increased DVA's total construction spending by approximately 50% from 2008 to 2010.³⁶ Over the same period (FY2009-2010), total government construction spending did not increase significantly in the raw FPDS data; other departments did not experience a spending surge comparable to DVA's.³⁷

Disabled veterans were a core target of the new veteran policy. As a result, the implicit effect of Obama's veteran policy was to increase VDSA spending channeled through the DVA. As Figure C1 shows, the absolute increase in DVA construction spending in FY2009-2010 (compared to FY2008) for VDSA set-asides was massive, rising from \$0.4B in FY2008 to \$1.6B in FY2010. Moreover, the same figure emphasize that a similar increase in construction spending for other types of set-asides is not observed in the data.

The Withdrawal of US Troops from Iraq This discussion raises the following question: Which firms benefited the most from this spending surge? In theory, any construction firm classified as veteran-disabled-owned could participate in these restricted tenders and compete for a VDSA contract. To select the best candidates as recipients of this unexpected surge in demand, we remind that at the beginning of FY2009—i.e., after the election victory— President Obama also announced the withdrawal of (most) US troops from

³⁴<https://www.govinfo.gov/content/pkg/PLAW-111publ5/pdf/PLAW-111publ5.pdf>.

³⁵In his first term, the president made veterans' care one of his top priorities. See more at: <https://www.americanprogress.org/issues/security/news/2013/02/13/53328/president-obamas-commitment-to-veterans-must-remain-a-second-term-priority/>.

³⁶As a data-quality check, we plot the same figure by using a different variable accounting for the same set-aside. We find a very similar jump, which allows us to rule out that the increase is only due to different coding from 2008 to 2010.

³⁷See *uspending.gov*.

Iraq. Accordingly, many veterans returned home from 2009 up to the end of 2011.³⁸ An increasing influx of disabled veterans launching startups in the construction industry is conceivable and was documented by media;³⁹ this may have been associated with a supply shock, in addition to the stronger incentive for existing firms to switch status and set themselves as a veteran-disabled-owned business to take advantage of the increased government demand.⁴⁰ In fact, the number of veteran-disabled-owned firms being awarded contracts in the whole FPDS population increased considerably after 2010.⁴¹

Therefore, the 2009-2010 veteran-disabled-owned firm pool is most likely composed of market entrants (i.e., newly formed firms) and incumbents. On the one hand, the new entrants result from the increasing veteran supply and incentives to designate an existing firm as veteran-disabled-owned.⁴² On the other hand, FPDS allows us to observe veteran-disabled-owned firms that were winning contracts before the demand surge. These firms were already in the market (i.e., being awarded at least one VDSA or non-VDSA construction contract in FY2008), and they had an incentive to continue working for the government because of the increased opportunities. Incumbents could not perfectly anticipate the recession, as the first GDP decline was registered at the end of summer 2008, that is, shortly before the start of FY2009, which saw the launch of Obama's veteran policy, and the resulting large increase in VDSA construction spending. However, there could be a huge selection with many new entrants. This is corroborated by Figure C3, which shows that the cumulative yearly number of bidders in VDSA auctions increases from 2009 up until 2014. By focusing only on incumbents, we assume that these firms could unexpectedly take advantage of more opportunities and may have had easier access to the large increase in construction demand from the DVA, despite increasing competition. Finally, even if there is a change in the intensity and in the composition of competition in 2009 to 2010, we argue that both treated and control incumbents in 2008 should be exposed in the same way to it. If anything, firms winning in 2009-2010 could be more competitive as they are winning contracts and should theoretically perform better in the firm-level procurement outcomes that we consider.

VIII.3 Long-Run Implications of Set-Asides for Veteran-Disabled Firms

We might expect this additional unanticipated spending in VDSA construction contracts to positively (or, conversely, negatively) affect firm-level performance, as recipients could end up from the additional demand to award more (fewer) contracts in non-set aside tender, in more (fewer) states, sell more (less) diverse goods/services, and win more (fewer) contracts if the market-enhancement (or deadweight-loss) effect dominates. Specifically, we compare the firm-level procurement dynamics of veteran-disabled-owned incumbents compared to non-winning incumbents. The observed demand spike, according to Figure C1, ends in FY2010. Starting from FY2011, we want to see how incumbents have behaved differently when

³⁸<https://www.theguardian.com/world/2009/feb/27/obama-iraq-war-end-august-2010>

³⁹Most veterans who start businesses do so in the construction industry because construction jobs best match veterans' skills. See <https://usveteransmagazine.com/2017/09/construction-industry-great-fit-veterans/>.

⁴⁰For example, an existing firm may have made a disabled veteran a major shareholder in the company.

⁴¹Source: www.usaspending.gov.

⁴²We do not have records before FY 2008.

having been exposed to a higher intensity of VDSA contracts on both the extensive margin (at least one award) and the intensive margin (total value of contracts). Thus, we can rewrite the Equation (2) as:

$$Y_{k,t} = \alpha + \sum_{j=2011}^{2018} \theta_j \text{SetAside}_{20092010}^m * \delta_j + \delta_t + \delta_k + \epsilon_{k,t} \quad (3)$$

Figures C4-C7 show the development of θ from 2011 through 2018 for *Non-SA-Share* (Figure C4), *Multiple Categories* (Figure C5), *Multiple States* (Figure C6), and *Total Sales* (Figure C7). Panels a) and b) use *VDSA Dummy* and $\log(1+VDSA \text{ value})$ as main regressors, respectively. Since we control for time fixed effects, θ_j is identified by comparing changes in the performance of firms over time. Overall and at both margins, recipients exhibit negative annual procurement performance over time in all dimensions: They become more specialized, more localized, and are awarded fewer contracts than non-exposed incumbents. Instead, the effect on *Not-Set Aside Share* is not significant at the intensive margin. The corresponding extensive margin effect, on the other hand, is positive but not of clear interpretation. It dissipates after three years, but it is significant again in the sixth and seventh year after the demand surge. The results, when significant, show that there is a greater reliance on set-aside than non-set-aside tenders when a firm is awarded targeted contracts. Accordingly, it provide weak evidence that the market-enhancement effect could somewhat overweight the deadweight-loss effect. In the Appendix C Figures C8-C11, we propose the same graphical evidence by restricting the focus to contracts awarded through competitive bidding with at least two bids, as these contracts are likely to be won more unexpectedly. The results are robust to the baseline in both quantitative and qualitative terms. However, the effects on the Non-Set-Aside Share on the extensive margin disappear almost completely. This finding further suggests that the claim of less reliance on set-aside should be interpreted extremely carefully.

IX Conclusions

Almost a quarter of the US federal contracting budget is set aside for small and disadvantaged firms only. The goal of the program is two-fold: to ensure adequate participation of targeted businesses in the public procurement market, and to achieve distributional goals. However, at least theoretically, the restrictions to competition imposed by set-asides could also induce costs, not only in terms of worse procurement outcomes for taxpayers, but also by weakening competition, as non-eligible firms could be forced to exit the market, while eligible firms could be dissuaded from becoming larger and more efficient. Accordingly, the effect of set-asides on contract and firm-level outcomes is ambiguous. This paper sought to shed light on this issue by examining a large and heterogeneous government procurement database that spans a decade.

We use a two-stage approach to explore tender and contract outcomes. To circumvent the lack of exogenous variation in our data, as a first step we draw on random forest techniques that exploit the wealth of information contained in our dataset to calculate a propensity score for the tender being set aside. We then estimate the effect of restricted tenders on competition and post-award performance using an IPWRA.

Set-asides are found to prompt more firms to bid—that is, the increase in bidders from targeting more than offsets the loss in bidders resulting from a smaller pool of potential competitors. During the execution phase, set-aside contracts incur higher cost overruns and delays. The more restrictive the set-aside (i.e., when contracts are restricted to firms that are small *and* disadvantaged), the stronger these effects. Focusing on one specific set-aside category and those firms unexpectedly facing a surge in set-aside demand, we do not find evidence of improved business performance in the long run—at least in terms of future procurement outcomes. Our findings suggest overall that short-term contractual inefficiency does not translate into long-term improvements to firm performance.

While our results for a large set of government contracts do not provide evidence against the null hypothesis that set-asides introduce inefficiencies to the procurement process, these findings do not argue in favor of stopping set-aside programs. Consistent with results in the literature on preferential programs, bid subsidies could improve efficiency while maintaining set-aside benefits and reducing attendant distortions. Examining preferential programs in procurement, existing studies (e.g., Athey et al., 2013) show that smaller firms can likely be induced to participate—thus increasing competition—without causing a drastic change in the composition of participating firms (i.e., large firms can still participate but do not obtain a bid subsidy). This averts the effects resulting from the change in the composition of bidders, which is an important driver behind our findings.⁴³

Despite providing empirical evidence along different dimensions and contributing to the literature on policies that restrict competition and their implications for the public budgets, our study has several limitations. First, we sacrifice some external validity for internal validity by excluding contracts from the analysis for which we could perfectly (or almost perfectly) predict the treatment status (Imbens and Rubin, 2015). Therefore, the generalizability of these results to other contracts and settings is limited. Another limitation is the grouping all types of SESAs into one specification. Since the different types may produce heterogeneous effects on outcomes, standalone effects and channels should be analyzed. Given the high number of variables and the low number of observations for each subsample, this approach leads to insurmountable overfitting and collinearity in our sample. Third, even though we investigate the effect of set-asides on competition, this work does not provide evidence on the actual strategies and composition of bidders due to the lack of bid-level data. Accordingly, bidding behavior should instead be estimated in a structural model to better investigate the channels of our results. Finally, to further investigate long-run implications for firms exposed to set-asides, it would be crucial for future research efforts to investigate further objective business dynamics—e.g., scale (Gugler et al., 2020), access to credit (di Giovanni et al., 2022) as well as productivity and survival (Cappelletti and Giuffrida, 2021).

⁴³The counterfactual analysis in Nakabayashi (2013) shows that eliminating the program would lead to a 40% decline in small business participation in the Japanese construction procurement market.

Bibliography

- Abadie, A., Athey, S., Imbens, G. W., and Wooldridge, J. (2017). When should you adjust standard errors for clustering? Working Paper 24003, National Bureau of Economic Research.
- Altonji, J. G., Elder, T. E., and Taber, C. R. (2005). Selection on observed and unobserved variables: Assessing the effectiveness of catholic schools. *Journal of Political Economy*, 113(1):151–184.
- Angrist, D. (1998). Estimating the Labor Market Impact of Voluntary Military Service Using Social Security Data on Military Applicants,” . *Econometrica* 66: 2.
- Athey, S., Coey, D., and Levin, J. (2013). Set-asides and subsidies in auctions. *American Economic Journal: Microeconomics*, 5(1):1–27.
- 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.
- Ayres, I. and Cramton, P. (1995). Deficit reduction through diversity: How affirmative action at the FCC increased auction competition. *Stan. L. Rev.*, 48:761.
- 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.
- Best, M. C., Hjort, J., and Szakonyi, D. (2017). Individuals and organizations as sources of state effectiveness. (23350).
- Blundell, R. and Dias, M. C. (2009). Alternative approaches to evaluation in empirical microeconomics. *Journal of Human Resources*, 44(3):565–640.
- 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.

