// NO.23-035 | 09/2023

DISCUSSION PAPER

// MATILDE CAPPELLETTI, LEONARDO M. GIUFFRIDA, SOHVI HEATON, AND DONALD S. SIEGEL

Strategic Management in Public Procurement: The Role of Dynamic Capabilities in Equity and Efficiency

Strategic Management in Public Procurement:

The Role of Dynamic Capabilities in Equity and Efficiency[1]

August 7, 2023

Matilde Cappelletti, Leonardo M. Giuffrida, Sohvi Heaton, Donald S. Siegel

ABSTRACT

A key issue in strategic management in the public sector is how government creates economic and social value through procurement. Unfortunately, most procurement studies are based on contract theories, which fail to incorporate the growing role of strategic management in performance. We fill this gap by analyzing longitudinal data on contracting to assess the equity and efficiency effects of a form of affirmative action used by governments: set-aside programs. Employing a machine learning-augmented propensity score weighting approach, we find that set-aside contracts are negatively associated with contract performance. These effects are attenuated by an agency's dynamic capabilities and the extent to which the agency uses more competitive procedures. Our findings illustrate how the dynamic capabilities of a federal agency can simultaneously enhance equity and efficiency.

JEL codes: D73, H57, O38, L22.

Keywords: Dynamic capabilities, resource-based view, public procurement, machine learning, random forest.

[1] Cappelletti: ZEW Mannheim and University of Mannheim; email: matilde.cappelletti@zew.de. Giuffrida: ZEW Mannheim, MaCCi, CESifo; email: leonardo.giuffrida@zew.de. Heaton: Leavey School of Business; email: sheaton@scu.edu. Siegel: Arizona State University; email: donald.siegel.1@asu.edu.

1. INTRODUCTION

Value creation through procurement in the public sector should be based both on optimal contractual arrangements (Cabral et al., 2019), as well as how strategic managers effectively implement such contractual arrangements. Thus, understanding value creation in public procurement requires insights from both contract theory and strategic management.

Existing studies of public procurement have traditionally been based on transactions cost theory, i.e., how participants' minimize transaction costs, while holding other factors constant (e.g., Poppo and Zenger, 2002; Luo and Kaul, 2019; Li et al., 2018). Implementation and organizational issues relating to government contracting have not received sufficient attention in the literature on strategic management in the public sector. These studies often implicitly assume that public agencies have the requisite capabilities to implement contractual arrangements effectively. Unlike their counterparts in the private sector, public organizations are constrained by regulation and often influenced by multiple political stakeholder groups pursuing different objectives (Heaton et al., 2023). This increases complexity and uncertainty in implementation, which requires organizational flexibility and capabilities. In the private sector, strategic management is applied to maximize economic performance and profitability. However, in the public sector, achieving goals beyond cost effectiveness are central reasons for adopting strategic management (Bryson et al., 2018). Consequently, the extent of publicness may be an important moderator of the relevance of private sector theories and evidence to contract performance.

In this paper, we fill this void by applying the dynamic capabilities framework to the public sector and by investigating the role of dynamic capabilities in the performance of affirmative action programs. The U.S. federal government spends over $600 billion on contracts every year, roughly equivalent to the size of Sweden's economy (Guy and Ely, 2022). Governments participate in the market as a buyer and, at the same time, use their purchasing power to advance social goals (McCrudden, 2004). To help provide a level playing field in the public procurement market for small businesses—which may lack competitive skills—governments have introduced initiatives that limits competition for a certain portion of public contracts to small businesses. Such contracts are called "small business set-asides" (SBSA), and they aim to boost small businesses participation in the federal contracting. However, set-asides may entail a tradeoff between promoting the growth of small and economically disadvantaged businesses and simultaneously ensuring efficient results for the end user – government (Schwartz, 2010). This requires a delicate balance that is not easy to achieve (McCue et al., 2015).

As discussed, maximizing such tradeoffs within regulatory constraints, while being accountable to diverse stakeholder groups with varying expectations, can create complexity and ambiguity. To overcome these challenges, public organizations need capabilities suited to managing such tensions. The dynamic capabilities framework (Teece et al., 1997; Teece, 2007) can provide useful insights into this as dynamic capabilities may play an important role in balancing competing logics. Dynamic capabilities are particularly beneficial under conditions that involve a high level of volatility, uncertainty, complexity, or ambiguity (VUCA) as such conditions require leaders who can develop organizational capabilities that stimulate innovative offerings and new business models (Schoemaker et al., 2018, p. 16).

We extend a solution that stems from the key insight in both strategic management and public administration that resources and capabilities must be managed efficiently to create value. Some organizations are better at managing and deploying resources and dealing with the idiosyncratic environments of the public sector in pursuit of value (Quelin et al., 2019), as exemplified in recent studies. For example, Heaton et al. (2022) found that public universities in the U.S. differ in their capability to allocate resources flexibly, which ultimately affects financial performance. Zheng et al. (2019) showed how some organizations are more deeply embedded in their environments, so more effective in managing political capital.

We apply this logic to make two predictions about a federal agency's dynamic capabilities and contract performance, in the context of small business set-asides. Given that contracting in a non-competitive environment can result in less-desirable outcomes (Cappelletti and Giuffrida, 2022), we expect that set-aside contracts do not perform as well as conventional contracts. Due to the fact that management of the contracting process is crucial for governments to capture value from outsourcing (Adida and Bravo, 2019; Handley and Angst, 2015), we expect that a federal agency's flexibility in contracting will positively influence contract performance. Since highly competitive environments involve a high intensity of rivalry in a dynamic and uncertain environment, dynamic capabilities are likely to be most valuable in such environments, where their benefits are likely to outweigh their costs (Makadok, 2001). Therefore, we expect that a federal agency's flexibility in contracting is more beneficial when the agency deals with more contracts with full and open competition.

We test these predictions using a sample of 120,448 contracts (45 percent of which are set-asides) between U.S. federal agencies and private businesses. Using rich data on contract, contractors, and federal agency characteristics, we find support for our theoretical predictions. Thereafter, we consider a number of alternative data measurements and economic specifications, and find the results are robust to these approaches.

This study makes four contributions to the theoretical literature. First, our paper deepens understanding of strategic management in the public sector and how and under what conditions it can foster better performance. Public procurement is moving toward a strategic role of achieving specific government objectives (Walker, 2015) from an administrative role of meeting regulatory requirements (Patrucco et al., 2017). The underlying hypothesis in the literature is that public organizations that are more strategic generally achieve better outcomes. However, there are few empirical studies on public sector strategic management and its connection with implementation and performance (e.g., Bryson et al., 2010). Thus, our paper improves our understanding of strategy formulation and implementation in the public sector.

Second, while most analyses of government contracts have focused on R&D contracts (Lichtenberg, 1988), we extend the logic to consider performance variations in non-R&D government contracts. As Table 1 shows, the participation rates of small businesses in service and construction contracts are relatively high. In 2020, the U.S. government spent about $400 billion on service contracts (Potter, 2022), while it usually invests about $50 billion in R&D procurement (Rassenfosse et al., 2019).

Service and construction contracts may involve a lot of uncertainty, such as unexpected adverse shocks. Thus, previous studies have used delays and cost overruns as performance measures (e.g., Giuffrida and Rovigatti, 2022). For R&D contracts, particularly Small Business Innovation Research programs, researchers use a number of different measures such as commercialization of the results of the funded research (Link and Scott, 2010), patenting and job creation (Siegel and Wessner, 2012), and firm-level sales (Lerner, 1999). Given the differences, existing models upon which these studies were based might not be applicable to the case of non-R&D contracts.

Third, we also contribute to the literature on dynamic capabilities. The majority of empirical studies on dynamic capabilities have been conducted in industrial sectors such as technology and semiconductors (Cosmi, 2020). Limited research has been done in the public sector, with a few exceptions (e.g., Heaton et al., 2022). By focusing on public procurement as a new empirical context, we identify a new boundary condition for the effect of dynamic capabilities (the competitive contracting environment).

Finally, this study makes an empirical contribution by introducing a new measure of dynamic capabilities in the public sector and employing a novel methodology. To the best of our knowledge, this is the first time that government dynamic capabilities have been assessed over time. We also contribute to the burgeoning literature on the use of machine learning techniques in strategy (Tidhar and Eisenhardt, 2018; Choi et al., 2018; Raj and Seamans, 2019; Miric et al., 2022), through our use of random forests, a machine learning technique.

INSERT TABLE 1 ABOUT HERE

2. THEORY AND HYPOTHESES

2.1. Government Capabilities

Although procurement is generally a regulated activity and open competition the norm, agencies have some discretion in determining contract-award recipients (Blount et al., 2018). A growing number of studies have highlighted the importance of organizational capabilities for successful outsourcing (e.g., Brudney et al., 2005; Hefetz and Warner, 2004; Lamothe and Lamothe, 2010). In particular, extant research recognizes that the management capacity of a contracting agency is critical for successful procurement (Brown and Potoski, 2006).

A government's contracting capability is commonly measured by assessing how it manages the establishment of contracts and execution of these transactions. For example, Brown and Potoski (2003) argue that the level of investment in developing and maintaining contracting capabilities explains differences in procurement performance. They define contract management capacities as consisting of three types of specific capacities: feasibility assessment capacity, implementation capacity, and evaluation capacity. Prior negative contracting experiences and the types of projects involving higher transaction costs can influence the level of investment in contracting capabilities (Ekstrom and Selviaridis, 2014). Yang et al. (2009) identify four types of contracting capabilities: agenda setting capacity, contract formulation capacity, contract implementation capacity, and contract evaluation capacity. They show that such contracting capacities are all positively associated with contract performance measured by cost, quality, and efficiency gains (Ekstrom and Selviaridis, 2014). Ingrahm et al. (2003) identify four levers that determine the level of management capacity: administrative infrastructure and technology, leadership, integrating management systems into unified wholes, and a system of managing for results.

However, what is missing in such conceptualizations of a government's contracting capacity is the evolutionary dynamics (Vanneste and Puranam, 2010). The broader exchange governance literature suggests that contractual governance is complemented during a long-term exchange by relational mechanisms based on trust and flexibility (Klein-Woolthuis et al., 2005). Given such interactions, dynamics of contractor and buyer learning need to be considered.

2.2. Dynamic Capabilities

Teece et al. (1997) define dynamic capabilities as "the firm's ability to integrate, build, and reconfigure internal and external competences to address rapidly changing environments." Such capabilities reside in organizational routines and processes. Managers use specific organizational processes to alter and reshape their resource base (Eisenhardt and Martin, 2000).

The dynamic capabilities framework suggests that dynamic capabilities enable an organization to change efficiently to respond to new market demand, which ultimately enhances firm performance. For example, Zahra and George (2002) consider absorptive capability as a dynamic capability and show that it positively affects the sustainability of a firm's competitive advantage. Pablo et al. (2007) also show how a public sector organization using a dynamic capability (learning through experimenting) can improve organizational performance.

The dynamic capabilities framework provides the basis for arguing that a federal agency's routine to reshape its resource base by responding to past contractor performance is a source of dynamic capability, and such a capability makes operations more responsive to changing environments within the regulation constraints. The public procurement literature highlights that managing performance-based contracting is a learning process. For example, Heinrich and Choi (2007) explain how public organizations can use the experience gained and adapt financial incentives. Mayer and Argyres (2004) point out that organizations learn how to contract with each other as they learn about transaction-specific contingencies and hazards over time. Such learning can be used to modify and extend contractual provisions in subsequent contracts. Given this dynamic approach to contracting, contractor-specific knowledge can be used to adjust rewards or penalty outcomes to increase contract performance (Behn and Kant, 1999).