- Bryson, A., Dorsett, R., and Purdon, S. (2002). The use of propensity score matching in the evaluation of active labour market policies. Lse research online documents on economics, London School of Economics and Political Science, LSE Library.
- Bulow, J. and Klemperer, P. (2002). Prices and the Winner's Curse. *The RAND Journal of Economics*, 33(1):1–21.
- Caliendo, M. and Kopeinig, S. (2008). Some practical guidance for the implementation of propensity score matching. *Journal of Economic Surveys*, 22(1):31–72.
- 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 7434, C.E.P.R. 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. and Giuffrida, L. M. (2021). Procuring survival. *ZEW-Centre for European Economic Research Discussion Paper*, (21-093).
- Carril, R. (2020). Rules versus discretion in public procurement. *mimeo*.
- 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):S0047272720300050.
- Carril, R., Gonzalez-Lira, A., and Walker, M. S. (2022). Competition under incomplete contracts and the design of procurement policies. Economics Working Papers 1824, Department of Economics and Business, Universitat Pompeu Fabra.
- Coviello, D., Guglielmo, A., and Spagnolo, G. (2017). The effect of discretion on procurement performance. *Management Science*, 64(2):715–738.
- 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.
- 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.
- Drake, C. (1993). Effects of misspecification of the propensity score on estimators of treatment effect. *Biometrics*, pages 1231–1236.
- Dupraz, Y. (2013). Using weights in Stata. Technical report, Paris School of Economics.
- EC (2017). European semester: Thematic factsheet–public procurement.
- Fang, K.-n., Zhu, J.-p., and Xie, B.-c. (2010). A research into the forecasting of fund return rate direction and trading strategies based on the random forest method. *Economic Survey*, 2:61–65.
- Garrido, M. M., Kelley, A. S., Paris, J., Roza, K., Meier, D. E., Morrison, R. S., and Aldridge, M. D. (2014). Methods for constructing and assessing propensity scores. *Health Services Research*, 49(5):1701–1720.
- Gerardino, M. P., Litschig, S., and Pomeranz, D. (2017). Can audits backfire? evidence from public procurement in chile. Working Paper 23978, National Bureau of Economic Research.
- Giuffrida, L. M. and Raiteri, E. (2021). Buyers’ workload and r&d procurement outcomes: Evidence from the us air force research lab. *ZEW-Centre for European Economic Research Discussion Paper*, (21-059).
- Giuffrida, L. M. and Rovigatti, G. (2022). Can the Private Sector Ensure the Public Interest? Evidence from Federal Procurement. *Journal of Economics & Management Strategy*, (Forthcoming).
- Goeree, J. K. and Offerman, T. (1999). Competitive Bidding in Auctions with Private and Common Values. Tinbergen Institute Discussion Paper 2000-044/1, Tinbergen Institute.
- Goller, D., Lechner, M., Moczall, A., and Wolff, J. (2019). Does the Estimation of the Propensity Score by Machine Learning Improve Matching Estimation? The Case of Germany’s Programmes for Long Term Unemployed. IZA Discussion Papers 12526, Institute of Labor Economics (IZA).
- 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.

- Hao, M., Li, Y., Wang, Y., and Zhang, S. (2010). Prediction of PKCO Inhibitory Activity Using the Random Forest Algorithm. *International Journal of Molecular Sciences*, 11:3413–33.
- Hart, O. and Moore, J. (1988). Incomplete contracts and renegotiation. *Econometrica*, pages 755–785.
- Hebous, S. and Zimmermann, T. (2019). Can government demand stimulate private investment? Evidence from US federal procurement. Technical Report 7534, CESifo Working Paper Series.
- 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.
- 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.
- Jääskeläinen, J. and Tukiainen, J. (2019). Anatomy of public procurement. *VATT Institute for Economic Research Working Papers*, 118.
- 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.
- Kastanioti, C., Kontodimopoulos, N., Stasinopoulos, D., Kapetaneas, N., and Polyzos, N. (2013). Public procurement of health technologies in Greece in an era of economic crisis. *Health policy*, 109(1):7–13.
- Krasnokutskaya, E. and Seim, K. (2011). Bid preference programs and participation in highway procurement auctions. *American Economic Review*, 101(6):2653–86.
- Lee, B. K., Lessler, J., and Stuart, E. A. (2010). Improving propensity score weighting using machine learning. *Statistics in Medicine*, 29(3):337–346.
- 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.
- 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.
- Maubert, A., Birtwisle, L., Bernard, J., Benizri, E., and Bereder, J. (2019). Can machine learning predict resectability of a peritoneal carcinomatosis? *Surgical Oncology*, 29:120–125.
- 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.
- Narla, A. and Rehkopf, D. H. (2019). Novel ranking of protective and risk factors for adolescent adiposity in US females. *Obesity Science & Practice*, 5(2):177–186.
- Nichols, A. and McBride, L. (2017). Propensity Scores and Causal Inference Using Machine Learning Methods. Technical report, Stata Users Group.
- OECD (2019). *OECD SME and Entrepreneurship Outlook 2019*. OECD Publishing, Paris.
- Pinkse, J. and Tan, G. (2005). The Affiliation Effect in First-Price Auctions. *Econometrica*, 73(1):263–277.
- Rubin, D. B. and Thomas, N. (1996). Matching using estimated propensity scores: relating theory to practice. *Biometrics*, pages 249–264.
- Sianesi, B. (2004). An evaluation of the Swedish system of active labor market programs in the 1990s. *Review of Economics and Statistics*, 86(1):133–155.
- Smith, J. A. and Todd, P. E. (2005). Does matching overcome LaLonde's critique of nonexperimental estimators? *Journal of Econometrics*, 125(1-2):305–353.
- Spiller, P. T. (2008). An Institutional Theory of Public Contracts: Regulatory Implications. NBER Working Papers 14152, National Bureau of Economic Research, Inc.
- 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".

Tukiainen, J., Blesse, S., Bohne, A., Giuffrida, L. M., Jääskeläinen, J., Luukinen, A., and Sieppi, A. (2021). What Are the Priorities of Bureaucrats? Evidence from Conjoint Experiments with Procurement Officials. *EconPol Working Paper*, 63.

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

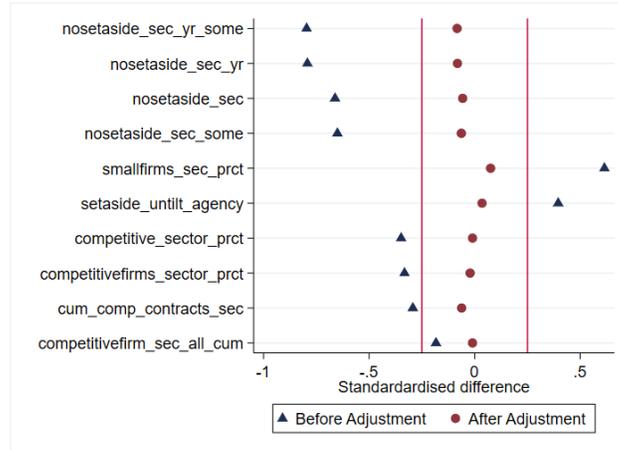
Wooldridge, J. M. (2010). *Econometric analysis of cross section and panel data*. MIT press.

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 Tables and Figures

Figure A1: Standardized differences for selected covariates, SBSA vs. NSA



Notes: This figure reports the standardized differences before and after the inverse probability weighted regression adjustment. The treatment for this sample is $SA = SBSA$; the control is non set-aside contract. We predict the propensity score using 76 variables. For traceability reasons, we report only the ten most important variables according to the random forest estimate of variable importance. The red vertical lines represent the 0.25 threshold. Below this threshold, covariate balance is achieved (Imbens and Rubin, 2015). For the variables used, standardized differences for all the covariates after adjustment are below the absolute value of 0.25.

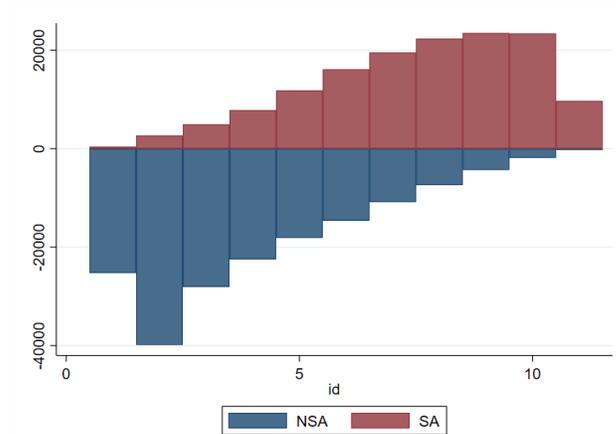


Figure A2: Propensity Score Distribution for the Treated and the Untreated

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.

Table A1: Descriptives for Treatment and Control Groups

	NSA (N= 66,290)		SBSA (N = 54,158)		SESA (N = 20,751)	
	Mean	Median	Mean	Median	Mean	Median
# of Bids	3.65	2.0	4.30	3.0	4.88	4.0
Same state	0.64	.	0.64	.	0.58	.
First Win	0.20	.	0.25	.	0.16	.
Share of Works	0.18	.	0.30	.	0.63	.
NP [N, Y]	0.45	.	0.33	.	0.46	.
SB [N, Y]	0.05	.	0.07	.	0.28	.
SA [N, Y]	0.50	.	0.61	.	0.26	.
Renegotiation [Y,N]	0.31	.	0.43	.	0.56	.
# Renegotiations	1.43	0.0	1.41	0.0	2.42	1.0
Total Cost (\$)	1529.93	80.5	570.42	77.0	1227.16	246.4
Award Amount (\$)	1189.35	73.2	462.10	67.3	999.93	214.8
Cost Overrun	338.59	0.0	107.56	0.0	228.69	0.0
Cost Overrun %	0.21	0.0	0.24	0.0	0.24	0.0
Expected Duration	287.48	222.0	229.83	152.0	280.53	212.0
Actual Duration	463.88	343.0	415.48	233.0	525.59	364.0
Time Overrun	176.37	0.0	185.64	0.0	245.05	45.0
Time Overrun %	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. socioeconomic programs (SESA). Dollar values are in thousands. *# of Bids* counts the number of offers received for a given contract. *Works* is a dummy equal to 1 if the contract code is indicating either a construction or a maintenance work. *NP* is a dummy equal to 1 if the solicitation procedure used is Negotiated Procedure. *SB* is an indicator equal to 1 if the solicitation is a Sealed Bid. *SA* is an indicator equal to 1 if the procedure is Simplified Acquisition. *Total Cost* stands for the sum of all in-scope and out-of-scope modifications related to each contract. *Expected cost* is constructed by adding all federal obligation when an in-scope modification is observed. *Extra Cost* is the difference between the expected and total cost. *Extra Cost (Ratio)* indicates the share of the cost overrun relative to the expected cost. *Renegotiation* only hinges on in-scope modifications of contract terms. *# of Renegotiations* variables report the number of in-scope contract modifications. *Expected Duration* and *Actual Duration* are the expected and actual duration of the contracts expressed in days. *Delay* is the difference between the two measures. *Delay (Ratio)* measures the ratio of the delay relative to the expected duration.

Table A2: Baseline Outcomes

	Log(# Offers)		Extra Cost		Delay	
	(1)	(2)	(3)	(4)	(5)	(6)
SBSA	0.139	0.132	0.024	0.027	0.054	0.051
	(0.0000)	(0.0005)	(0.0000)	(0.0003)	(0.0000)	(0.0011)
	0.933	0.933	0.225	0.225	0.896	0.896
	120440	120440	119672	119672	120424	120424
SESA	0.152	0.176	0.060	0.073	0.093	0.056
	(0.0000)	(0.0012)	(0.0000)	(0.0009)	(0.0000)	(0.0034)
	0.913	0.913	0.218	0.218	0.932	0.932
	87036	87036	86593	86593	87027	87027
Buyer* Year FE		✓		✓		✓
Market Controls		✓		✓		✓
Contract Controls		✓		✓		✓

Notes: Results for the average treatment effect on the treated of the inverse probability weighting regression adjustment on the two main outcomes, *Extra Cost* and *Delay*. The treated are contracts awarded with some kind of restricted solicitations, in the top panel contracts are set aside for small businesses (SBSA), in the bottom they are set aside for other type of small businesses (SESA). Several controls accounting for buyer, contract and market characteristics 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 (using only pre-treatment buyer, contract and market characteristics). Standard errors are adjusted with 1,000 bootstrap repetitions.

Table A3: Firm-level Outcomes

	Same State [N,Y]		First Win [N,Y]	
	(1)	(2)	(3)	(4)
SBSA	-0.001	-0.015	-0.005	-0.000
	(0.0000)	(0.0003)	(0.0000)	(0.0001)
	0.641	0.641	0.212	0.212
	120009	120009	120448	120448
SESA	-0.025	-0.022	-0.035	0.002
	(0.0000)	(0.0006)	(0.0000)	(0.0003)
	0.627	0.627	0.181	0.181
	86620	86620	87041	87041
Buyer* Year FE		✓		✓
Market Controls		✓		✓
Contract Controls		✓		✓

Notes: Results for the average treatment effect on the treated of the inverse probability weighting regression adjustment on two additional outcomes, namely *Same State [N,Y]* and *First Win [N,Y]*. The treated are contracts awarded with some kind of restricted solicitations, in the top panel contracts are set aside for small businesses (SBSA), in the bottom they are set aside for other type of small businesses (SESA). Several controls accounting for buyer, contract and market characteristics 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 (using only pre-treatment buyer, contract and market characteristics). Standard errors are adjusted with 1,000 bootstrap repetitions.

Table A4: Ex-post Outcomes with Seller Fixed Effects

	Extra Cost				Delay			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
SBSA	0.027 (0.0003) <i>120448</i>	0.027 (0.0003) <i>40114</i>	0.025 (0.0000) <i>13452</i>	0.025 (0.0000) <i>22703</i>	0.051 (0.0011) <i>120448</i>	0.049 (0.0011) <i>40114</i>	0.058 (0.0000) <i>13452</i>	0.058 (0.0000) <i>22703</i>
SESA	0.073 (0.0009) <i>87041</i>	0.077 (0.0009) <i>14464</i>	0.065 (0.0001) <i>4493</i>	0.065 (0.0001) <i>6220</i>	0.057 (0.0034) <i>87041</i>	0.061 (0.0034) <i>14464</i>	0.108 (0.0001) <i>4493</i>	0.108 (0.0001) <i>6220</i>
Buyer* Year FE	✓	✓	✓	✓	✓	✓	✓	✓
Market Controls	✓	✓	✓	✓	✓	✓	✓	✓
Contract Controls	✓	✓	✓	✓	✓	✓	✓	✓
Seller FE		✓	✓	✓		✓	✓	✓
Seller*Year FE			✓				✓	
Seller*NAICS FE				✓				✓

Notes: Results for the average treatment effect on the treated of the inverse probability weighting regression adjustment on the two main outcomes, *Extra Cost*, *Delay*. We show the results when adding seller fixed effects, firm year fixed effects as well as seller-NAICS code fixed effects. Column (1) and (5) report the baseline regression for the two outcomes. The treated are contracts awarded with some kind of restricted solicitations; in the top panel, contracts are set aside for small businesses (SBSA), in the bottom they are set aside for other type of small businesses (OSA). Several controls accounting for buyer, contract and market characteristics 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 (using only pre-treatment buyer, contract and market characteristics). Standard errors are adjusted with 1,000 bootstrap repetitions.

Table A5: Additional Outcomes

	Renegotiation [N,Y]		# Renegotiation	
	(1)	(2)	(3)	(4)
SBSA	0.046	0.050	0.014	0.030
	(0.0000)	(0.0002)	(0.0000)	(0.0005)
	0.363	0.363	0.308	0.308
	120448	120448	120448	120448
SESA	0.060	0.045	-0.067	-0.056
	(0.0000)	(0.0006)	(0.0000)	(0.0022)
	0.366	0.366	0.347	0.347
	87041	87041	87041	87041
Buyer* Year FE		✓		✓
Market Controls		✓		✓
Contract Controls		✓		✓

Notes: Results for the average treatment effect on the treated of the inverse probability weighting regression adjustment on two additional outcomes, namely *Renegotiation* and *# Renegotiation*. The treated are contracts awarded with some kind of restricted solicitations, in the top panel contracts are set aside for small businesses (SBSA), in the bottom they are set aside for other type of small businesses (SESA). Several controls accounting for buyer, contract and market characteristics 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 (using only pre-treatment buyer, contract and market characteristics). Standard errors are adjusted with 1,000 bootstrap repetitions.

Table A6: Alternative Methods

	Extra Cost			Delay		
	(1)	(2)	(3)	(4)	(5)	(6)
	IPWRA (RF)	PSM (RF)	IPWRA (logit)	IPWRA (RF)	PSM (RF)	IPWRA (logit)
SBSA	0.024 (0.0000)	0.025 (0.0090)	0.030 (0.0001)	0.055 (0.0003)	0.054 (0.0247)	0.044 (0.0004)
N	120448	120448	120448	120448	120448	120448
SESA	0.058 (0.0001)	0.068 (0.0195)	0.030 (0.0002)	0.090 (0.0004)	0.034 (0.0532)	0.235 (0.0007)
N	87041	87041	87041	87041	87041	87041

Notes: Results for methodological robustness checks on the inverse probability weighting regression adjustment (IPWRA) on the two main outcomes, *Extra Cost* and *Delay*. The treated are contracts awarded with some kind of restricted solicitations, in the top panel contracts were set aside for small businesses (SBSA), in the bottom they were set aside for other type of small businesses (OSA). Column (1) reports baseline results, Column (2) the traditional propensity score matching (PSM) approach with the previously estimated propensity score of the probability of being treated. Note that Stata 16 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) reports results from IPWRA estimated with the propensity score predicted with a logistic regression. Several controls accounting for buyer, contract and market characteristics are included in the regression. Standard errors are adjusted with 1,000 bootstrap repetitions in Stata 16.

Table A7: Functional Robustness Checks

	Log(# Offers)		Extra Cost		Delay	
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: baseline sample						
SBSA	0.139 (0.0000)	0.132 (0.0005)	0.024 (0.0000)	0.027 (0.0003)	0.054 (0.0000)	0.051 (0.0011)
	0.933 <i>120440</i>	0.933 <i>120440</i>	0.225 <i>119672</i>	0.225 <i>119672</i>	0.896 <i>120424</i>	0.896 <i>120424</i>
SESA	0.152 (0.0000)	0.176 (0.0012)	0.060 (0.0000)	0.073 (0.0009)	0.093 (0.0000)	0.056 (0.0034)
	0.913 <i>87036</i>	0.913 <i>87036</i>	0.218 <i>86593</i>	0.218 <i>86593</i>	0.932 <i>87027</i>	0.932 <i>87027</i>
Panel B: seed 2022						
SBSA	0.262 (0.0000)	0.254 (0.0006)	0.029 (0.0000)	0.032 (0.0004)	0.047 (0.0000)	0.028 (0.0012)

	0.933 <i>120448</i>	0.933 <i>120448</i>	0.225 <i>120448</i>	0.225 <i>120448</i>	0.896 <i>120448</i>	0.896 <i>120448</i>
SESA	0.289 (0.0000)	0.298 (0.0015)	0.062 (0.0000)	0.075 (0.0010)	0.103 (0.0000)	0.055 (0.0037)
	0.913 <i>87041</i>	0.913 <i>87041</i>	0.218 <i>87041</i>	0.218 <i>87041</i>	0.932 <i>87041</i>	0.932 <i>87041</i>
Panel C: seed 16432						
SBSA	0.147 (0.0000)	0.138 (0.0005)	0.024 (0.0000)	0.027 (0.0003)	0.049 (0.0000)	0.049 (0.0011)
	0.933 <i>120448</i>	0.933 <i>120448</i>	0.225 <i>120448</i>	0.225 <i>120448</i>	0.896 <i>120448</i>	0.896 <i>120448</i>
SESA	0.146 (0.0000)	0.162 (0.0010)	0.060 (0.0000)	0.073 (0.0009)	0.063 (0.0000)	0.006 (0.0033)
	0.913 <i>87041</i>	0.913 <i>87041</i>	0.218 <i>87041</i>	0.218 <i>87041</i>	0.932 <i>87041</i>	0.932 <i>87041</i>
Panel D: 5 variables						
SBSA	0.170 (0.0000)	0.162 (0.0005)	0.025 (0.0000)	0.031 (0.0003)	0.049 (0.0000)	0.051 (0.0010)
	0.933 <i>120448</i>	0.933 <i>120448</i>	0.225 <i>120448</i>	0.225 <i>120448</i>	0.896 <i>120448</i>	0.896 <i>120448</i>
SESA	0.190 (0.0000)	0.203 (0.0012)	0.063 (0.0000)	0.077 (0.0010)	0.122 (0.0000)	0.086 (0.0037)
	0.913 <i>87041</i>	0.913 <i>87041</i>	0.218 <i>87041</i>	0.218 <i>87041</i>	0.932 <i>87041</i>	0.932 <i>87041</i>
Panel E: 11 variables						
SBSA	0.146 (0.0000)	0.139 (0.0005)	0.020 (0.0000)	0.024 (0.0003)	0.051 (0.0000)	0.048 (0.0011)
	0.933 <i>120448</i>	0.933 <i>120448</i>	0.225 <i>120448</i>	0.225 <i>120448</i>	0.896 <i>120448</i>	0.896 <i>120448</i>
SESA	0.169 (0.0000)	0.182 (0.0011)	0.060 (0.0000)	0.070 (0.0010)	0.087 (0.0000)	0.048 (0.0036)
	0.913 <i>87041</i>	0.913 <i>87041</i>	0.218 <i>87041</i>	0.218 <i>87041</i>	0.932 <i>87041</i>	0.932 <i>87041</i>
Panel F: 22 variables						
SBSA	0.121 (0.0000)	0.114 (0.0005)	0.024 (0.0000)	0.026 (0.0003)	0.050 (0.0000)	0.041 (0.0011)
	0.933	0.933	0.225	0.225	0.896	0.896