Contracting based on performance can be costly, which makes it difficult to apply in different contexts for several reasons. First, writing performance-based statements of work is difficult because such statements are unique to each project. Second performance-based contracting also requires a cultural change on both the contractor's and buyer's side. Third, raising performance standards can be expensive and the increased cost is not necessarily offset by increased performance (Fuhs, 1998). Fourth, procurement regulations—particularly in the E.U.—make it hard to incorporate contractor reputation into the selection process for transparency and competition concerns.[2]

2.3 Contracting for Equity and Contract Performance

Contracting for equity refers to the process of creating the environment for disadvantaged businesses to participate in government procurement. A small business set-aside, for example,

[2] Nonetheless, Tukiainen et al. (2021) find that the majority of the European public procurement officials surveyed prefer avoiding bidders with poor past performance as they are likely to incur delays and unexpected extra costs.

provides a preference for small businesses in the procurement and contracting process. Set-asides are based on public tenders exclusively reserved for small businesses, which limits competition. Traditional market models of contracting assume that contractors will deliver higher quality and more cost-effective services when exposed to competition since market forces will ensure desirable contractor selection and performance. Following this logic, transferring production to a less competitive market may yield limited economic benefits (Johnston et al., 2004).

We expect that contracting in a less competitive environment, such as set-asides, will result in a loss in efficiency. When there are a limited number of providers there is a greater risk of collusive schemes among participants (Stigler, 1964). Alternatively, contractors may focus on maximizing the probability of winning by organizing politically with government agencies (Johnston et al., 2004). Previous studies have provided supporting evidence for these conjectures. For instance, using timber auction data, Athey et al. (2013) find that small business set-asides produce a decline in revenue and efficiency for the Forest Service. However, Denes (1997) finds no evidence that set-asides for small businesses increase the cost of government contracts based on U.S. Army Corps data 1990-1991. Based on longitudinal U.S. federal contract data, Cappelletti and Giuffrida (2022) find that set-asides have a negative impact on contract-level outcomes.[3] Therefore, we hypothesize that:

[3] Although economic theories suggest more competition brings better outcomes in equilibrium, empirical studies find mixed results because perceived market competition also influences bidding (Fischbacher et al., 2009). Using US Army Corps data from 1990-1991, Denes (1997) finds no evidence that set-asides for small businesses increase the cost of government contracts. Some small firms that would have not considered bidding might enter the procurement auctions because of set aside programs (i.e., perceived higher chances of winning without large firms), which makes set-asides more competitive than non-set asides (Cappelletti and Giuffrida, 2022). Nakabayashi (2013) provides evidence for a positive impact of small business set-aside programs on the Japanese public construction market. Eliminating set-aside programs would lead to a 40% decline in small business participation. Nakabayashi (2013) also estimates the detrimental impacts from lack of competition would be higher than the efficiency loss caused by small businesses winning contracts. Similarly, Tkachenko et al. (2019) find that set-aside auctions have more efficient outcomes for procurement in terms of lower award prices for homogeneous goods.

Hypothesis 1. *Set-aside contracts have weaker performance than conventional contracts.*

2.4. An agency's flexibility in contracting and contract performance

Public management scholars often argue that management of the contracting process is crucial for governments to capture value from outsourcing. Federal agencies differ in contract management capacity. Therefore, any performance test needs to incorporate such a heterogeneity in contract management capability (Lamothe and Lamothe, 2010).

Organizations differ in their capability to exercise flexibility in contracting. Flexibility in contracting refers to the organization's ability to allow flexibility in selection. This means that the agency assesses tradeoffs the cost and non-cost actors, with the intent of awarding to the contractor that provide the agency with optimal value (Sade, 2005, p. 34). Organizations have different practices for information integration and sharing and different levels of management support, which are key to the flexible management of contracting (Selviaridis and Wynstra, 2014). There are also many impediments to allowing flexibility in contracting such as a significant change in culture required by the contracting agency, lack of training, and inability to achieve sufficient competition (Transportation Research Board, 2009). The Environmental Protection Agency that reported 16.6 percent of contracts were based on flexible arrangements (EPA, 2003), while the General Services Administration had a goal of 50 percent performance-based contract awards in 2005 (Boykin, 2005).

We expect that a federal agency's flexibility in contracting will positively influence contract performance. Many scholars have reported a positive relationship between performance-based contracting and desired outcomes. For example, Koning and De Meerendonk (2014) find an improvement in service quality following an increase in the scoring-weight given to a contractor in the scoring rule of public procurements of work-to-welfare programs. Similarly, Decarolis et al. (2016) also find a significant increase in quality following the announcement of the introduction of a past-performance based vendor rating system in a large utility company, rather than a corresponding increase in price. Butler et al. (2020) demonstrate that some mechanisms based on past performance can hinder supplier participation, but if appropriately designed can significantly increase both entry and quality of performance without increasing costs to the contracting agency. Based on a review of selected contracts worth about $1B, the EPA's inspector general office found that over $290M awarded in performance-based contracts could have been put to better use if the EPA had not granted award terms for less-than-superior service (EPA, 2008).

Therefore, contractors of a federal agency with stronger dynamic capabilities should have better performance, as measured by fewer contract amendments (i.e., cost overruns and delays):

Hypothesis 2. *A federal agency's flexibility in contracting positively affects contract performance.*

2.5. An agency’s flexibility in contracting, Contracting Environment, and Contract Performance

The effect of dynamic capabilities is context-specific. In competitive environments where rivalry is fierce, companies must innovate and explore new markets, find new ways to compete, and examine how they will differentiate themselves from competitors (Zahra, 1993, p. 324). Such environments increase the need for a firm’s capability to sense opportunities and seize the identified opportunities before competitors. A competitive environment requires dynamic capabilities. Even minor changes in its competitive environment demand dynamic capabilities if a firm is to survive (Tallman, 2015). Dynamic capabilities are therefore likely to be most valuable in highly competitive environments, where their benefits are likely to outweigh their costs (Makadok, 2001).

Scholars argue that a dynamic incentive mechanism based on past performance must complement standard competitive auctions to obtain decent value for money, especially when exchange terms are not contractible and there are many competing contractors (e.g., Calzolari and Spagnolo, 2009; Albano et al., 2017). As public agencies deal more often with openly competed contracts as required by regulations, it is more likely that the same agency needs to utilize its degree of bureaucratic discretion when navigating the procurement process (Nelson, 2017). Given the increased number of bids in open competition compared with restricted competition, the capability to select the small business that would be the best fit for entering into a contract is more important.

Indeed, precisely for this reason, we expect that a federal agency's flexibility in contracting is more beneficial when the agency deals with more contracts entered into through full and open competition. Once contracting officials set-aside contracts for small companies, excluding other potential contractors, the government can use competitive procedures to select a contractor. Regulations allow contracting officers to determine the extent and method by which the contract will be competed although full and open competition is typically encouraged (Kang and Miller, 2022). The utilization of competitive procedures differs by department/agency. For example, the Navy and Air Force reported the lowest competition rates among military departments, ranging between 36 and 45 percent between 2014 and 2019. The Defense Logistics Agency consistently reported the highest competition rates among the military departments, ranging from 71 to 84 percent (Duddy et al., 2020). Therefore, we formulate the following hypothesis:

Hypothesis 3. *The competitive environment moderates the relationship between a federal agency's dynamic capability and set-aside contract performance, such that dynamic capabilities have a stronger impact on performance in a more competitive environment.*

3. FEDERAL CONTRACTING FOR EQUITY: INSTITUTIONAL DETAILS

As noted above, set-aside programs for small businesses are designed to award a certain portion of federal contracts only to small businesses so that they do not have to compete with larger and more established competitors that have more resources and economies of scale (Blount et al., 2018). The small-business set-aside is by far the largest set-aside program in terms of dedicated federal budget. In 2022, recent statutory goals set by federal agencies are 23 percent of prime contracts annually for small businesses, exceeding $100 B.[4]

The goals of set-aside programs are both social and economic. Facilitating access is a way to create more equitable capitalism and the government needs to foster a level playing field for all businesses in the public procurement process (Snider et al., 2013). One small firm remarked that its view of set-asides was

> *Absolutely positive…it allows us to compete on a more level playing field. I think it's been a great program. You look at the numbers of small businesses in the United States, [and] you hear time and time again that so much of the income and GDP comes from small businesses.* (quoted in Lucyshyn and Rigilano, 2017, p.23).

Productive efficiency offered by contracting is another objective. Set-asides show the tradeoff between economic efficiency in the procurement of public goods and services and social goals (Schwartz, 2010). This requires a delicate balance.

The conflicting goals of set-asides can create a number of implementation challenges. Milton Friedman once said, "One of the great mistakes is to judge policies and programs by their intentions rather than their results." The government has established and implemented set-aside programs with the best of intentions. However, as with many policies, unintended consequences can occur. The government must strike a balance that helps small businesses while ensuring that contracting meets are met in an efficient way. Some argue that the current small business set-aside policy does not strike the optimal balance (Lucyshvn and Rigilano,

[4] For example, the SBIR program, a major set-aside program for R&D, reached $3.3B in 2019. See: https://www.sbir.gov/awards/annual-reports.

2017). Set-aside policy can create the potential for significant economic inefficiency. One mid-sized business executive offered the following remarks:

> *The government is always prone, when it hears about any inequities, to create more categories, more numbers, more demographic barriers, or segments. We continue to see the proliferation of size standards and demographic categories. At some point you have to ask, does the creation of these categories become counterproductive? By segmenting the industry space, do you force turbulence? Do you force unnecessary churn in the market?* (quoted in Lucyshyn and Rigilano, 2017, p.33).

Challenges in managing set-asides may add to the inefficiency in many ways. There has been increased reliance on multiple award contracts due partly to a lack of acquisition workforce, which means only a small group of small firms are favored.

> *But because Multiple Award Indefinite Delivery, Indefinite Quantity tend to have relatively long periods of performance, often up to five years, and few "on ramps," the contracts tend to limit participation. A government contract officer asserted that "SBSAs are giving small business work, but you have to be among the select few; there are some winners but there will be a lot of losers (*quoted in Lucyshyn and Rigilano, 2017, p.28*).*

In addition, small businesses bear a larger burden from regulation than large businesses (Crain and Crain, 2010). Employing a transaction costs framework, Dollinger et al. (1991) argue that minority firms often face higher transaction costs than their established counterparts. In 2014, Congresswoman Eshoo claimed that the "thousands of pages of procurement regulations discourage small innovative businesses from even attempting to navigate the rules" (Eshoo Press Release, 2014). One small firm expressed similar views:

> *Just putting together a proposal, getting a clearance, a facility clearance, so that you can bid on something that may require access to classified information. These are almost insurmountable hurdles for companies that are small businesses.* (quoted in Lucyshyn and Rigilano, 2017, p.31)

Such challenges can lead to negative outcomes. Cost overrun is one of the major issues in federal procurement. Meier (2010) notes the United States Department of Defense (DOD) cost overruns from 2000-2008 were over $200 million, up by more than 700 percent. GAO (2008) comments that from 2006-2008, DOD had overruns of $295 million and a delay of 21 months in systems delivery (Cable, 2016). Giuffrida and Rovigatti (2022) report that 59 percent of U.S. federal contracts that were completed between 2005 and 2015 experienced a delay, additional costs, or both. The government has attempted to provide more efficient services procured through private contractors, but then overwhelmingly the government turned to "cost-plus" or cost reimbursement type contracts, which provide less incentive to control costs and in fact reward overruns (Bajari and Tadelis, 2001; Berrios, 2006, p. 15).

Overall, the institutional details reveal the importance of managerial issues to improve contract performance.

4. EMPIRICAL METHODS

4.1. Data and Sample

We tested our hypotheses using data from the federal procurement data system (FPDS)—a publicly available database tracking U.S. federal procurement spending and available on www.usaspending.gov.[5] We extract data on all contracts executed between fiscal years 2008 and 2018. During this period, the federal acquisition regulation (FAR) required contracting agencies to report awarded contracts with an estimated value above $3,500[6] (including subsequent modifications) to the FPDS repository. The dataset provides a wealth of information including over 200 variables on contract, contractor, and buyer (federal agencies) characteristics. For example, contract-level data includes 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. Seller-level information includes the seller's identifier, the location of its headquarters, and weather it meets any small business standards. Buyer-level data has detailed information about the awarding agency, such as the awarding office within an agency.

[5] The FPDS has been used in previous studies in management (e.g., Calvo et al. 2019) and economics (e.g., Liebman and Mahoney, 2017).

[6] This was the cut-off amount for the period included in this study (FY2008-2018).