	<i>120448</i>	<i>120448</i>	<i>120448</i>	<i>120448</i>	<i>120448</i>	<i>120448</i>
SESA	0.151	0.163	0.058	0.068	0.099	0.066
	(0.0000)	(0.0012)	(0.0000)	(0.0009)	(0.0000)	(0.0032)
	0.913	0.913	0.218	0.218	0.932	0.932
	<i>87041</i>	<i>87041</i>	<i>87041</i>	<i>87041</i>	<i>87041</i>	<i>87041</i>
Panel G: excluding variables						
(previous performance at agency level)						
SBSA	0.141	0.133	0.022	0.024	0.046	0.047
	(0.0000)	(0.0005)	(0.0000)	(0.0003)	(0.0000)	(0.0011)
	0.933	0.933	0.225	0.225	0.896	0.896
	<i>120448</i>	<i>120448</i>	<i>120448</i>	<i>120448</i>	<i>120448</i>	<i>120448</i>
SESA	0.131	0.145	0.062	0.071	0.081	0.035
	(0.0000)	(0.0012)	(0.0000)	(0.0010)	(0.0000)	(0.0036)
	0.913	0.913	0.218	0.218	0.932	0.932
	<i>87041</i>	<i>87041</i>	<i>87041</i>	<i>87041</i>	<i>87041</i>	<i>87041</i>
Panel H: excluding variables						
(competition at the sector and division level)						
SBSA	0.137	0.128	0.022	0.027	0.059	0.061
	(0.0000)	(0.0005)	(0.0000)	(0.0003)	(0.0000)	(0.0010)
	0.933	0.933	0.225	0.225	0.896	0.896
	<i>120448</i>	<i>120448</i>	<i>120448</i>	<i>120448</i>	<i>120448</i>	<i>120448</i>
SESA	0.127	0.141	0.051	0.072	0.095	0.046
	(0.0000)	(0.0011)	(0.0000)	(0.0009)	(0.0000)	(0.0031)
	0.913	0.913	0.218	0.218	0.932	0.932
	<i>87041</i>	<i>87041</i>	<i>87041</i>	<i>87041</i>	<i>87041</i>	<i>87041</i>
Buyer* Year FE		✓		✓		✓
Market Controls		✓		✓		✓
Contract Controls		✓		✓		✓

Notes: Results for the average treatment effect on the treated of the inverse probability weighting regression adjustment on the two main outcomes, *Extra Cost* and *Delay*. Each panel refers to a different robustness check. The treated are contracts awarded with some kind of restricted solicitations, in the top panel contracts are set aside for small businesses (SBSA), in the bottom they are set aside for other type of small businesses (SESA). Several controls accounting for buyer, contract and market characteristics 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 (using only pre-treatment

buyer, contract and market characteristics). Standard errors are adjusted with 1,000 bootstrap repetitions.

Table A8: Sample Splits

	Log(# Offers)		Extra Cost		Delay	
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: baseline sample						
SBSA	0.139	0.132	0.024	0.027	0.054	0.051
	(0.0000)	(0.0005)	(0.0000)	(0.0003)	(0.0000)	(0.0011)
	0.933	0.933	0.225	0.225	0.896	0.896
	<i>120440</i>	<i>120440</i>	<i>119672</i>	<i>119672</i>	<i>120424</i>	<i>120424</i>
SESA	0.152	0.176	0.060	0.073	0.093	0.056
	(0.0000)	(0.0012)	(0.0000)	(0.0009)	(0.0000)	(0.0034)
	0.913	0.913	0.218	0.218	0.932	0.932
	<i>87036</i>	<i>87036</i>	<i>86593</i>	<i>86593</i>	<i>87027</i>	<i>87027</i>
Panel B: > 14 days						
SBSA	0.089	0.089	0.020	0.021	0.026	0.027
	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0001)
	0.757	0.757	0.170	0.170	0.651	0.651
	<i>356043</i>	<i>356043</i>	<i>353558</i>	<i>353558</i>	<i>354902</i>	<i>354902</i>
SESA	0.133	0.131	0.052	0.050	0.065	0.063
	(0.0000)	(0.0001)	(0.0000)	(0.0001)	(0.0000)	(0.0007)
	0.695	0.695	0.147	0.147	0.658	0.658
	<i>231820</i>	<i>231820</i>	<i>230413</i>	<i>230413</i>	<i>230778</i>	<i>230778</i>
Panel C: > 7 days and > \$ 25 K						
SBSA	0.123	0.123	0.022	0.022	0.053	0.053
	(0.0000)	(0.0000)	(0.0000)	(0.0001)	(0.0000)	(0.0002)
	0.943	0.943	0.218	0.218	0.928	0.928
	<i>124711</i>	<i>124711</i>	<i>123914</i>	<i>123914</i>	<i>124694</i>	<i>124694</i>
SESA	0.149	0.147	0.055	0.053	0.061	0.057
	(0.0000)	(0.0002)	(0.0000)	(0.0001)	(0.0000)	(0.0006)
	0.929	0.929	0.211	0.211	0.965	0.965
	<i>90377</i>	<i>90377</i>	<i>89910</i>	<i>89910</i>	<i>90366</i>	<i>90366</i>
Panel C: > 30 days and > \$ 25 K						
SBSA	0.122	0.122	0.022	0.022	0.074	0.073
	(0.0000)	(0.0000)	(0.0000)	(0.0001)	(0.0000)	(0.0003)
	0.923	0.923	0.240	0.240	0.866	0.866
	<i>111163</i>	<i>111163</i>	<i>110432</i>	<i>110432</i>	<i>111151</i>	<i>111151</i>
SESA	0.142	0.140	0.058	0.057	0.099	0.095

	(0.0000)	(0.0001)	(0.0000)	(0.0001)	(0.0000)	(0.0008)
	0.897	0.897	0.233	0.233	0.900	0.900
	80614	80614	80187	80187	80607	80607
Panel E: entire sample						
SBSA	0.085	0.085	0.021	0.021	-0.261	-0.264
	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0004)
	0.760	0.760	0.142	0.142	2.004	2.004
	431036	431036	428198	428198	429320	429320
SESA	0.121	0.118	0.047	0.045	-0.402	-0.403
	(0.0000)	(0.0001)	(0.0000)	(0.0001)	(0.0000)	(0.0009)
	0.705	0.705	0.120	0.120	2.328	2.328
	288616	288616	286969	286969	287067	287067
Buyer* Year FE		✓		✓		✓
Market Controls		✓		✓		✓
Contract Controls		✓		✓		✓

Notes: Results for the average treatment effect on the treated of the inverse probability weighting regression adjustment on the two main outcomes, *Extra Cost* and *Delay*. Each panel refers to a different sample selection robustness check. The treated are contracts awarded with some kind of restricted solicitations, in the top panel contracts are set aside for small businesses (SBSA), in the bottom they are set aside for other type of small businesses (SESA). Several controls accounting for buyer, contract and market characteristics 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 (using only pre-treatment buyer, contract and market characteristics). Standard errors are adjusted with 1,000 bootstrap repetitions.

B Appendix: Additional Information on Set-Asides

Other prominent SESA programs besides the VDSA 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 SESAs tenders if they have gone through a certification process. A firm that is certified for SESAs is automatically eligible to participate in SBSA solicitations.

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 SESA. These percentages change for the DVA, mostly due to the VDSA program—a major SESA category: only 12.8% are awarded through SBSA, and 46.8% are awarded through SESA.

Table B1: Types of Set-Asides

Set-aside Type	All Departments		Dept. Veteran Affairs (VA)		All Dept. (Except VA)	
	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 used in the sample. Note that *NSA* stands for unrestricted solicitations. *SBSA* for solicitations set aside for small businesses. *SESA* include other types of restricted solicitations, i.e. socioeconomic programs, such as those set-aside for women-owned small businesses or disadvantaged small businesses.

C Appendix: Long-Run-Analysis Figures

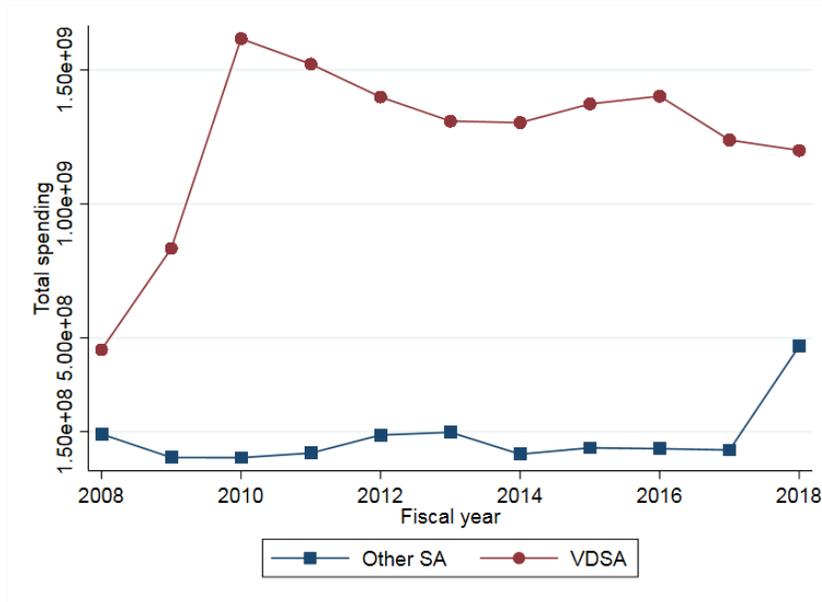


Figure C1: Total Spending in VDSA vs Other Set-Asides by the DVA, Construction

Notes: This figure reports the total yearly spending by the Department of Veteran Affairs on construction and maintenance. We plot separately the amount spent on Veteran-Disabled Set Aside (VDSA) and the spending on all other types of small business set-asides (SBSA) and other socioeconomic set-asides (SESA).

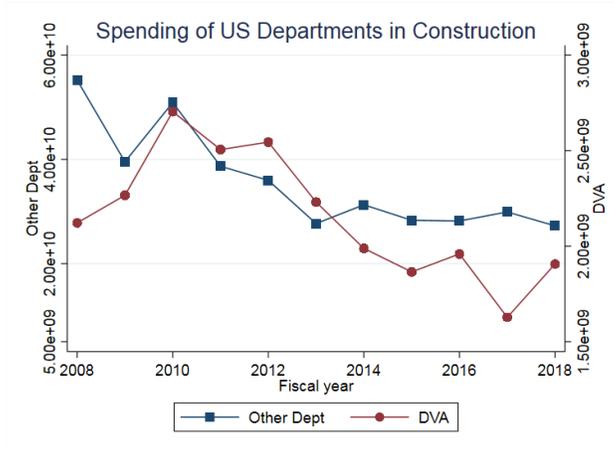


Figure C2: Total Spending by the DVA vs all Other Departments, Construction

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

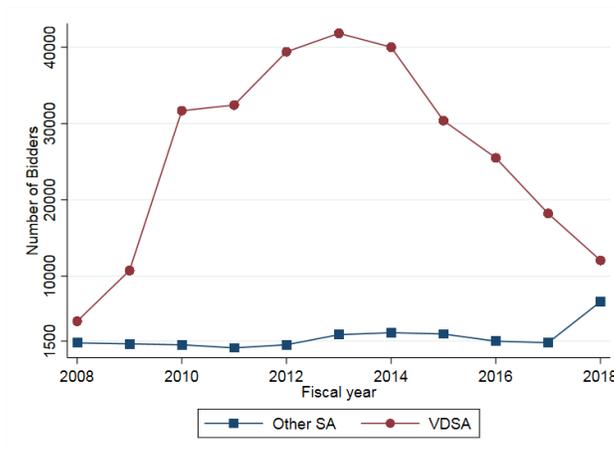


Figure C3: Number of bids in VDSA vs Other Set-Asides for contracts awarded by the DVA, Construction

Notes: This figure reports the total number of bids awarded by Department of Veteran Affairs on construction and maintenance. We plot separately the bids on Veteran-Disabled Set Aside (VDSA) and on all other types of small business set-asides (SBSA) combined with other socioeconomic set-asides (SESA).

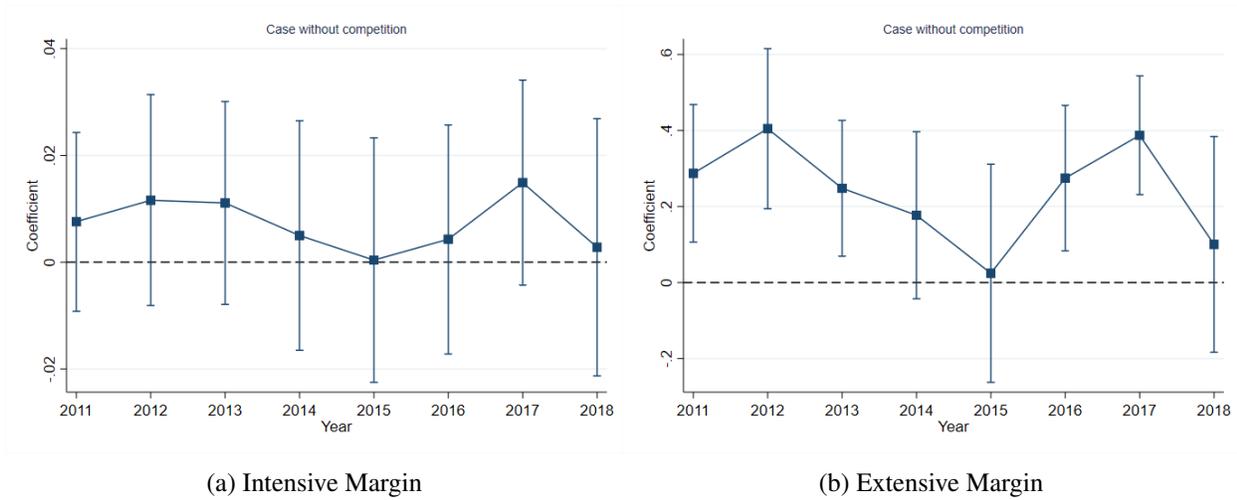


Figure C4: Non-Set-Aside Share

Notes: This figure reports the coefficients and confidence interval of the main regressor $SetAside_k^m$ interacted by fiscal year fixed effect. In panel (a) $SetAside_k^m$ is the log total set-aside award value by the firm k . In panel (b), it is a dummy indicating at least one set-aside contract at time t . The outcome *Non-Set-Aside Share* refers to the log share of non-set-aside awards over total awards.

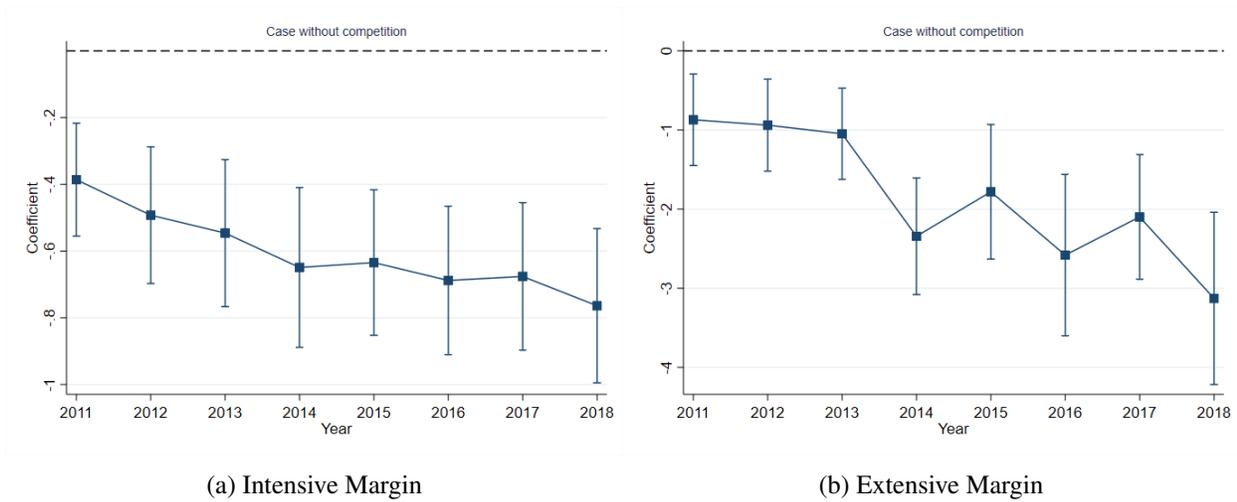
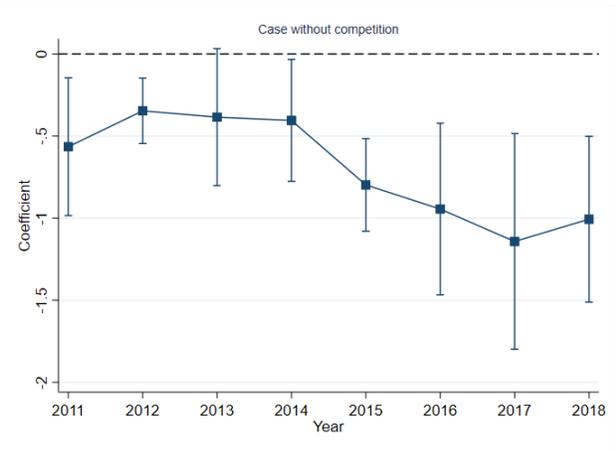
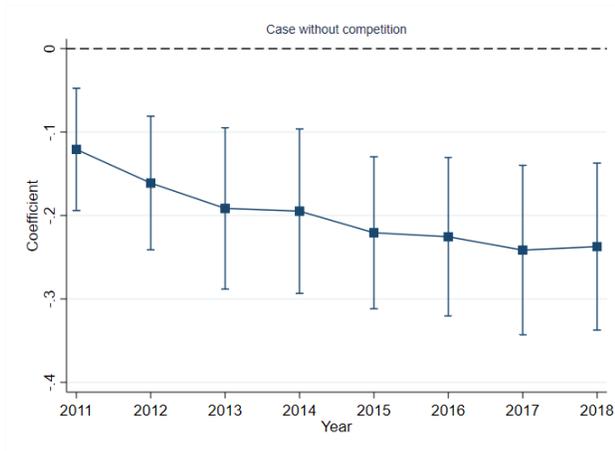


Figure C5: Multiple Products

Notes: This figure reports the coefficients and confidence interval of the main regressor $SetAside_k^m$ interacted by fiscal year fixed effect. In panel (a) $SetAside_k^m$ is the log total set-aside award value by the firm k . In panel (b), it is a dummy indicating at least one set-aside contract at time t . The outcome *Multiple Products* refers to the number of different procurement categories associated with a firm sales.

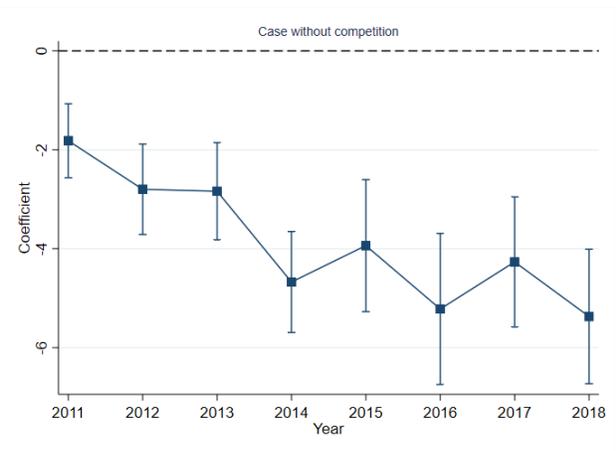
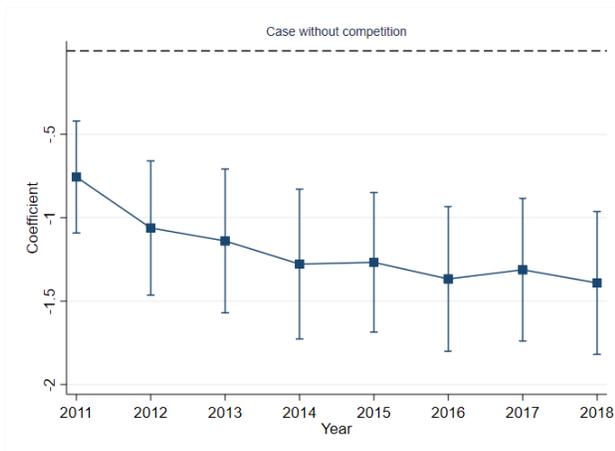


(a) Intensive Margin

(b) Extensive Margin

Figure C6: Multiple States

Notes: This figure reports the coefficients and confidence interval of the main regressor $SetAside_k^m$ interacted by fiscal year fixed effect. In panel (a) $SetAside_k^m$ is the log total set-aside award value by the firm k . In panel (b), it is a dummy indicating at least one set-aside contract at time t . The outcome *Multiple states* is the number of states in which a firm performed its activity.