Our sample consists of service and construction contracts—jointly accounting for roughly half of the total yearly federal procurement outlay in our data. Because of their long duration before actual completion, service and construction contracts may be subject to renegotiations for cost overruns and delays during the execution phase. Renegotiation is commonly used in the literature to inversely proxy public contract performance (see below). Therefore, we exclude contracts where renegotiations are not meaningful to their outcomes, i.e., research and development, physical deliveries, and leasing.[7] Similarly, we disregard indefinite-delivery contracts as they provide for an indefinite quantity of goods and services over a specified period of time, so delays and additional costs cannot be interpreted as indicators of poor performance.

To select companies that are similar in terms of quality, we include only contracts that are competitively awarded as the participation criteria boost competition ex ante (e.g., Athey et al. 2011).[8] Moreover, we select contracts performed in the U.S., awarded through simplified acquisition[9], negotiated procedures, or sealed-bid auctions, and with a fixed-price format.[10] With respect to contract duration and size, we include contracts with an expected duration of more than two weeks *and* expected cost of more than $25,000; we do so to exclude contracts that are too small to be compared meaningfully. This ultimately resulted in a sample of 120,448 contracts (45 percent of which are set-asides) with a total value of $104 billion ($9.5 billion yearly on average), 466,954 bids submitted, and 51,123 unique winners in

[7] For R&D contracts, we would need a performance measure different from extra cost and delays. Previous studies that investigated R&D contracts have used the number and the quality of patents as outcome variables (e.g., Decarolis et al., 2021). For similar reasons, we exclude other types of contracts. For example, relevant outcomes for physical goods are unit prices, as in Best et al. (2017). However, the FPDS withholds this type of unit price information. For leases or rentals, contract renegotiations are not an indicator of poor outcomes.

[8] We consider a tender competitive program where the extent of competition is labelled in the FPDS as "full and open".

[9] Simplified acquisition procedures allow contracting officers more flexibility and less paperwork in supplier selection. It is mandatory for contracts with an estimated value below the simplified procurement threshold, i.e., $100,000 before 2011, $150,000 until 2017, and $250,000 thereafter.

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

the period of 2008-2018. Thus, set-asides are a fairly significant proportion of federal contracts, which underscores the importance of assessing tradeoffs between equity and efficiency.

4.2. Variables

Dependent variable. Our dependent variable is contract performance, which is measured as *Cost Overrun*, which is defined as the sum of extra cost renegotiated in addition to the awarded budget. We calculate it as follows (in $): *Cost Overrun = Final Cost - Award Amount*, where *Award Amount* refers to the project's expected budget (i.e., contract value) and *Final Cost* refers to actual costs. As a robustness check, we consider *Time Overrun* as an alternative outcome, which we compute as *Time Overrun = Final Duration - Expected Duration*. It is measured in days and is the difference between the actual and estimated completion dates. These metrics capture contract outcomes in public contracting, as renegotiations of fixed-price contracts for extra costs and time are considered suboptimal for all parties involved (Spiller, 2009; Bajari et al., 2014) and are commonly used as post-award performance measures in both project management (Herweg et al., 2018) and empirical economics (Decarolis 2014).[11] In this paper, to facilitate effect interpretation, we compute the ratios of the two measures to their respective benchmarks (i.e., award amount and expected duration) as shown below and use the ratios as our dependent variables.

$$Extra\ Cost = \frac{Cost\ Overrun}{Award\ Amount} \quad \text{and} \quad Delay = \frac{Time\ Overrun}{Expected\ Duration}.$$

Flexibility in contracting. We operationalize dynamic capability with a proxy for flexibility in contracting. Dynamic capabilities, as defined above, reside in organizational routines and processes. Managers use specific organizational processes to alter and reshape

[11] Spiller (2008) argues that given the formal and bureaucratic nature of public procurement, any renegotiation of a contract clause would lead to additional adjustment costs, reducing the incentive for both contractors and public authorities to adjust. Bajari et al. (2014) provide empirical evidence for this hypothesis and estimate the adjustment costs in their construction data to be between 8 and 14 percent of the winning bid.

their resource base (Eisenhardt and Martin, 2000). Given that a dynamic capability is the capacity of an organization to purposefully create, extend, and modify its resource base (Helfat et al., 2007), we consider a federal agency's routine to reshape its resource base by responding to past contractor performance as a source of dynamic capability. Put differently, systematic learning and updating strategic responses based on past performance of contractors is an instance of a dynamic capability. Dynamic capabilities are a learned process through which a company systematically generates and modifies its operating routines to improve efficiency (Zollo and Winter, 2002). Such a capability makes operations more responsive to changing environments.[12]

More specifically, we consider an agency's flexibility in contracting as a proxy for dynamic capabilities. We assume that an agency has strong dynamic capabilities when it alters its resource base by selecting contractors who provide resources based on their past performance with the agency. Thus, agencies with strong dynamic capabilities may respond to their contractors' past performance. To construct our measure of dynamic capabilities, we consider only contracts awarded through simplified acquisition procedures—i.e., small contracts—to better compare procurement activities across the different federal agencies.[13] We define "good" contractors as those associated with no poor contracts (measured in a dichotomous fashion as those without extra costs or delays), within an agency-year pair, and mixed contractors as those with some renegotiations, up to 60 percent of the contracts.

---

[12] One example for a performance-based contract is the Navy's T-45 aircraft system 3- ready for training. The Navy uses two metrics: the contractor needs to have a minimum number of aircraft ready for training at 7:00 AM each business day in order to achieve a 57 percent aircraft availability and "sortie completion," which requires that the contractor meet 98 percent of the requirements for the scheduled training flights. As an incentive, the contract pays a performance bonus (maximum of $5 million annually) if the contractor exceeds the performance metrics. If the contractor only meets-or fails to meet-the minimum metrics, the contractor then receives none of the annual performance bonus (GAO, 2004). Some performance-based contracts are in name only because enforcement mechanisms are lacking and payments continue to be disbursed when performance is poor. Put differently, performance-based contracts do not accurately reflect the extent to which performance-based strategies are actually applied. A DoD Inspector General report revealed that in 33 instances (out of 60 performance-based contracts), the DoD failed to clearly define criteria for successful completion of tasks but disbursed payments to the contractors (DoD IG, 2013; Lucyshyn and Rigilano, 2017).

[13] In addition to the small size, the contracts we employ to construct our proxy of dynamic capability are relatively short-term—i.e., a median expected duration is 172 days.

Consequently, we define "bad" contractors as those that renegotiate more than 60 percent of their contracts.[14] As a result, our main proxy of dynamic capability ($DC_{b,t}$) builds as follows: 1 – the share of bad firms at *t-1* selected again at *t* by the same agency b. [15] In this vein, an increase in $DC_{b,t}$ implies an increase in dynamic capability.

Contracting environment. We define an agency-level contracting environment, i.e., $CE_{b,t}$, in terms of competition level measured as the share of competed contracts awarded by an agency *b* via full and open competition at *t* (Giuffrida and Rovigatti, 2022; Decarolis et al., 2020). A higher share of competed contracts indicates that the market for that specific work or service is more competitive. Although full and open competition is generally encouraged, federal agencies differ in the utilization rate of such competition.[16]

Agency controls. We include several additional variables to control for various agency and time-based characteristics. Larger agencies might be more competent and therefore might deal with more complex contracts, which are mechanically more prone to renegotiation (Decarolis et al., 2020). Such dynamic, unobserved work complexity at the agency level might also (negatively) correlate with the employment of set-asides and create omitted variable bias. We use three different measures to capture agency size overtime. First, *Contracting Offices* is the number of distinct contracting units within the agency. This variable also controls for the complexity of the agency structure. In addition, more offices might indicate that the agency has an intense procurement activity. It could also signal that the agency is spread over the U.S. territory. Second, *Procurement Budget* is the agency's total

[14] For example, we label a company that has received three contracts as good if none or one of the contracts performed poorly and bad otherwise—i.e., if at least two contracts were renegotiated.
[15] We stress that no rule binds federal agencies to discard (select again) bad (good) contracts and the selection is either made by contracting officers or results from the competition process. Also note that there is no institutional definition of bad contract or poor contractor in procurement regulation. In Section 5.2, we show the robustness of our results to such definition.
[16] For example, the Consumer Product Safety Commission in 2012 awarded only 1.8% of its contracts through a competitive procedure, while the U.S. Trade and Development Agency in 2016 awarded over 93% of its contracts with the same procedure.

annual procurement budget including *all* purchase categories. Third, *Contracts* is the agency's total annual number of contracts awarded including *all* purchase categories. The latter measure is complementary to procurement budget, as an agency might have a small budget yet award several small contracts. We believe that these three measures together provide a good picture of an agency's contracting activity through different nuances. Finally, we include as an additional covariate *R&D Budget (%)*, which measures the annual cumulative R&D contract spending as a percentage of the total procurement budget. As discussed above, R&D contracts inherently involve more flexibility and a higher level of contracting skills. This may correlate with our measure of dynamic capabilities and thus we control for the level of R&D activity.

Summary Statistics. Table 2 presents the summary statistics of our variables. In both the treatment and the control groups, the contract amounts are highly skewed: half of contracts have an award price below $70,000 while the average award amount is approximately one order of magnitude higher. The average total cost including costs for subsequent modifications is over $0.5 mil. for set-asides and about $1.5 mil. for non-set-asides. The average expected contract duration is 230 and 288 days and the actual duration is 416 and 464, for set-asides and non-set-asides respectively. The significant cost and time increase is summarized by outcome distributions across treatment arms: the mean extra costs of 0.211 (i.e., extra costs totaling 21 percent of the award amount) and the mean delays of 0.876 (i.e., about 88 percent of the expected duration) for non-set aside contracts. The outcomes are poorer for set-aside contracts with extra costs of 0.242 and delays of 0.921. The descriptive statistics suggest that on average set-aside contracts are associated with poorer outcomes. Panel C shows the characteristics of the agencies in our sample. Each agency spends an average of $1.67 bil. per F.Y. on over 9,000 different projects.

INSERT TABLE 2 ABOUT HERE

Panels D to F of Table 2 present the correlation matrix of the variables used in our study. The results show that *Extra Cost* and *Delay*, our proxies for contract execution inefficiency, are correlated but not strongly. This suggests that they are different types of poor contract outcomes. $DC_{b,t}$ and $DC_{b,t}^{good}$ are almost perfectly correlated, which confirms the robustness of our measure. This means that agencies that avoid selecting bad firms consistently select good firms instead. In addition, the $DC_{b,t}$ variable is not strongly correlated with $CE_{b,t}$, which indicates that the latter measures are something different from a dynamic capability and can be used instrumentally without a confounding effect. With regard to the covariates, the correlation among them seems to be low enough to avoid any potential bias caused by multicollinearity. Although the proxies for agency size are highly correlated, the correlation with the capability variables is marginal, making multicollinearity unlikely.

4.3 Empirical Strategy

Consider the following model of contract performance, $Y_{i,b,t}$ associated with contract i awarded by agency b in year t.

$$Y_{i,b,t} = \beta_0 + \beta_1 SA_i + \beta_2 DC_{b,t} + \beta_3 SA_i * DC_{b,t} + \beta_4 SA_i * DC_{b,t} * CE_{b,t} + \beta_5 \boldsymbol{Z_{b,t}} + \beta_6 \boldsymbol{Z_i} + \beta_7 \boldsymbol{Z_{m,t}} + \epsilon_{i,b,t} \qquad (1)$$

In equation (1), SA_i is the treatment indicator, which takes the value one if the contract is awarded through a small-business set-aside program, and zero otherwise.[17] The coefficients of interest are $\beta_1, \beta_2, \beta_3$ and β_4 which indicate the degree of the effect of SBSA (direct or moderated) on contract outcomes, while β_5, β_6, and β_7 are coefficients of the control variables.

Identification strategy. To test our hypotheses, an ideal experiment would take a treatment or control group of contracts that are similar in all dimensions. Unfortunately, a natural

[17] Note that we consider only set-asides for small firms. 0 refers to contracts without competitive constraints as a counterfactual scenario and excludes other set-aside categories that are more restrictive in terms of participation criteria (e.g., women-owned small businesses) and account for a smaller share of total expenditures.

experiment does not allow us to control for all possible sources of endogeneity. Moreover, as shown in Table 2, treated and control contracts differ substantially in several observable characteristics. This suggests that selection into the treatment might bias the effect of the treatment on outcomes, if estimating equation (1) by OLS. Our preferred method, the inverse probability weighting (IPW) approach, is an appropriate identification method as it takes into account both the above-described selection into treatment and the valuation problem (Caliendo and Kopeinig, 2008). The latter stems from our interest in the counterfactual situation, i.e., how the treated contracts would have been fulfilled had they not been set-aside. Some studies on public procurement have already employed the IPW (e.g., Bruce et al., 2019; Decarolis et al., 2021). Our Online Appendix A provides more detailed explanations of the IPW method with additional information on the random forest, the selection of covariates, and the validity of the assumption of the IPW.