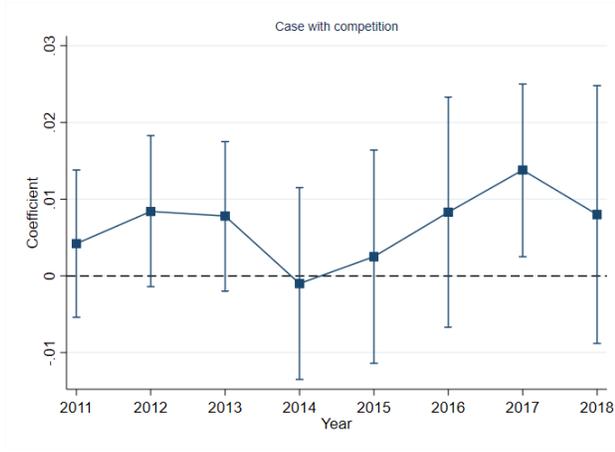


(a) Intensive Margin

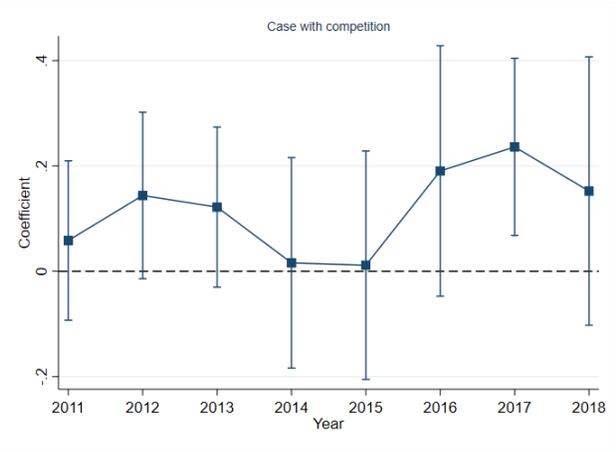
(b) Extensive Margin

Figure C7: Total Sales

Notes: This figure reports the coefficients and confidence interval of the main regressor $SetAside_k^m$ interacted by fiscal year fixed effect. In panel (a) $SetAside_k^m$ is the log total set-aside award value by the firm k . In panel (b), it is a dummy indicating at least one set-aside contract at time t . The outcome *Total Sales* is the total sales to the government.



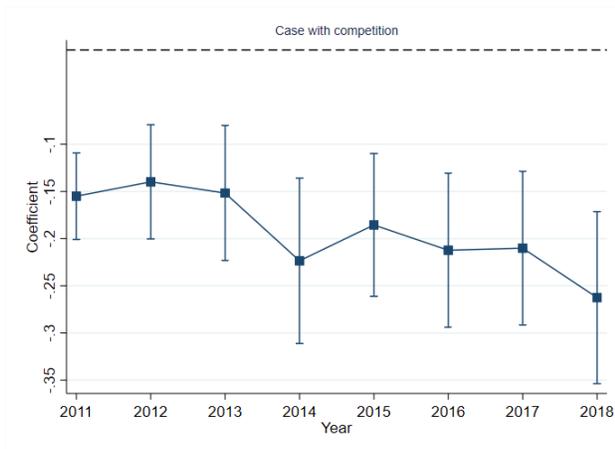
(a) Intensive Margin



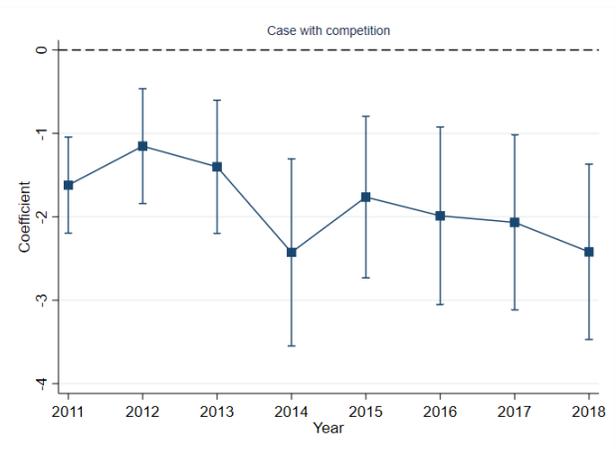
(b) Extensive Margin

Figure C8: Non-Set-Aside Share (Competed Tenders)

Notes: This figure reports the coefficients and confidence interval of the main regressor $SetAside_k^m$ interacted by fiscal year fixed effect. In panel (a) $SetAside_k^m$ is the log total set-aside award value by the firm k . In panel (b), it is a dummy indicating at least one set-aside contract at time t . The outcome *Non-Set-Aside Share* refers to the log share of non-set-aside awards over total awards. We define a tender as “competed” when at least two offers were placed in the auction.



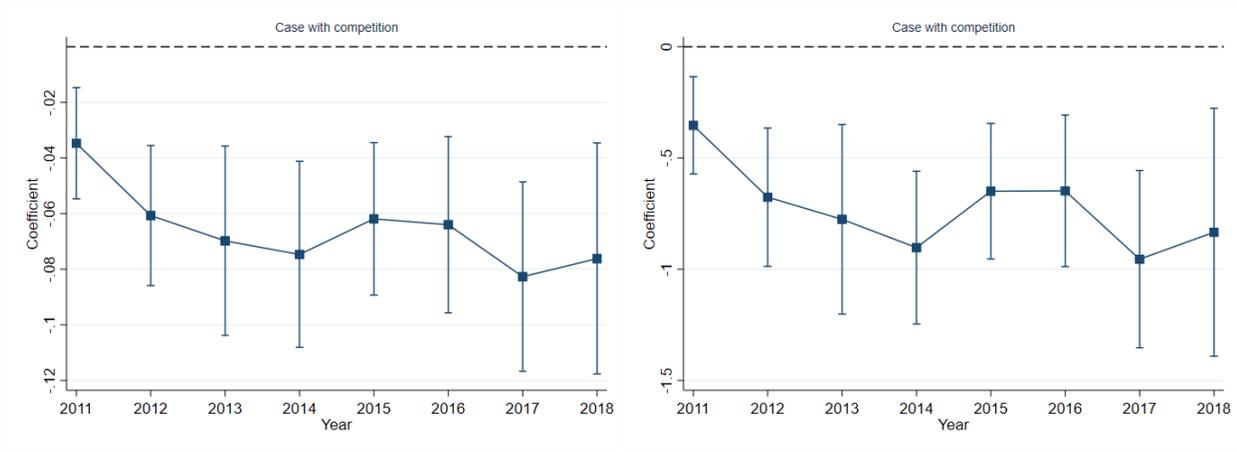
(a) Intensive Margin



(b) Extensive Margin

Figure C9: Multiple Products (Competed Tenders)

Notes: This figure reports the coefficients and confidence interval of the main regressor $SetAside_k^m$ interacted by fiscal year fixed effect. In panel (a) $SetAside_k^m$ is the log total set-aside award value by the firm k . In panel (b), it is a dummy indicating at least one set-aside contract at time t . The outcome *Multiple Products* refers to the number of different procurement categories associated with a firm sales. We define a tender as “competed” when at least two offers were placed in the auction.

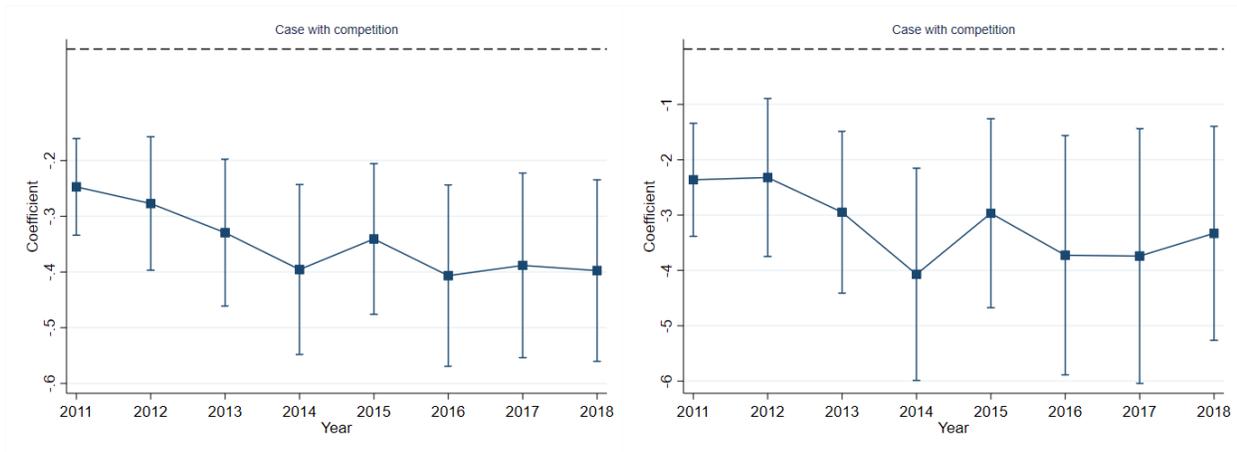


(a) Intensive Margin

(b) Extensive Margin

Figure C10: Multiple States (Competed Tenders)

Notes: This figure reports the coefficients and confidence interval of the main regressor $SetAside_k^m$ interacted by fiscal year fixed effect. In panel (a) $SetAside_k^m$ is the the log total set-aside award value by the firm k . In panel (b), it is a dummy indicating at least one set-aside contract at time t . The outcome *Multiple states* is the number of states in which a firm performed its activity. We define a tender as “competed” when at least two offers were placed in the auction.



(a) Intensive Margin

(b) Extensive Margin

Figure C11: Total Sales (Competed Tenders)

Notes: This figure reports the coefficients and confidence interval of the main regressor $SetAside_k^m$ interacted by fiscal year fixed effect. In panel (a) $SetAside_k^m$ is the the log total set-aside award value by the firm k . In panel (b), it is a dummy indicating at least one set-aside contract at time t . The outcome *Total Sales* is the total sales to the government. We define a tender as “competed” when at least two offers were placed in the auction.

D Appendix: Details on the Contract-Level Empirical Strategy

Random Forest Random forest is a machine learning method developed by Breiman (2001) whose 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 buyer, contract, and market characteristics. Random forests build on decision trees. A decision tree is 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 $p(W_i)$, where W_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 D1, 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. 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*.⁴⁴

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. The first method is the minimal depth. The lower the number of minimal depth, the more important is the variable. The number indicates the depth of the node - the closer the node to the root of the tree, the lower the minimal depth. Hence, a low minimal depth means that the variable

⁴⁴To 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.