Our approach, the IPW, works as follows: In the first stage, we estimate the propensity score as the probability of assignment to the set-aside treatment as a function of the observed covariates. Hinging on the wealth of our data, this approach allows us to perform an unbiased counterfactual analysis by eliminating a large number of differences observable by both the econometrician and the buyer between the treatment and the control group. In the second-stage, we adjust for imbalances between the treatment and control groups by reweighting them through the propensity scores calculated in the first stage. We provide further details on both estimation stages in what follows.

Following Cappelletti and Giuffrida (2022), we estimated the propensity score $e(z)$ using random forest—a machine learning algorithm. Random forest takes multiple random independent samples from the data and then develops a decision tree for each sample based on how the data are classified within the data set (Breiman, 2001). In recent years, researchers have shown that machine-learning techniques can overcome some of the limitations of logistic or probabilistic regression in predicting propensity score.[18]

By estimating the propensity score, we approximate the implementation of the Rule of Two and the information set that the contracting officer considers in setting aside. The Rule of Two requires contracting officers to set aside any acquisition over the simplified acquisition threshold for small business participation when there is a reasonable expectation that (1) offers will be obtained from at least two responsible small business concerns and (2) the award will be made at fair market prices.[19] Given that a market (i.e., a set of potential bidders for a specific contract) is loosely defined in set-aside guidelines, we refrain from imposing one specific market definition. Instead, to account for market characteristics, we construct discrete measures of potential competition both by division (e.g., Pacific, Mountain, South Atlantic) and by sector. The latter is defined by the most granular product or service code (i.e., the four-digit code as reported in FPDS).

In the second stage, in order to ensure that set-asides are balanced across similar contracts in terms of observables, we weight them by the inverse propensity score. By doing so, we effectively compare contracts equally likely to be assigned to each treatment group based on observables. In particular, we estimate the average treatment effect on the treated (ATET) by weighing the outcome of treated contracts by one and that of control contracts by $e(z)/(1-e(z))$ and run Equation (1). One advantage of weighting using propensity scores rather than matching on covariates is that it reduces the curse of dimensionality and improves the precision of estimates (see Online Appendix A for further details on the random forest implementation in our setting).

[18] See Cappelletti and Giuffrida, 2022 for more details on the advantages of random forest compared to other machine learning techniques.
[19] See https://contractingacademy.gatech.edu/tag/rule-of-two/.

5. RESULTS

In Table 3, we present the second-stage IPW results relating to our hypotheses. Our dependent variable is contract performance measured by extra cost. Consistent with Cappelletti and Giuffrida (2022), Model (1) shows that set-aside has a negative and significant impact on cost overruns (s.e.= 0.008), which supports hypothesis 1. Setting aside a contract increases the extra cost by 2.6 percentage points (p.p.) (or 12 percent) on the award amount, compared to the counterfactual, i.e., had the contract not been set aside.

INSERT TABLE 3 ABOUT HERE

Hypothesis 2 proposed that a federal agency's dynamic capabilities improve contract performance. Model (2) confirms our hypothesis 2, showing a positive and significant coefficient of dynamic capabilities (-0.309, s.e.= 0.042). An increase of one standard deviation in $DC_{b,t}$ decrease the extra cost by 6.5 p.p., or by 29 percent. Instead, in Model (3), we find that dynamic capabilities' effect on cost overruns is not significant when interacted with set-aside. This zero effect suggests that high dynamic capability levels can compensate for the negative impact of set aside. Model (4) includes the linear interaction among dynamic capabilities, set-aside status, and contracting environment, which is negative and significant. When $CE_{b,t}$ increases by one standard deviation, an increase by one standard deviation in $DC_{b,t}$ decreases the extra cost by 6.5 p.p., or by 29 percent. This indicates that the influence of dynamic capabilities on the relationship between set-asides and contract performance hold positive and is stronger for agencies with a higher level of competitive contracts. Figure D1 in Online Appendix D presents a visual representation of the thee-way interaction term of Model (4) and report the predictive margins. We observe that, for set-aside contracts, the impact of $DC_{b,t}$ decreases regardless of $CE_{b,t}$ being low or high (red and blue line, respectively). However, the effect is lower and decreases more steeply when the agency environment is more competitive. This evidence corroborates hypothesis 3.

INSERT FIGURE 1 ABOUT HERE

5.1. Robustness Checks

There are several econometric issues that may arise based on our estimation approach. These concerns include (1) sensitivity of results to different outcome measures and (2) the empirical appropriateness of the dynamic capability variable.[20]

First, we broke down our aggregate measure of contract performance for each type of performance (extra cost, delays) to study the isolated effects of each performance. The variable Extra Cost is a direct measure of additional costs incurred by the government and, indirectly, taxpayers. Delays can induce indirect costs in terms of disruptions for government agencies—e.g., in the case of a delay of a professional service—or in the provision of a public good—e.g., for the delivery of a new road. The results, reported in Table 4, are qualitatively consistent with the original results obtained when employing cost overruns as a dependent variable, even after the direct effects of set-aside and dynamic capability lose statistical significance at the conventional levels in the triple-mediated model specification (Model 4).

INSERT TABLE 4 ABOUT HERE

Second, as reported in Table 5, we use different thresholds for defining a poor contract (i.e. a bad contractor). We report an additional set of results that employ $DC_{b,t}$ defining firm as bad if they renegotiate more than 50 percent or 70 percent of their contracts (instead of the 60 percent baseline cutoff) within an agency-year combination in Panel A and B, respectively. When comparing the robustness check with our baseline results (i.e., Panel A of Table 3), we find the difference between most coefficients is not statistically significant, showing that the estimates are not sensitive to changes in the measures of dynamic capabilities.

INSERT TABLE 5 ABOUT HERE

[20] For more robustness checks, see Online Appendix B.

In sum, our main results are robust to a host of alternative specifications and continue to support all our hypotheses.

6. DISCUSSION

The U.S. federal government spends more than $600 billion each year on contracts, roughly equivalent to the size of the Swedish economy. Governments leverage their purchasing power to advance social goals (McCrudden, 2004). Federal contracting for equity embodied by set-aside programs, although of substantial importance (more than $100B every year), is not enough studied and not fully understood yet. In this paper, we shed light on this topic. Conventional prescriptions from the contracting literature do not apply in a straightforward manner to public contracting. That is because federal agencies (our "buyers" in this context) have to deal with multiple stakeholder groups with conflicting objectives in a highly regulated market, which requires managerial capabilities to deal with strategic trade-offs. Put differently, federal contracting for equity is an excellent setting in which to examine the effects of dynamic capabilities.

Our theoretical model predicts that a federal agency's dynamic capability has a positive impact on contract performance. The model also predicts that the competitive environment moderates the relationship between a federal agency's dynamic capability and contract performance, such that dynamic capabilities have a stronger impact on performance in a more competitive environment. We tested these predictions using comprehensive longitudinal data on public procurement at U.S. federal agencies and a novel regression method.

We find empirical support for the above predictions. Generally, our results support the hypothesis that government agencies can pursue both equity and efficiency objectives by utilizing their market power through dynamic capabilities side-by-side, instead of at the expense of one another. We consider a federal agency's routine to reshape its resource base as a source of dynamic capabilities. This involves a learning process of managing inter-firm exchanges of productive resources and capabilities. Agencies can use experience gained and adapt incentives for contractors. Such a capability makes operations more responsive to changing environments.

More specifically, we find that set-asides are more negatively associated with contract performance than regular contracts. As hypothesized, a federal agency's dynamic capabilities have a direct, positive influence on contract performance. Such positive impact is stronger for agencies that use more open and full competition.

Note that several of our hypotheses were only weakly supported. For example, results relating to the main effect of dynamic capabilities are positive and significant (hypothesis 2), but the relationship does not hold in the context of set-asides. However, since being a set-aside contract has a negative effect on performance, a zero effect on performance when interacting with dynamic capabilities is a positive outcome on performance. A possible explanation is that this is due to the constraining environment of set-asides that limits competition. In such a more restricted environment, having dynamic capabilities may be less beneficial. This is consistent with Heaton et al. (2022), suggesting that the effect of a university's dynamic capabilities is negatively moderated by the level of governance in the state where the university operates.

6.1. Theoretical Implications[21]

Our results have three important theoretical implications. First, our findings enhance our understanding of value creation resulting from transactions between public and private organizations. While prior studies have identified the importance of contractual governance, especially in R&D contracts, this study highlights the role of managerial capabilities in implementing non-R&D contracts, which are often subject to uncertainties and complexities.

[21] For managerial implications, see Online Appendix C.

More specifically, this study clarifies existing research on the impact of a form of affirmative action (set-aside programs) aiming for equity and efficiency in performance. Our findings support the role of government capability in enhancing the linkage between set-asides and contract performance by empirically demonstrating that the agency's flexibility in contracting negatively moderates the relationship between set-asides and contract performance. Our results corroborate the findings of previous studies that highlight the role of government capability in contract performance (Bruce et al., 2019).

In addition, despite the explanatory potential of dynamic capabilities, the framework has received relatively little attention in the public management literature (Pablo et al., 2007; Piening, 2013; Cosimi, 2020). Yet, some studies hint that "the dynamic capability view fits well with the current volatile environment which many reformed public organizations experience" (Rosenberg Hansen and Ferlie, 2016, p. 12). Similarly, Pablo et al. (2007) also argue that dynamic capabilities are critical factors for the success of public organizations. If public organizations are not able to renew their resource base through dynamic capabilities, they find it difficult to respond promptly and effectively to divergent challenges in the changing environments that public organizations face today (Piening, 2013; Cosimi, 2020). We contribute to this line of the literature by providing empirical evidence supporting the argument. In addition, given that "the concept of dynamic capabilities is insufficiently underpinned by empirical data," (Easterby-Smith and Prieto, 2008, p. 237) our study advances the field and improves our understanding of boundary conditions under which dynamic capabilities are most effective. More specifically, we identified the moderating effect of the contracting environment (competition level measured as the proportion of competitive contracts) as an important boundary condition for the effect of dynamic capabilities.

We also make an important empirical contribution to the literature on dynamic capabilities. Although change is at the core of dynamic capabilities (Winter, 2003), a relatively limited number of studies have used longitudinal data to construct a measure of dynamic capabilities (Laaksonen and Peltoniemi, 2018). Studies often operationalize dynamic capabilities based on financial data such as R&D expenditure (e.g., Helfat, 1997) and marketing expenditure (e.g., Narasimhan et al., 2006) while others rely on managers' evaluation of their company's proficiency (e.g., Daneels, 2008). Unlike these studies, using longitudinal data, our operationalization of dynamic capabilities taps into organizational-specific paths in organizational learning and capability building over time. In addition, we contribute to the growing body of research applying machine learning in strategy research by employing a random forest technique (Tidhar and Eisenhardt, 2018; Choi et al., 2018; Raj and Seamans, 2019; Miric et al., 2022). Our machine learning approach allows for more-precise effect sizes and nonlinearities to better measure (Tidhar and Eisenhardt, 2020) and predict (Lee at al., 2018).

6.2. Study Limitations and Future Research

A limitation of this study is that it does not consider potential long-term effects of set-asides. In the long run, small businesses might win more contracts as they learn by doing. On the other hand, set-asides may discourage them from growing and developing new skills and encourage them to stay small to continue to benefit from the subsidies. Cappelletti and Giuffrida (2022) found no evidence of a positive scale effect of small businesses from veteran set-aside program is low: firms in the program do not grow differently as others with regular contracts. On the contrary, winners become increasingly dependent on set-aside tenders for their operations. Future research could use different performance measures such as set-aside graduation rates that capture longer-term effects.

Another limitation concerns the dynamic capability measure. We assume that a federal agency responds to a contractor's poor performance in year *t-1* in year *t*. However, it is possible that some agencies do not respond so quickly. We believe that this assumption is reasonable, given that our sample consists of relatively short-term contracts (the median is five and a half months). Future studies could expand the sample to include longer-term contracts and control for the speed of an agency's response.

A final limitation relates to our machine-learning approach. The many advantages of using the random forest for prediction come with the usual caveat that it is not easy to interpret estimates as it is not always clear what algorithm has been used to generate such estimates.

BIBLIOGRAPHY

Adida E, Bravo F (2019) Contracts for healthcare referral services: Coordination via outcome-based penalty contracts. *Management Science* 65(3):1322–1341.

Al-Matari, E.M. (2020) Do Characteristics of the Board of Directors and Top Executives Have an Effect on Corporate Performance among the Financial Sector? Evidence Using Stock. *Corporate Governance: The International Journal of Business in Society*, *20*, 16–43.

Albano, G. L., B. Cesi and A. Iozzi, 2017. "Public Procurement with Unverifiable Quality: The Case for Discretionary Competitive Procedures", *Journal of Public Economics*, 145, pp. 14–26.

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

Athey, S., Levin, J., and Seira, E. (2011). Comparing open and sealed bid auctions: Evidence from timber auctions. *The Quarterly Journal of Economics*, *126*(1), 207–257.

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

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

Behn, R. D., and Kant, P. A. (1999). Strategies for avoiding the pitfalls of performance contracting. *Public Productivity & Management Review*, 470–489.

Berrios, R. (2006). Government contracts and contractor behavior. *Journal of Business Ethics*, *63*(2), 119–130.

Best, M. C., Hjort, J., and Szakonyi, D. (2017). Individuals and organizations as sources of state effectiveness. NBER Working paper No. 23350.

Blount, I. Y., Seetharaman, J., and Brown, T. L. (2018). The impact of strategy on public sector set-aside programs: the case of the state of Ohio. *Journal of Public Procurement*.

Boykin, D. 2005. What performance-based contracting really means for procurement goals. American City & County. https://www.americancityandcounty.com/2005/04/18/what-performance-based-contracting-really-means-for-procurement-goals/

Brown, TL. and Potoski, M. (2003). Transaction Costs and Institutional Explanations for Government Service Production Decisions. *Journal of Public Administration Research and Theory* 13(4): 441–68.

Breiman, L. (2001). Random forests. *Machine Learning*, 45(1):5–32.
Bruce, J. R., de Figueiredo, J. M., and Silverman, B. S. (2019). Public contracting for private innovation: Government capabilities, decision rights, and performance outcomes. *Strategic Management Journal*, 40(4), 533–555.
Brudney, J. L., Fernandez, S., Ryu, J. E., and Wright, D. S. (2005). Exploring and explaining contracting out: Patterns among the American states. *Journal of Public Administration Research and Theory*, *15*(3), 393–419.
Bryson, J. M., Berry, F. S., and Yang, K. (2010). The state of strategic management research. A selective literature review and set of future directions. *American Review of Public Administration*, 40(5), 495–521.
Bryson, J.M., Hamilton E. L. and Van Slyke, David M. (2018) Getting strategic about strategic planning research, Public Management Review, 20:3, 317–339.
Butler, J. V., Carbone, E., Conzo, P., and Spagnolo, G. (2020). Past performance and entry in procurement: An experimental investigation. *Journal of Economic Behavior & Organization*, *173*, 179–195.
Cable, C. L. (2016). *The relationship between US Federal Governments cost estimates, service contract costs, and cost overruns*. Doctoral dissertation, University of Phoenix.
Cabral, S., Mahoney, J. T., McGahan, A. M., and Potoski, M. (2019). Value creation and value appropriation in public and nonprofit organizations. *Strategic Management Journal*, 40(4), 465–475.
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.
Cappelletti, M. and Giuffrida, L. M. (2022). Targeted Bidders in Government Tenders. ZEW – Centre for European Economic Research Discussion Paper No. 22–030, Available at SSRN: https://ssrn.com/abstract=4168912 or http://dx.doi.org/10.2139/ssrn.4168912
Choi, J., Menon, A., and Tabakovic, H. (2021). Using machine learning to revisit the diversification–performance relationship. *Strategic Management Journal*, 42(9), 1632– 1661. https://doi.org/10.1002/smj.3317
Cosimi, L. (2020). A behavioral micro foundation of public administration: do cognitive capabilities and social capital impact on dynamic capabilities in local governments? Business administration. Université Côte d'Azur; Università degli studi (Pise, Italie).
Crain, W. M., and Crain, N. V. (2014). The cost of federal regulation to the U.S. economy, manufacturing small business. Washington, DC: National Association of Manufacturers.
Danneels, E. (2008). Organizational antecedents of second-order competences. *Strategic Management Journal*, **29**, pp. 519–543.
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., Spagnolo, G., and Pacini, R. (2016). *Past performance and procurement outcomes* (No. w22814). National Bureau of Economic Research.

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.
Decarolis, F., Rassenfosse, G. d., Giuffrida, L. M., Iossa, E., Mollisi, V., Raiteri, E., and Spagnolo, G. (2021). Buyers' role in innovation procurement: Evidence from U.S. military R&D contracts. *Journal of Economics and Management Strategy*. 30, 697– 720. https://doi.org/10.1111/jems.12430
Denes, T. A. (1997). Do small business set-asides increase the cost of government contracting?. *Public Administration Review*, 441–444.
Dollinger, M. J. and Daily, CM. (1991). Purchasing from Small Minority-Owned Firms: Corporate Problems (Center for Advanced Purchasing Studies/National Association of Purchasing Management, Tempe, Arizona).
Duddy, B., Landucci, T. and Knechtel, J. (2020). 'Influencing Defense Acquisition Delivering Capabilities Faster'. *Defense Acquisition Research Journal*, **27**, 436–74.
Easterby-Smith, M., and Prieto, I. M. (2008). Dynamic capabilities and knowledge management: an integrative role for learning?. *British journal of management*, *19*(3), 235–249.
Eisenhardt, K. M., and Martin, J. A. (2000). Dynamic capabilities: what are they?. *Strategic management journal*, *21*(10-11), 1105–1121.
EPA. 2018. Report: EPA Can Achieve Better Value by Complying With Performance-Based Contract Requirements https://www.epa.gov/office-inspector-general/report-epa-can-achieve-better-value-complying-performance-based-contract
EPA, 2003. EPA could increase savings and improve quality through greater use of performance-based service contracts https://www.epa.gov/sites/default/files/2015-12/documents/pbsc_final_audit_report20030331.pdf
Eshoo Press Release. (2014). Reps. Eshoo & Connolly unveil draft federal IT procurement reform bill. Retrieved from http://eshoo.house.gov/press-releases/reps-eshoo-connolly-unveil-draft-federal-it-procurement-reform-bill/.
Fischbacher, U., Fong, C. M., and Fehr, E. (2009). Fairness, errors, and the power of competition. *Journal of Economic Behavior & Organization*, *72*, 527–545.
Fuhs, J. W. (1998). *How the implementation of performance based contracting has affected program management within the Department of Defense*, NAVAL POSTGRADUATE SCHOOL MONTEREY CA.
Gansler, J. S., Lucyshyn, W., and Vorhis, C. (2011). *Performance-based services acquisition*. Acquisition Research Program.
Gervais, S., and Odean, T. (2001). Learning to be overconfident. *The review of financial studies*, *14*(1), 1–27.
Giuffrida, L. M., and Rovigatti, G. (2022). Supplier selection and contract enforcement: Evidence from performance bonding. *Journal of Economics & Management Strategy*, 1– 40. https://doi.org/10.1111/jems.12492
Gordon, M. D., Morris, J. C. and Steinfeld, J. (2019). 'Deepwater or Troubled Water? Principal–Agent Theory and Performance-Based Contracting in the Coast Guard's Deepwater Modernization Program'. *International Journal of Public Administration*, **42**, 298–309.

Guy, M. E., and Ely, T. L. (2022). *Essentials of public service: An introduction to contemporary public administration*. Melvin & Leigh, Publishers.
Handley SM, Angst C (2015). The impact of culture on the relationship between governance and opportunism in outsourcing relationships. *Strategic Management Journal* 36(9):1412–1434.
Heaton, S., Teece, D., and Agronin, E. (2022). Dynamic capabilities and governance: An empirical investigation of financial performance of the higher education sector. *Strategic Management Journal*, *44*(2), 520–548.
Hefetz, A., and Warner, M. (2004). Privatization and its reverse: Explaining the dynamics of the government contracting process. *Journal of public administration research and theory*, *14*(2), 171–190.
Helfat, C.E. (1997). Know-how and asset complementarity and dynamic capability accumulation: the case of R&D. *Strategic Management Journal*, **18**, pp. 339–360.
Heinrich, C. J., and Choi, Y. (2007). Performance-based contracting in social welfare programs. *The American Review of Public Administration*, *37*(4), 409–435.
Herweg, F. and Schwarz, M. A. (2018). Optimal cost overruns: Procurement auctions with Renegotiation. *International Economic Review*, 59(4):1995–2021.
Ingraham, P. W., Joyce, P. K. and Donahue, A. K. (2003) Government Performance: Why Management Matters, Baltimore, MD: Johns Hopkins University Press.
Imbens, G. W. and Rubin, D. B. (2015). Causal inference in statistics, social, and biomedical sciences. Cambridge University Press.
Jensen, M. C. (1993). 'The Modern Industrial Revolution, Exit, and the Failure of Internal Control Systems'. *The Journal of Finance*, **48**, 831–80.
Johnston, J. M., Romzek, B. S., and Wood, C. H. (2004). The challenges of contracting and accountability cross the federal system: From ambulances to space shuttles. *Publius: The Journal of Federalism*, *34*(3), 155–182.
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**, 1495–556.
Klein-Woolthuis, R. K. W., Hildebrand, R. and Nooteboom, B. (2005) Trust, contract and relationship development, Organization Studies. 26(6): 813–840.
Koning, P., and Van De Meerendonk, A. (2014). The impact of scoring weights on price and quality outcomes: An application to the procurement of Welfare-to-Work contracts. *European Economic Review*, *71*, 1–14
Laaksonen, O., and Peltoniemi, M. 2018. The essence of dynamic capabilities and their measurement. *International Journal of Management Reviews*. 20: 184–205.
Lamothe, M. and Lamothe, S. (2010). 'Competing for What? Linking competition to performance in social service contracting. *The American Review of Public Administration*, **40**, 326–50.
Lee, C., Menon, A., and Tabakovic, H. (2018). Using machine learning to predict high-impact general technologies [Working paper]. Ithaca, NY: Cornell University.
Lerner, J. (1999). The government as venture capitalist: The long-run impact of the SBIR program. *Journal of Business*, *72*(3), 285–318.
Li, J., Li, P., Wang, B. (2019). The liability of opaqueness: State ownership and the likelihood of deal completion in international acquisitions by Chinese firms. *Strategic Management*