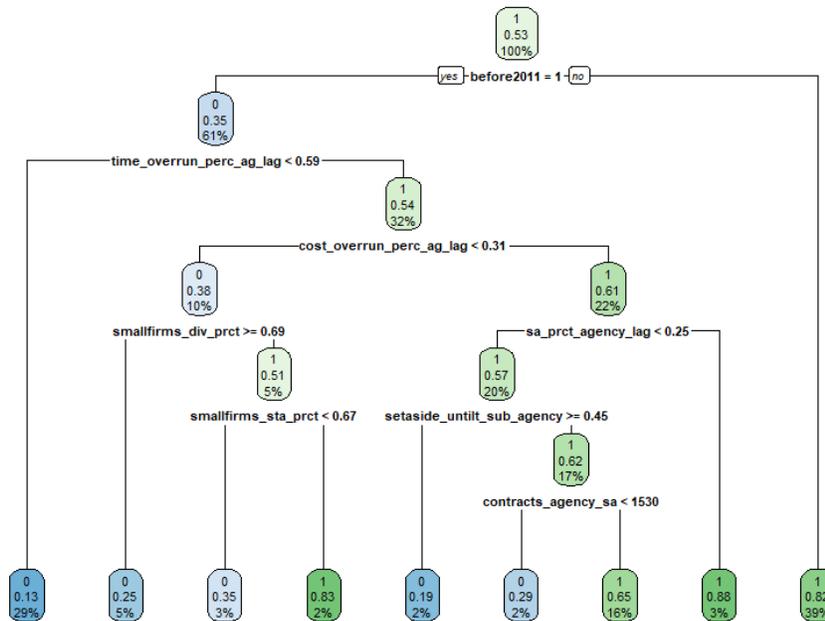


Figure D1: Example of a classification tree

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 D2.

splits many observations into two groups. In the example reported in Figure D1, *Before2011* would have depth 1. The second method is the mean decrease in accuracy, which we can obtain by looking at the loss in accuracy (i.e., the increase in the out-of-bag error) when permuting the given variable (Liaw et al., 2002). The third method is the mean decrease Gini, which we can obtain by looking at the decrease in Gini, a measure of the purity of the node. The larger the Gini decrease or impurity, the more important the variable is. At every split, the Gini impurity for the two subsequent nodes is less than the Gini impurity of the previous one. By adding up the Gini decreases for each individual variable over all trees, we obtain the mean decrease Gini, which is often consistent with the permutation importance measure (Breiman and Cutler, 2011 and Hao et al., 2010). The mean decrease Gini is usually the preferred method because of its low computational costs. Hence, we also employ the mean decrease Gini criterion for selecting the most relevant variable at each node.

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 a stable importance of variables.⁴⁵ When growing more

⁴⁵Note that we use the estimate of the out-of-bag error to set the optimal number of trees. Since we observe that the lowest number for the out-of-bag error is achieved when growing 600 trees or more, 600 is the optimal number of trees. At this point,

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 5,000 trees, as we are interested in the determinants of the treatment. Thus, it is essential to have stable variable importance.

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 80% of the original dataset, and we keep the remaining 20% 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., 11 since we have 114 variables). Then, we use half as many variables, and finally, twice as many. We commit to this value since we obtain the lowest prediction error for 11 variables. Note, however, that we estimate the propensity score for the three different numbers of randomly selected variables and obtain stable estimates of the ATET. This shows that the results are not sensitive to the different specifications of the propensity score.⁴⁶

Treatment Predictors In this subsection, we identify key variables for the decision to restrict competition to small firms. Predicting the propensity score using random forests enables us to do so. Furthermore, we compare three different measures of variable importance.

In Figure D2, 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. We find that the most critical factors in predicting the treatment are the different measures and proxies for the level of competition. Contracting officers seem to be influenced by the share of competitive contracts out of all contracts. Also among the most influential factors is the *Non-Set-Aside Sector*, namely, the share of small firms always winning without set-aside out of all small firms in a sector (and in a sector per year). Finally, we observe that variables indicating how much an agency has awarded through set-aside until the time of the award of a given contract in the FY also play an important role.⁴⁷

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. We use importance as a variable selection tool for

the error equals 15%, which means that each observation will be classified wrongly 15% of the time. We report this step only for completeness because we grow 5,000 trees instead of only 600, given that we are interested in stable variable importance. The only downside to growing more trees is computational costs.

⁴⁶We want to be extra careful and make sure that the results are robust because they are sensitive to the different propensity score specifications when using inverse probability weighting (Zhao, 2004). Since we implement the doubly-robust IPWRA, however, the model should be less influenced by the specification of the propensity score.

⁴⁷Note that this is true both for the amount awarded and the number of contracts awarded through set-aside.

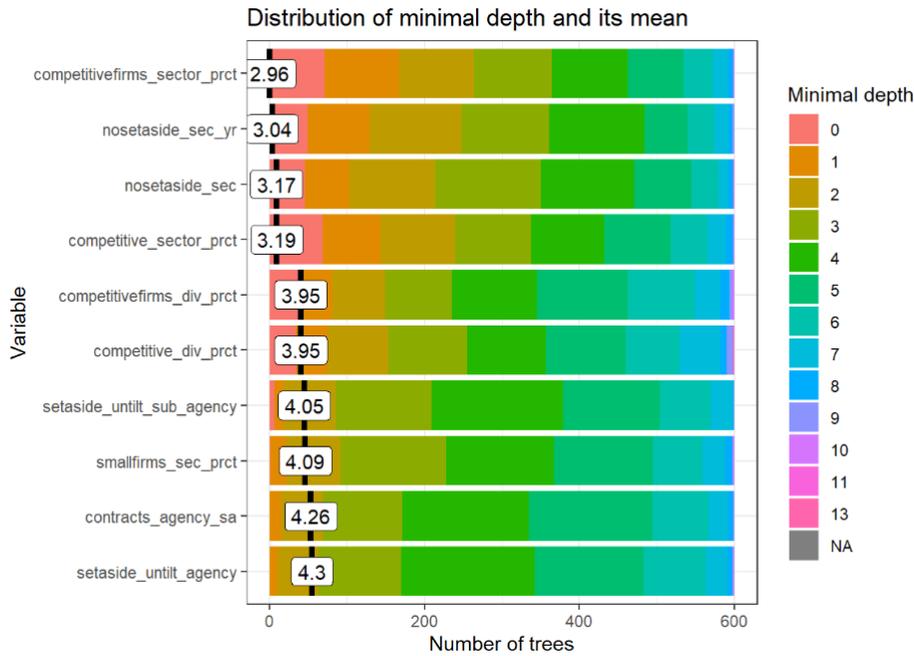


Figure D2: Variable Importance According to Minimal Depth

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

guiding one of the robustness checks of the IPWRA method. 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.

Since *Non-Set-Aside Sector* is reported by all three criteria as among the first five covariates, we can safely state that it is one of the driving factors in the choice of setting aside contracts. When observing the results reported of the mean decrease in accuracy criterion, we notice that variables can be divided into four importance levels. For instance, we see a large gap between the first and the second-ranked variables, namely, the cumulative sum of competitive contracts awarded in the same sector and the *Non-Set-Aside Sector Year*. Several indicators of the competitiveness of the market also play a major role. They are mostly located in the second and third levels of importance. Looking at differences between the orders of importance reported by the three measures, one variable stands out: the ratio of small firms per state and sector. While it appears fairly at the top for the mean decrease in accuracy, it shows up at a much later stage for the mean decrease Gini. Keeping in mind that the mean decrease Gini is computed similarly to the minimal depth, it does not seem surprising then that the selected important variables are almost the same and have the same order for both measures. For the mean decrease Gini method, it is worth noting that the time dimension, namely whether the contract is awarded before or after 2011, is ranked tenth and, thus, estimated to be an important variable for predicting the treatment.

Variable importance analysis does not detect corruption, bid collusion, or similar instances. Therefore, we cannot rule out something of this kind. However, we can carefully state that, on average, contracting

officers seem to follow the rules imposed by the FAR when setting aside contracts, although contracting officers are allowed some degree of discretion.

IPWRA The second stage of our empirical strategy relies on two main assumptions:

- 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.

$$Y(0), Y(1) \perp\!\!\!\perp D \mid W, \forall W \quad (4)$$

- 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.

$$0 < P(D = 1|W) < 1 \quad (5)$$

In Section VI, 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 ATET. 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:

$$\hat{\Delta}_{ct} = \frac{\bar{W}_t - \bar{W}_c}{\sqrt{\frac{\sigma_c + \sigma_t}{2}}} \quad (6)$$

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

$$s_{weight}^2 = \frac{\sum \omega_i}{(\sum \omega_i)^2 - \sum \omega_i^2} \sum \omega_i (W_i - \bar{W}_{weight})^2 \quad (7)$$

The main idea of IPWRA is to combine the inverse probability weighting method with regression adjustment to obtain a doubly-robust estimator. The robustness of this approach lies in the fact that we only need one of the following to be correctly specified: (i) the conditional mean or (ii) the propensity score. Thus, this weighting has the advantage that if one of the following is true:

- if the regression model is *incorrect*, but the propensity model is *correct*.
- if the propensity model is *incorrect*, but the regression model is *correct*.

, then the estimated ATET is a consistent estimator of the true ATET for the relevant outcomes of interest.

Clearly, if they are both correct, $\hat{\Delta}$ will still be a consistent estimator. But if they are both incorrect, $\hat{\Delta}$ will be inconsistent.

In other words, the doubly-robust estimator offers protection against misspecification. Moreover, in large samples, if the propensity score is correctly specified, the model will have a smaller variance than a simple inverse weighted estimator. The opposite is true, i.e. the model will have a larger variance, if the regression model is specified correctly. However, as explained above, the advantage of using the doubly-robust estimator is that it offers protection in case one of the two models is not correctly specified.

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

$$\min_{\delta_1, \beta_1} \sum_{i=1}^N SetAside_i (outcome_i - \delta_1 - \beta_1 X_i)^2 * ipw_i \quad (8)$$

and

$$\min_{\delta_0, \beta_0} \sum_{i=1}^N (1 - SetAside_i) (outcome_i - \delta_0 - \beta_0 X_i)^2 * ipw_i. \quad (9)$$

Having predicted the outcome values for both treated and untreated contracts, we next estimate the ATET using regression adjustment as follows:

$$ATET = N^{-1} \sum_{i=1}^N SetAside_i [(\hat{\delta}_1 + \hat{\beta}_1 X_i) - (\hat{\delta}_0 + \hat{\beta}_0 X_i)], \quad (10)$$

where $W_{i,z,b}$ and $X_{i,z,b,s}$ denote the vectors of covariates for each individual. As explained earlier, we specify two different sets of covariates, $W_{i,z,b}$ for the propensity score estimation with only *ex ante* observables and $X_{i,z,b,s}$, where $W_{i,z,b} \in X_{i,z,b,s}$.

E Appendix: The Additional Procurement Outcomes

As mentioned in Section V, 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.⁴⁸ Note that we consider a renegotiation as *in-scope* if the reason for the modification is consistent with the original contract terms.⁴⁹ *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 *Number of 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 *Number of 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 contracts is that we do not consider homogeneous goods and lack a proper set of contracts for comparative purposes.

⁴⁸ 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).

⁴⁹ 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).

Online Appendix: List of Treatment Predictors

The control variables we use to predict treatment (i.e., SBSA or SESA) can be divided into three main categories: Market, Buyer, and Contract Characteristics.⁵⁰ We define the market for a given procurement as the industry associated with the procurement category (i.e., NAICS code) and the state of performance. First, under the category of market characteristics, we identify the competitiveness of small firms in the region and industry. In a competitive market, buyers might set aside more contracts because the rule of two is easier to satisfy.⁵¹ We also control for the percentage of small firms in a sector that win without the type of set-aside under study and the percentage of small firms per year in a given NAICS sector, region, and state. We define region as the four geographical units of the US, namely West, Midwest, South and Northeast.