Journal, 40 (2): 303-327
Lichtenberg, F. R. (1988). The Private R and D Investment Response to Federal Design and Technical Competitions. *The American Economic Review*, **78**, 550–59. Available at http://www.jstor.org/stable/1809152.
Liebman J.B., Mahoney N. (2017) Do expiring budgets lead to wasteful year-end spending? evidence from federal procurement. *American Economic Review* 107(11):3510–49.
Link, A. N. and Scott, J. T. (2009). ‘Private Investor Participation and Commercialization Rates for Government-sponsored Research and Development: Would a Prediction Market Improve the Performance of the SBIR Programme?’. *Economica*, **76**, 264–81.
Lucyshyn, W. and Rigilano, J. (2017). *Evaluating the Impact of Small Business Set-Asides on Acquisition Efficiency*, Acquisition Research Program.
Luo, J. and Kaul, A. (2019). Private action in public interest: The comparative governance of social issues. *Strategic Management Journal*, 40(4): 476-502.
Makadok, R. (2001). Toward a synthesis of the resource-based and dynamic-capability views of rent creation. *Strategic management journal*, *22*(5), 387–401.
Mayer, K. J., and Argyres, N. S. (2004). Learning to contract: Evidence from the personal computer industry. *Organization science*, *15*(4), 394–410.
McCrudden, C. (2004). Using public procurement to achieve social outcomes. In *Natural resources forum* (Vol. 28, No. 4, pp. 257–267). Oxford, UK: Blackwell Publishing Ltd..
McCue, C. P., Prier, E. and Swanson, D. (2015). ‘Five dilemmas in public procurement’. *Journal of Public Procurement*, **15**, 177–207.
Meier, S. R. (2010). Causal inferences on the cost overruns and schedule delays of large-scale U.S. federal defense and intelligence acquisition programs. Project Management Journal, 41(1), 28-39.
Miric, M., Jia, N., and Huang, K. G. (2022). Using supervised machine learning for large-scale classification in management research: The case for identifying artificial intelligence patents. *Strategic Management Journal*, 44(2), 491– 519. https://doi.org/10.1002/smj.344100
Narasimhan, O., Rajiv, S. and Dutta, S. (2006). Absorptive capacity in high-technology markets: the competitive advantage of the haves. *Marketing Science*, **25**, pp. 510– 524.
Nelson, G. L. (2017). *Rational choice institutionalism and federal contracting officers’ discretionary authority for allocating construction contracts among federally designated disadvantaged and socioeconomic small business concerns: an exploratory research design*. Doctoral dissertation, Washington State University
Pablo, A., Reay, T., Dewald, J.R. and Casebeer, A.L. (2007). Identifying, enabling and managing dynamic capabilities in the public sector. *Journal of Management Studies*, 44: 687–708.
Patrucco, A. S., Luzzini, D., Ronchi, S., Essig, M., Amann, M. and Glas, A. H. (2017). Designing a public procurement strategy: lessons from local governments. *Public Money & Management*, **37**, 269–76.
Piening, E. P. (2013). Dynamic capabilities in public organizations: A literature review and research agenda. *Public management review*, *15*(2), 209–245.
Poppo, L., and Zenger, T. (2002). Do formal contracts and relational governance function as ´substitutes or complements?. *Strategic management journal*, *23*(8), 707–725. https://doi.org/10.1002/smj.249

Potter, R. A. (2022). ‘How much of rulemaking is done by contractors?’. *Brookings*, 16 February. Available at https://www.brookings.edu/research/how-much-of-rulemaking-is-done-by-contractors/ (accessed 3 November 2022).
Quelin, B.V., Cabral, S., Kivleniece, I., and Lazzarini, S.G. (2019). The private scope in public-private collaborations: An institutional and capability-based perspective. *Organization Science*. 30(4): 831–846.
Raj, M. and Seamans, R. 2019. Primer on artificial intelligence and robotics. *Journal of Organization Design* 8(1): 1–14.
Rassenfosse, G. de, Jaffe, A. and Raiteri, E. (2019). ‘The procurement of innovation by the U.S. government’. *PloS one*, **14**, e0218927.
Rosenberg Hansen, J., and Ferlie, E. (2016). Applying strategic management theories in public sector organizations: Developing a typology. *Public Management Review*, *18*(1), 1–19.
Schoemaker, P. J., Heaton, S., and Teece, D. (2018). Innovation, dynamic capabilities, and leadership. *California management review*, *61*(1), 15-42.
Schwartz, J. I. (2010). Procurement as a Tool for Economic Stimulus: The United States’ Experience, 2008-2010. Paper Presented at the International Public Procurement Conference, August, Seoul, South Korea.
Selviaridis, K. and Wynstra, F. 2015. Performance-based contracting: A literature review and future research directions. International Journal of Production Research. 53(12): 3505-3540
Siegel, D. S., and Wessner, C. (2012). Universities and the success of entrepreneurial ventures:
Evidence from the small business innovation research program. *The Journal of Technology Transfer*, *37*, 404–415.
Snider, K. F., Kidalov, M. V., and Rendon, R. G. (2013). Diversity governance by convenience?
Federal contracting for minority-owned. *Public Administration Quarterly*, *37*(3), 393–432.
Spiller, P. T. (2009). An institutional theory of public contracts: Regulatory implications. In *Regulation, deregulation, reregulation*. Edward Elgar Publishing.
Stigler, G.J. (1964). A theory of oligopoly.” *Journal of Political Economy* 72.1: 44–61.
Tallman, S. (2015). ‘Dynamic capabilities’. Wiley Encyclopedia of Management, 1–3.
Teece, D., Pisano, G., Shuen, A. (1997). Dynamic capabilities and strategic management. *Strategic Management Journal*, 18(7): 509-533.
Teece, D., (2007), Explicating dynamic capabilities: The nature and microfoundations of (sustainable) enterprise performance. *Strategic Management Journal*, 28(13): 1319-1350.
Tidhar, R. and Eisenhardt, K. (2020). Get rich or die trying…finding revenue model fit using machine learning and multiple cases. Strategic Management Journal. 41(7): 1245–1273.
Transportation Research Board of National Academics (2009) Performance-based contracting for maintenance. Washington, D.C.: National Cooperative Highway Research Program
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* 21-033, ZEW Discussion Paper Series.
Vanneste, B. S., Puranam, P. (2010). Repeated interactions and contractual detail: Identifying the learning effect. Organization Science. 21(1): 186–201.

Winter, S. G. (2003). Understanding dynamic capabilities. *Strategic Management Journal*, **24**, 991– 995.
Wooldridge, J. M. (2010). *Econometric analysis of cross section and panel data*. MIT press.
Zahra, S. A. (1993). Environment, corporate entrepreneurship, and financial performance: A taxonomic approach. *Journal of business venturing*, *8*(4), 319–340.
Zahra, S. A. and George, G. (2002). Absorptive capacity: A review, reconceptualization, and extension. *Academy of management review*, *27(2)*, 185-203.
Zheng, W. Ni, N., Crilly, D. (2019). Non-profit organizations as a nexus between government and business: evidence from Chinese charities. *Strategic Management Journal*. 40(4)
Zollo, M. and Winter, S. (2002). Deliberate learning and the evolution of dynamic capabilities. *Organization Science*, 13(3): 339–351.

TABLE 1 Full Dataset**:** Contract Type (Observations=8,253,682)

Contract type	**Proportions of total contracts**	**Small business set-asides**
Products	43.6%	17.2%
Research and Development	5.7%	30.8%
Services	20.4%	24.3%
Construction and Maintenance	30.3%	32.8%

Note: Column 2 refers to the proportion of spending over all U.S. federal contracts between 2008 and 2018, while Column 3 is the percentage of spending awarded to set-asides. We aggregate contracts according to their sector code. The types of contracts included are products, R&D, services/construction, and maintenance. Set-asides in R&D include SBIR (Small Business Innovation Research) contracts.

TABLE 2 Summary Statistics

Panel A: Contract-level, Set-aside (Treatment Group)	**(1) N**	**(2) mean**	**(3) median**	**(4) sd**	**(5) min**	**(6) max**
Award Amount ($,000)	54,158	462.1	67.30	2,662	25	213,691
Total Cost ($,000)	53,698	570.4	77	3,273	7.81e-08	277,267
Exp. Duration (days)	54,158	229.8	152	251.0	15	2,558
Total Duration (days)	54,144	415.5	233	497.2	1	3,921
Extra Cost	53,698	0.242	0	0.731	-0.337	4.482
Delay	54,144	0.921	0	1.911	-0.381	11.98
# Bids	54,158	4.297	3	9.404	1	999

Panel B: Contract-level, Non-set-aside (Control Group)	**(1) N**	**(2) mean**	**(3) median**	**(4) sd**	**(5) min**	**(6) max**
Award Amount ($,000)	66,290	1,189	73.17	11,032	25	1.157e+06
Total Cost ($,000)	65,974	1,530	80.51	15,281	0	1.718e+06
Exp. Duration (days)	66,290	287.5	222	334.6	15	3,895
Total Duration (days)	66,280	463.9	343	527.8	1	3,983
Extra Cost	65,974	0.211	0	0.708	-0.337	4.482
Delay	66,280	0.876	0	2.020	-0.381	11.98
# Bids	66,290	3.654	2	11.447	1	999

Panel C: Agency-Year Level	**(1) N**	**(2) mean**	**(3) median**	**(4) sd**	**(5) min**	**(6) max**
DC	417	0.821	0.857	0.205	0	1
DC^good	417	0.779	0.808	0.222	0	1
CE	533	0.508	0.500	0.254	0	1
Proc. Budget ($ B)	542	1.667	0.033	8.421	.00003	90.23
# Contr. Offices	542	56.51	3	136.8	1	840
# Contracts (000)	542	9.320	0.214	36.152	0.001	301.335
R&D Budget (%)	542	0.039	0.005	0.093	0	0.952

Panel D: Contract-level, Set-aside (Treatment Group) – Matrix of Correlations	**(1)**	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**	**(7)**
(1) Award Amount ($,000)	1.000						
(2) Total Cost ($,000)	0.939	1.000					
(3) Exp. Duration (days)	0.135	0.147	1.000				
(4) Total Duration (days)	0.166	0.201	0.604	1.000			
(5) Extra Cost (%)	-0.005	0.078	0.256	0.547	1.000		
(6) Delay (%)	0.061	0.087	-0.061	0.567	0.340	1.000	
(7) # Bids	0.043	0.042	0.037	0.053	0.016	0.025	1.000

Panel E: Contract-level, Non-set-aside (Control Group) – Matrix of Correlations	**(1)**	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**	**(7)**

(1) Award Amount ($,000)	1.000						
(2) Total Cost ($,000)	0.894	1.000					
(3) Exp. Duration (days)	0.126	0.124	1.000				
(4) Total Duration (days)	0.140	0.147	0.658	1.000			
(5) Extra Cost (%)	0.003	0.062	0.143	0.462	1.000		
(6) Delay (%)	0.018	0.029	-0.117	0.473	0.303	1.000	
(7) # Bids	0.023	0.020	-0.007	0.003	-0.002	-0.004	1.000

Panel F: Agency-year level – Matrix of correlations

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
(1) DC	1.000						
(2) DC^good	0.962	1.000					
(3) CE	0.136	0.126	1.000				
(4) Proc. Budget ($ B)	-0.068	-0.152	0.019	1.000			
(5) # Contr. Offices	-0.112	-0.233	0.029	0.706	1.000		
(6) # Contracts (000)	-0.078	-0.179	0.021	0.887	0.861	1.000	
(7) R&D Budget (%)	-0.030	-0.059	-0.037	0.221	0.080	0.164	1.000

Note: This table contains summary statistic at the contract level. We report the statistics for small business set-aside in Panel A, and for non set-aside contracts in Panel B. Dollar values are in thousands. Award amount is the sum of all federal obligation. Total Cost is the sum of all modifications related to each contract. Exp. Duration and Total Duration report the expected and actual duration of the contracts, in days. Delay measures the ratio of delay (total duration minus expected duration) relative to the expected duration. Extra Cost indicates the share of the cost overrun (total minus expected cost) relative to the expected cost. # of Bids counts the number of offers received for a given contract. Panel C reports statistics at the agency-year level. DC and DC^good are proxies for dynamic capabilities, CE for competitive environment. Proc. Budget, in billions, reports the total procurement budget by agency and year. # Contr. Offices corresponds to the unique number of contracting offices within an agency. # Contracts, in thousands, reports the total number of contracts awarded by an agency. R&D Budget (%) reports the percentage of procurement spending that is awarded to R&D contracts. We report the matrix of correlations for the same variables as Panel A, B and C in panel D, E, F respectively. We report them for small business set-aside (SA), the treatment group, in Panel D, and for non set-aside contracts, the control group, in Panel E. Panel F reports summary statistics at the agency-year level.

TABLE 3 Baseline Results

	(1) Extra Cost	(2) Extra Cost	(3) Extra Cost	(4) Extra Cost
SA	0.026	0.027	0.044	0.092
	(0.008)	(0.008)	(0.080)	(0.082)
DC		-0.309	-0.298	-0.294
		(0.042)	(0.077)	(0.077)
SA*DC			-0.023	0.031
			(0.098)	(0.097)
SA*DC*CE				-0.223
				(0.046)
Observations	90,095	85,209	85,209	85,209

Note: Results for the average treatment effect on the treated of the inverse probability weighting Extra Cost. The treated are contracts awarded with restricted solicitations, i.e., contracts that are set aside (SA). DC is a proxy for dynamic capabilities. We construct DC by looking at the percentage of “bad” contractors that are selected again by the same agency in the following period. We report in the first row the coefficient estimates, standard errors are in parentheses CE is a measure for the competitive environment.