Second, the category of buyer characteristics that include variables measuring prior experience⁵². We construct a proxy that takes into account the agency's experience with set-asides. We choose this proxy to control for the contracting agency's performance on past delays and additional costs of awarded contracts.⁵³

In addition, contracting officers might be encouraged to set aside more procurements if the agency falls behind its annual goal in contracting.⁵⁴ We also include 66 dummies for each awarding agency to account for time-invariant characteristics of the awarding agency and the propensity of awarding agencies to set-aside contracts.

Third, in the contract characteristics category, we include *FY*, a categorical variable for 2008-2018, *MultiYear*, an indicator that equals 1 if the contract is expected to last longer than one year, and two variables, *LastWeek* and *LastMonth*, that indicate whether the contract is awarded in the last week or month of the FY.⁵⁵

In addition, *Public Works* is a dummy that equals 1 if the contract is for the performance of a public works project, and *Oversight* is a dummy that equals 1 if the contract is subject to supervision by the contracting authority (in addition to the surety company via performance bonds if the contract is construction-

⁵⁰See the full list for each category and all variable descriptions in Appendix A. We cannot include seller information in this step because the winning firm is not determined until after the treatment. Nonetheless, we include such variables in subsequent steps as controls for the contract outcome

⁵¹We believe that a good proxy for the degree of competition is the percentage of competitive small firms to all small firms, both in the sector and in the contract division. We define a firm as competitive if (i) it is small and (ii) it has won at least one tender where (iii) at least two bids are submitted.

⁵²It is often impractical to specify complete contracts (Hart and Moore, 1988). However, some buyers may be better at drafting more complete contracts, resulting in less renegotiation.

⁵³We include the fraction of set-aside contracts that reported cost or schedule overruns prior to each award in all set-aside contracts.

⁵⁴We control for this with *Set Aside Until t*, the cumulative sum of set-aside contracts divided by the cumulative sum of all contracts per agency (and per sub-agency) until the start of the given contract *i*. Note that the given contract *i* is not included. Due to the possibility of time-invariant agency characteristics, i.e., a tendency to award or not award set-aside contracts, we include a lag variable of the share of the previous FY, i.e., the share of set-aside contracts out of all contracts awarded by a given agency (and sub-agency)

⁵⁵Other time-related variables are *Before2011*, which indicates whether the contract is awarded before 2011, and *Period*, 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

related).⁵⁶ Finally, we include *Quartile Award Amounts* 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.⁵⁷ Such a proxy is critical because complex contracts are less likely to be set aside.

Seller Characteristics

- *active_years*: Counts how many years the firm has been active, at any time t . It equals the time of the contract minus the time of the first contract in the dataset.
- *alreadydelay*: Dummy indicating whether there was a delay in the past.
- *alreadyoverrun*: Dummy indicating whether there was already a cost overrun in the past.
- *alreadyundercost*: Dummy indicating whether the firm completed the contract cheaper in the past.
- *atleastsetaside*: Dummy indicating whether the firm has won at least one set-aside in total.
- *atleastsetaside_year*: Dummy indicating whether there was already a set-aside in the past (if the contract i is won with a set-aside, this contract is not included).
- *controlled*: Dummy indicating whether the firm is controlled by a parent firm.
- *delayprc_firm*: Ratio of the number of contracts performed by the firm with delays out of the total number of contracts performed by the same firm.
- *fasterfirm*: Firm completed a contract faster than expected in the past.
- *overrunprc_firm*: Ratio of the number of contracts performed by the firm with extra costs out of the total number of contracts performed by the same firm.
- *negdelaysprc_firm*: Ratio of the number of contracts performed ahead of time by the firm out of the total number of contracts performed by the same firm.
- *partnership*: Dummy indicating whether the seller is a partnership.
- *p25_rec_ag_avgwon**: Quartile by fiscal year of the average amount won by the recipient during the year per agency.
- *p25_rec_ag_num**: Quartile by fiscal year of the number of contracts won by the recipient during the year per agency.

⁵⁶Giuffrida and Rovigatti (2022) and Calvo et al. (2019), and Carril (2020) empirically investigate the effect of supervision on contract outcomes and report in more detail the different types of supervision. Following their approach, we run separate regressions for different subgroups to examine the effect of oversights on procurement outcomes.

⁵⁷Here we use the four-digit NAICS code to classify industries based on information about the product or service.

- p25_rec_ag_won*: Quartile by fiscal year of the amount won by the recipient during the year per agency.
- p25_rec_yr_won*: Quartile by fiscal year of the amount won by the recipient during the year.
- p25_rec_div_avgwon*: Quartile by fiscal year of the average amount won by the recipient during the year per division.
- p25_rec_div_num*: Quartile by fiscal year of the number of contracts won by the recipient during the year per division.
- p25_rec_div_won*: Quartile by fiscal year of the amount won by the recipient during the year per division.
- p25_rec_sec_avgwon*: Quartile by fiscal year of the average amount won by the recipient during the year per sector.
- p25_rec_sec_num*: Quartile by fiscal year of the number of contracts won by the recipient during the year per sector.
- p25_rec_sec_won*: Quartile by fiscal year of the amount won by the recipient during the year per sector.
- p25_rec_yr_avgwon*: Quartile by fiscal year of the average amount won by the recipient during the year.
- p25_rec_yr_num*: Quartile by fiscal year of the number of contracts won by the recipient during the year.
- public: An indicator variable equal to one if the seller is a public entity or a publicly-held company.
- samedivision: An indicator variable equal to one if recipient division and performance division coincide.
- sameregion: An indicator equal to one if the recipient region and performance region coincide.
- samestate: An indicator equal to one if the recipient state and performance state coincide.
- undercostprc_firm: Ratio of the number of contracts performed below costs by the firm out of the total number of contracts performed by the same firm.

Contract Characteristics

- *before2011*: Indicator equal to 1 if the contract was awarded before 2011.

- *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 (DoD) Agency.
- *construction*: An indicator variable equal to one if the contract is based on the execution of a construction project (service code is starting with Y).
- *construction_maintenance*: An indicator variable equal to one if the contract is based on the execution of a construction or maintenance project (service code is starting with either Y or Z).
- *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 because of its size or scope.
- *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.
- *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.
- *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.
- *multiyear*: Indicator variable equal to one if the expected duration is over 365 days.
- *nationalinterest*: Indicator variable equal to one if the contract is created for the national interest, e.g. projects to limit damages provoked by hurricanes.
- *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.
- *performance_based_service_acqu**: This variable describes the requirements in terms of results required rather than the methods of performance of the work b. Uses measurable performance standards (i.e. terms of quality, timelines, quantity etc.) and quality assurance surveillance plans (refer to 46.103(a) and 46.401(a)) c. Specifies procedures for reductions of fee or for reductions to the price of a fixed-price contract when services are not performed or do not meet contract requirements (refer to 46.407) d. Includes performance incentives where applicable. For FPDS reporting purposes, a minimum of 80% of the anticipated obligations under the procurements action must meet the above requirements for FY 2004 and prior and a minimum of 50% of the anticipated obligations under the procurements action must meet the above requirements for FY 2005 and later.
- *performancedivision**: Indicator for each of the nine divisions (Pacific, Mountain, South Central, etc.) of the place of performance of the contract.
- *performanceregion**: Indicator for each of the four regions (West, South, Midwest, and North East) of the place of performance of the contract.
- *period**: Variable accounting for the Great Recession. It equals to -1 if the contract was awarded before December 2007. It equals to 1 if the contract was awarded after June 2009. And it equals to 0 if the contract was awarded in between these periods.
- *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.

Market Characteristics

- *competitive_div_prct*: Share of competitive contracts out of all contracts awarded to small firms.
- *competitive_sector_prct*: Share of competitive contracts out of all contracts awarded to small firms.
- *competitivefirms_sector_prct*: share of small competitive firms out of all small firms per year and per sector.
- *competitivefirms_div_prct*: Share of competitive small firms out of all small firms per year.
- *cum_comp_contracts*: Cumulative sum of competitive contracts out of all contract for each year.⁵⁸

⁵⁸Cumulative measures always start from the first contract of the current fiscal year.

- *nosetaside_sec*: Share of small firms in a sector that wins without set-aside (firm has never won with a set-aside) out of all small firms.
- *nosetaside_sec_some*: Share of small firms in a sector that wins without set-aside (although the same firm has won a contract with a set-aside before) out of all small firms.
- *nosetaside_sec_yr*: Share of small firms in a sector per fiscal year that win without set-aside (firm has never won with a set-aside) out of all small firms in that sector per year.
- *nosetaside_sec_yr_some*: Share of small firms in a sector per fiscal year that win without set-aside (although the same firm has won a contract with a set-aside before) out of all small firms in that sector per year .
- *p25_workload**: quartiles per year of workload_untilt
- *p25_workload_sec**: quartiles per year per sector of workload_untilt
- *p25_workload_contracts**: quartiles per year of workload_contracts_untilt.
- *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.
- *smallfirms_sec_prct*: Share of small firms per sector per year.
- *smallfirms_div_prct*: Share of small firms per division per year.
- *smallfirms_reg_prct*: Share of small firms per region per year.
- *smallfirms_sta_prct*: Share of small firms per state per year.
- *workload_contracts_untilt*: Cumulative count of all active contracts for the winning firm (considering only contracts active at the action date, time t , of the other contract i).
- *workload_untilt*: Cumulative sum of all active contracts for the winning firm (considering only contracts active at the action date, time t , of the other contract i).

Buyer Characteristics

- *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.
- *contracts_agency_sa*: Cumulative count of set-aside contracts grouped by awarding agency and year (and sorting by action date) .
- *cost_overrun_perc_ag_lag*: Lag of average cost overrun per agency per year (average of *Delay (Ratio)*).
- *sa_prct_agency_lag*: Ratio of the previous fiscal year, i.e. set-aside contracts out of all contracts awarded by a given agency.
- *sa_prct_sub_agency_lag*: Ratio of previous fiscal year, i.e. set-aside contracts out of all contracts awarded by a given subagency.
- *time_overrun_perc_ag_lag*: Lag of average time overrun per agency per year (average of *Extra Cost (Ratio)*).
- *years_firm_agency*: Number of years in which the same firm wins with the same agency.



Download ZEW Discussion Papers:

<https://www.zew.de/en/publications/zew-discussion-papers>

or see:

<https://www.ssrn.com/link/ZEW-Ctr-Euro-Econ-Research.html>

<https://ideas.repec.org/s/zbw/zewdip.html>



IMPRINT

**ZEW – Leibniz-Zentrum für Europäische
Wirtschaftsforschung GmbH Mannheim**

ZEW – Leibniz Centre for European
Economic Research

L 7,1 · 68161 Mannheim · Germany

Phone +49 621 1235-01

info@zew.de · zew.de

Discussion Papers are intended to make results of ZEW research promptly available to other economists in order to encourage discussion and suggestions for revisions. The authors are solely responsible for the contents which do not necessarily represent the opinion of the ZEW.