TABLE 4 Robustness Check: Alternative Outcome

	(1) Delay	(2) Delay	(3) Delay	(4) Delay
SA	0.056	0.056	0.014	0.127
	(0.023)	(0.024)	(0.214)	(0.211)
DC		-0.030	-0.058	-0.050
		(0.141)	(0.232)	(0.232)
SA*DC			0.054	0.178
			(0.267)	(0.275)
SA*DC*CE				-0.514
				(0.133)
Observations	90,711	85,805	85,805	85,805

Note: Results for the average treatment effect on the treated of the inverse probability weighting on Delay. The treated are contracts awarded with restricted solicitations, i.e., contracts that are set aside (SA). DC is a proxy for dynamic capabilities. We construct DC by looking at the percentage of "bad" contractors that are selected again by the same agency in the following period. We report in the first row the coefficient estimates, standard errors are in parentheses CE is a measure for the competitive environment.

TABLE 5 Robustness Check: Share of Renegotiated Contract

Panel A: Dynamic Capability, Bad: Share > 50%	**(1)** Extra Cost	**(2)** Extra Cost	**(3)** Extra Cost	**(4)** Extra Cost
SA	0.026	0.027	0.043	0.072
	(0.008)	(0.008)	(0.063)	(0.064)
DC		-0.252	-0.239	-0.238
		(0.033)	(0.064)	(0.064)
SA*DC			-0.024	0.053
			(0.084)	(0.083)
SA*DC*CE				-0.227
				(0.050)

Panel B: Dynamic Capability, Bad: Share > 70%	**(1)** Extra Cost	**(2)** Extra Cost	**(3)** Extra Cost	**(4)** Extra Cost
SA	0.026	0.026	0.065	0.129
	(0.008)	(0.008)	(0.085)	(0.088)
DC		-0.317	-0.291	-0.287
		(0.046)	(0.079)	(0.079)
SA*DC			-0.049	-0.012
			(0.103)	(0.102)
SA*DC*CE				-0.224
				(0.045)
Observations	90,095	85,209	85,209	85,209

Note: Results for the average treatment effect on the treated of the inverse probability weighting on Extra Cost. The treated are contracts awarded with restricted solicitations, i.e., contracts that are set aside. DC is a proxy for dynamic capabilities. We construct it by looking at the percentage of “bad” firms that are selected again by the same agency in the following period. The two DC measures are constructed by defining "bad" a firm with a share of 50% and 70% renegotiated contracts, which are reported in Panel A and B respectively. We report the coefficient estimates in the first row, standard errors are in parentheses. CE is a measure for the competitive environment

ONLINE APPENDIX

APPENDIX A: Details on the First Stage - Random Forest and Propensity Scores

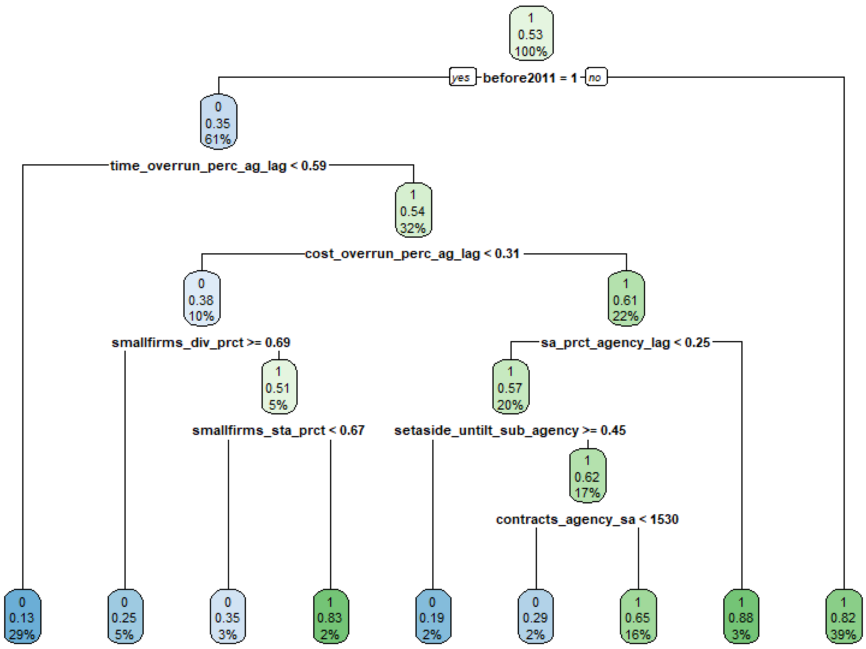
This appendix provides further arguments for using the random forest to predict propensity scores. Following Cappelletti and Giuffrida (2022), we leverage a random forest to predict propensity scores of contracts being set aside as a first stage. Developed by Breiman (2001), the random forest is a machine learning method that aims at predicting an outcome based on the available covariates. In this work, we use the random forest to predict whether a contract is likely to be set aside based on buyer, contract, and market characteristics.

Random forests are built on decision trees. A single decision tree consists of a series of binary (yes/no) questions that lead to a class prediction for each observation—in this case, whether the contract belongs to the treatment or control group. Classification trees are used when the outcome variable is binary, while regression trees are employed 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 that receives a binary treatment, the random forest will combine multiple classification trees.

Important features of the trees are nodes and branches, the meaning of which is best illustrated by Figure A1. Starting from the green ellipse at the top of the figure, "Before2011 = 1" is the first node or split. At this point, 100 percent of the sample is at this node, and 53 percent of the sample has received the treatment. If the statement is true, i.e., if the contract is awarded prior to 2011, i.e. "Before2011 = 1", then we move to the left branch. We then reach the second node, which includes 61 percent of the sample, 35 percent of which receiving the treatment. At this node, we can further split the sample by posing another question, namely whether the selected variable takes a value below 0.59. We continue in a similar manner until we reach a final node at the bottom of the decision tree. If the variable for a particular contract takes a value below 0.59, the decision tree predicts that the contract will not be set aside.

Although decision trees are a useful method for prediction and classification, they can often lead to overfitting. Random forests solve this problem because they work as follows: Multiple trees are grown, each time using a different bootstrap sample of the data. Then a majority decision is made to predict the outcome of each observation. By selecting a random subset of features at each split (or node), random forests add an extra level of randomness compared to decision trees. . In the latter, the variable selected at each node is the best among all possible variables in the data. Instead, in a random forest, only a subset of variables is selected at random at each node, and the variables selected are the best among a subset of variables.

FIGURE A1 Example of a classification tree



Note: The decision tree was generated using the R package "rpart" and predicts the likelihood of receiving treatment for each observation, namely whether the contract is set aside or not. The sample and variables used to build this decision tree are identical to those used in the random forest model in the main analysis. However, note that this decision tree serves as a mere illustrative example of how decision trees look like. Consequently, the decision tree shown here may not necessarily be indicative of the trees generated through a random forest model.

Breiman and Cutler (2011) advise to grow between 1,000 to 5,000 trees if the number of variables is high and if the researcher is interested in the stable importance of the variables. In our application, we decide to grow 1,000 trees to have a stable significance of the variables since we are interested in the determinants of treatment, but also to keep the computational cost low.

For tuning the optimal number of randomly selected variables at each node, we again follow Breiman and Cutler (2011). The authors recommend experimenting with varying numbers of randomly selected variables in combination with a relatively small number of trees. An exact number of trees is not given. We, therefore, use 200 trees. We train the data with a sample that is 80 percent of the original dataset and test the error prediction on the remaining 20 percent aside. We choose three different numbers of variables to run three trials. First, we calculate the 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. As we obtain the lowest prediction error for 11 variables, we settle on this value.

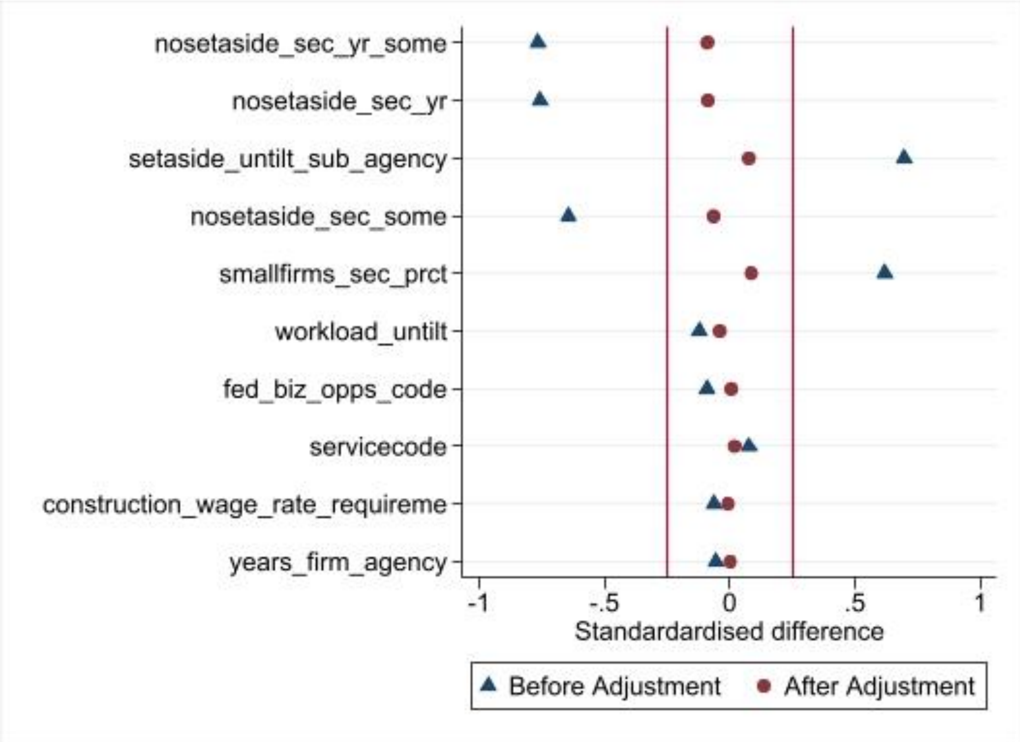
Selection of Covariates - According to Caliendo and Kopeinig (2008), treatment predictors are key to an unbiased estimate of the propensity score. From the FPDS, we select variables that (i) we expect to correlate simultaneously with treatment and outcomes of interest, (ii) are measured before treatment, and (iii) are orthogonal to any anticipation of treatment. Although the use too many variables can increase the variance of the estimators, even if the coefficients remain unbiased and consistent (Bryson et al., 2002), Rubin and Thomas (1996) argue against removing variables for the sake of parsimony. Instead, they should be included whenever the econometrician believes they correlate with the covariates and outcomes. This argument underpins the chosen random forest approach. After selecting the variables that meet the criteria of unbiasedness, we let the random forest use the most relevant variables for propensity score prediction, rather than making a variable selection ourselves (see below). We select 76 variables for propensity score prediction. The full list of variables can be found in the Online Appendix.

The IPW- Following the literature on the potential outcomes framework (e.g., Rubin, 1974), we consider the binary variable whose value depends on multiple predictors. Under the condition of covariates, the propensity score then describes the probability of each subject being assigned to the treatment given the observed covariates K_i. For causal comparisons, we adopt the standard assumption of a unit's stable treatment value (Rubin 1980), which states that each unit's potential outcomes are unaffected by the treatment assignments of other units and that each unit has potential outcomes—i.e., $Y_i(SA),\ SA = z = \{0;\ 1\}$, corresponding to the possible treatment levels, of which only one is observed: $Y_i = SA * Y_i(1) + (1 - SA) * Y_i(0)$. Under the unconfoundedness assumption—i.e, $Y(SA), Y(1 - SA) \perp SA|K_i$, we have $\Pr(Y(SA)|K_i) = \Pr(Y|K_i,\ SA = z)$ for z: 0,1 so $\tau(x)$ is the average treatment effect (ATE) conditional on k: $\tau(k) = E[Y(1) - Y(0)|K_i = k]$. Estimation of either comparison requires the probabilistic assignment assumption, $0 < e(X) < 1$, which states that the study population is restricted to values of covariates for which there can be both control and treated units.

Testing of Assumptions - To test whether the overlap (or common-support) assumption is satisfied, we examine the distributions of propensity scores for both treated and untreated overlap. According to the assumption, every unit has a probability greater than zero of being allocated to either the treatment or the control group. We exclude from the relevant population those units whose probability of receiving the treatment can be perfectly (or nearly perfectly) predicted (Wooldridge, 2010). As a result, we limit our sample to include only those units whose propensity score falls within the range of 0.01 to 0.99, meaning that we exclude 9.5 percent of observation. Second, it is important to check the balance of the covariates after the prediction of the propensity score. Ho et al. (2007) claim that the propensity score was appropriately specified when the covariates are balanced, which means that the treatment effect estimates can be valid and unbiased (Zhao et al., 2016). We test whether the overlap assumption holds by examining the covariate balance between treated and control contracts. It should be noted that the overlap assumption is frequently made in observational studies. This assumption implies that any individual in the sample could receive any level of treatment and that we cannot perfectly predict the probability of receiving treatment.

To this end, we can compute the normalized difference statistic (also called standardized differences) after applying IPW (Imbens and Rubin, 2015). This procedure allows us to assess the comparability of treated and control units in the weighted sample (Austin, 2009). In addition, the normalized difference statistic is more robust than the simple calculation of the t-statistic or the test of mean difference as it does not depend directly on sample size (Wooldridge, 2010). If the normalized differences do not exceed 0.25 (absolute value), the covariate balance 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 percent are above 0.1, and none are above 0.15. Figure A2 reports the standardized differences before and after weighting for the ten most important variables, as selected by the random forest. For example, we report the standardized difference for the proportion of small firms that also receive orders without set-aside in a given sector and year. For this variable, the standardized difference moves from -0.796 before adjustment for IPW to -0.083 after adjustment. This result is extremely promising because, as Garrido et al. (2014) suggest, the balance of critical covariates is even more relevant than the balance of other independent variables.

FIGURE A2 Standardized differences for selected covariates



Note: This figure displays the standardized differences both pre- and post- inverse probability weighted regression adjustment. In this sample, the treatment group consists of set-aside contracts, while the control group includes non set-aside contracts. The propensity score is estimated using 76 variables, but for simplicity, we present the ten most important variables as determined by the random forest's estimate of variable importance. The red vertical lines depict the 0.25 threshold, which indicates that covariate balance is attained when the standardized differences for all variables fall below this threshold (Imbens and Rubin, 2015). For the variables used, all covariates exhibit standardized differences below the absolute value of 0.25 after adjustment.

APPENDIX B: Additional Robustness Checks

In addition to the robustness checks presented in the paper, we consider the measure of dynamic capabilities. Since renegotiation entails transaction costs, as argued above, it is rational for agencies not only to screen out bad contractors from new procurements but also to select again good contractors. We measure this by $DC_{b,t}^{good}$, which is defined as the share of good firms (i.e., those who never renegotiate) at *t-1* that are selected again at *t*. When using $DC_{b,t}^{good}$, a federal agency's responsiveness to past good performance of the contractor, as an alternative measure for dynamic capabilities, the results are qualitatively and quantitatively almost identical, as shown in Table B1. Note that the effect is not mechanical, as not selecting bad contractors does not automatically mean selecting good contractors, nor searching for new contractors.

TABLE B1 Robustness Check: Alternative Dynamic Capability Measurement

	(1) **Extra Cost**	**(2)** **Extra Cost**	**(3)** **Extra Cost**	**(4)** **Extra Cost**
SA	0.026	0.027	0.037	0.047
	(0.008)	(0.008)	(0.055)	(0.056)
DC^good		-0.280	-0.272	-0.274
		(0.030)	(0.062)	(0.062)
SA*DC^good			-0.017	0.083
			(0.081)	(0.082)
SA*DC^good*CE				-0.218
				(0.053)
Observations	90,095	85,209	85,209	85,209

Note: Results for the average treatment effect on the treated of the inverse probability weighting on Extra Cost. The treated are contracts awarded with restricted solicitations, i.e., contracts that are set aside (SA). DC^good is an altenative proxy for dynamic capabilities. We construct DC^good by looking at the percentage of "good" contractors that are selected again by the same agency in the following period. We report in the first row the coefficient estimates, standard errors are in parentheses CE is a measure for the competitive environment.

Results also remain virtually unchanged when we employ an alternative way to assess a contractor's performance when building $DC_{b,t}$ that uses more than a one-year lag of a contractor's performance. Given that our panel of firms-agency-year is unbalanced, our preferred specification relies on a one-lag definition of dynamic capabilities. However, if a contractor has a longer history of good performance and only a single year of poor performance, a federal agency with high dynamic capability might still not classify the contractor as "bad" and find a new contractor for subsequent procurement processes by relying on a more informative history of good information. Therefore, as an additional check, we use a longer lag of a contractor's performance to better capture a longer firm-agency relationship. To capture such a possibility, we re-construct our dynamic capability variables in the following ways. We identify an agency selecting a bad contractor again for two consecutive periods of being a poor contractor instead of one period.[22]

TABLE B2 Robustness Check: t-1 Equals t-2

	(1) Extra Cost	(2) Extra Cost	(3) Extra Cost	(4) Extra Cost
SA	0.026	0.025	0.073	0.119
	(0.008)	(0.009)	(0.148)	(0.148)
DC		-0.552	-0.525	-0.515
		(0.075)	(0.136)	(0.136)
SA*DC			-0.052	-0.028
			(0.159)	(0.159)
SA*DC*CE				-0.142
				(0.037)
Observations	90,095	79,608	79,608	79,608

Note: Results for the average treatment effect on the treated of the inverse probability weighting Extra Cost. The treated are contracts awarded with restricted solicitations, i.e., contracts that are set aside (SA). DC is a proxy for dynamic capabilities. We construct DC^bad by looking at the percentage of "bad" firms that are selected again by the same agency after two consecutive periods of being bad. We report the coefficient estimates in the first row, standard errors are in parentheses. CE is a measure for the competitive environment.

As shown in Table B2, the results are again consistent with our main results. Similarly, we consider an agency to have a weak DC if it selects a poor contractor that performed badly at *t-2*, ignoring what happened at *t-1*. This check addresses the concern that some of the contracts

[22] Note that $DC_{b,t}^{bad}$ is then equal to 1- share of bad firms at *t-1* selected again at *t* by the same agency *b*.

we used to build our $DC_{b,t}$ measure have lasted longer than one year. Because of their length, they might not yet have been completed in the following year, making it difficult for the agency to react to the performance of an ongoing contract. Table B3 reports the results, which remain largely in line with our initial results.

TABLE B3 Robustness Check: Using t-2 and Ignoring t-1

	(1) Extra Cost	**(2) Extra Cost**	**(3) Extra Cost**	**(4) Extra Cost**
SA	0.026	0.026	-0.023	0.007
	(0.008)	(0.009)	(0.086)	(0.086)
DC		-0.357	-0.389	-0.385
		(0.054)	(0.095)	(0.095)
SA*DC			0.062	0.111
			(0.105)	(0.107)
SA*DC*CE				-0.167
				(0.044)
Observations	90,095	79,608	79,608	79,608

Note: Results for the average treatment effect on the treated of the inverse probability weighting on the two main outcomes: Extra Cost and Delay. The treated are contracts awarded with restricted solicitations, i.e., contracts that are set aside. DC is a proxy for dynamic capabilities. We construct DC by looking at the percentage of "bad" firms that are selected again by the same agency two periods after the previous contract. We report the coefficient estimates in the first row, standard errors are in parentheses. CE is a measure for the competitive environment.

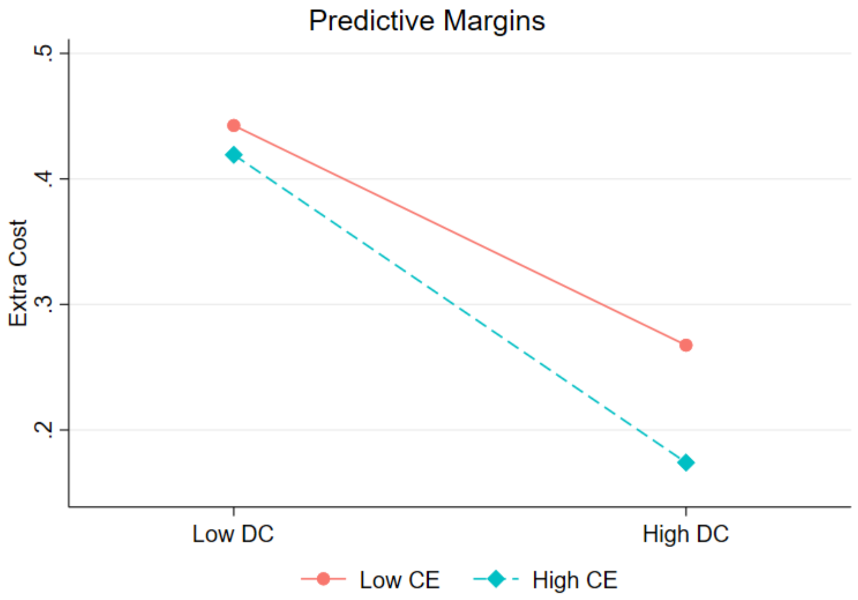
APPENDIX C: Managerial Implications

Policy makers should consider public procurement as an opportunity to incorporate equity considerations and likewise to achieve efficiency in the public interest. This requires new coordination in policy development and implementation. New coordination must reverse inflexibility around processes and rules in contracting. A myriad of acquisition rules can lead to undesirable outcomes such as vendor lock problems. Federal agencies can utilize their dynamic capabilities and respond to inefficiencies by employing tools such as the termination for convenience clause when the project is going in the wrong direction. More generally, new coordination may involve analyzing the agency's resource base, extending the agency's pool of resources, quickly learning how to implement changes to its standard procurement procedures, and choosing procurement strategies based on its current deficiencies and the nature of contractors. Such contingencies can facilitate evaluation and eventual prescriptions on procurement, and enable governments to respond faster to public needs.

Although we investigated the effects of an affirmative action program in the public sector, our study may have implications for the management of affirmative action programs such as diversity management policy in the private sector. Our research implies that such programs can lead to achieving intended objectives, such as providing a level playing field for under-represented groups while simultaneously hiring qualified candidates for their positions through an organization’s capability to manage such programs. For example, companies can adopt more flexible benefit programs designed to satisfy the different needs of diverse applicants. Creating an environment for fair and healthy competition among minority applicants can result in improvement in the performance of candidates and organizations.

APPENDIX D: Additional Figures

FIGURE D1 Three-way Interaction Plot



Note: The figure shows a visual representation for the three-way interaction “SA*DC*CE” on *Extra Cost*, as reported in Column (4) of Table 3. We estimate the interaction for set-aside contracts for high (above median) and low (below median) levels of CE, at low (0.2) and high (0.8) level of DC. We report the predictive margins.

ONLINE APPENDIX BIBLIOGRAPHY

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.
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.
Caliendo, M. and Kopeinig, S. (2008). Some practical guidance for the implementation of propensity score matching. *Journal of Economic Surveys*, 22(1):31–72.
Cappelletti, M. and Giuffrida, L. M. (2022). Targeted Bidders in Government Tenders. ZEW–Centre for European Economic Research Discussion Paper No. 22-030, Available at SSRN: https://ssrn.com/abstract=4168912 or http://dx.doi.org/10.2139/ssrn.4168912
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.
Ho, D. E., Imai, K., King, G., & Stuart, E. A. (2007). Matching as nonparametric preprocessing for reducing model dependence in parametric causal inference. *Political analysis*, *15*(3), 199-236.
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.
Rubin, D. (1974). Estimating causal effects of treatments in randomized and nonrandomized studies. Journal of Educational Psychology, 66(5), 688–701.
Rubin, D. (1980). Estimating causal effects of treatments in randomized and nonrandomized studies. *Biometrics*, 36(2), 293–298.
Rubin, D. B. and Thomas, N. (1996). Matching using estimated propensity scores: relating theory to practice. *Biometrics*, 249–264.
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.
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.

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