Essays in Empirical Industrial Organization and Public Economics

Inauguraldissertation

zur Erlangung des Grades eines Doktors
der Wirtschaftswissenschaften
der Universität Mannheim

vorgelegt von

Matilde Cappelletti

Frühjahrs-/Sommersemester 2025

Abteilungssprecher:	Prof. Thomas Tröger
Referent:	Prof. Michelle Sovinsky
Koreferent:	Prof. Nicolas Ziebarth
Vorsitzende der Disputation:	Prof. Henrik Orzen
Tag der Disputation:	25. August 2025

Acknowledgements

When I was in high school, I spent a year abroad. While I was deciding where to go, a wise person told me, “The country does not matter, it is all about the people you meet.” That insight has shaped how I have navigated my time in Mannheim. With this in mind, I want to thank everyone who supported, guided, and shared my PhD journey.

First and foremost, I am incredibly grateful to my advisor, Michelle Sovinsky, for recognizing and highlighting my strengths. Thank you, Michelle, for believing in my single-authored paper when it was just a single-page proposal. This project would not have seen the light without you. I have learned so much about research (and how to talk about research) from you. I am also extremely grateful to my advisor, Nicolas Ziebarth, for always being available for comments and honest opinions on my work and research ideas. Both my advisors’ courses, feedback, guidance, and encouragement were crucial to my doctoral journey and to helping me become a better researcher.

I wrote this dissertation while working as a researcher at the ZEW – Leibniz Centre for European Economic Research in Mannheim. I am incredibly grateful to all the current and past colleagues at the Department of Corporate Taxation and Public Finance for their helpful feedback, as well as for shared coffees, lunches, and conferences. I want to especially thank my fellow Junior Research Group “Public Procurement” members, Leonardo Giuffrida and Adriano De Leverano. Both of you have motivated me and taught me a great deal about research. Thank you, Adri, for being a great office mate. I have really enjoyed our conversations. Thank you, Leo, for being an incredible mentor since the last semester of my master’s degree. You have been a constant source of learning and motivation.

This dissertation is also the result of insightful and stimulating collaborations with my coauthors, Leonardo and Gabriele Rovigatti. Thank you for being amazing coauthors.

At the university and the ZEW, I have greatly benefited from the advice and feedback of many researchers. In particular, I want to thank Friedrich Heinemann, Zareh Asatryan, Laura Grigolon, Kevin Remmy, Effrosyni Adamopoulou, and Helena Perrone. Your comments and advice have been instrumental in this journey. An important thank you goes to everyone I have met at seminars and conferences who took the time to help improve my research. In particular, I would like to thank Lucy Xialou Wang. Thank you, Lucy, for being an excellent mentor and for your thoughtful feedback on my research.

A sincere thank you to everyone who has provided administrative, library, and IT support, both at the university and at the ZEW. In particular, thank you to Leili Erfanian, Martina Hamann, and Marion Lehnert for your help and patience.

I also want to thank my fellow PhD students at the CDSE and the ZEW. Over the years, I have enjoyed lunches, coffees, and conversations with many of you. Thank you all for making my time in Mannheim special. I have enjoyed the great PhD Meetings at Michelle's chair: thanks to Andrés Plúas-López, Jakob von Ditfurth, and Alejandro Medina for the valuable comments and the great fun, even during the stressful times. Also, a special thanks to Cäcilia Lipowski for sharing many research conversations and bike rides, and to Chiara Malavasi for the helpful hand and all the adventures, from Hamburg to Mannheim.

To all my friends back home who have been by my side since our school days: thank you for sticking by me, even from afar, even as we grow and change. A special thank you to Viviana Roggero for all the milestones we have celebrated together since elementary school.

I am lucky to have met many wonderful people throughout my university journey, which began in 2013 in Bolzano-Bozen. My fellow unibz friends, Rosa Alma Chizzini and Eleonora Sfrappini, were not only fantastic study partners during our bachelor's degree but also two cornerstones of my PhD journey, even from afar. Thank you, Rosa, for following all the PhD discussions. Hiking a 5,416-meter Himalayan mountain pass together is not the hardest thing I have accomplished with you by my side (physically or figuratively), and that speaks volumes about how important your friendship is. Thank you, Ele, for always answering my calls and being there through all the highs and lows of both life and the PhD. I have always felt so heard and understood. I have loved sharing this journey with you, and I look forward to sharing more of it in the future, as friends first and coauthors second.

My time pursuing a Master's degree in Mannheim also brought many great friendships that made the journey more enjoyable. Thank you, Franziska Krug, Kathrin Schmitt, and Mona Köhler, for the (almost) yearly hiking holiday over the years. A special thank you to Franzi for being the best roommate I could have asked for and my first *Heim* in Mannheim.

A heartfelt thank you goes to Ursula Berresheim and Marina Hoch. From the master's degree to the study groups during the first year of the PhD, you have become treasured friends. This PhD journey would not have been the same without your friendship: The long conversations after work, your readiness to listen, and much more.

None of this would have been possible without the unconditional love and support of my family. To my sister, for being proud of me, even without fully knowing what I do. To my niece, for distracting me from my adult worries with joyful games and carefree runs. To my parents, for letting me go, yet always keeping me close to their hearts. To my dad, for reading my papers. To my mom, for always reminding me of the hurdles I have conquered in the past, even when I did not believe I could.

To my partner, Stefan Hartmann, for everything. Our shared PhD journey has felt like the *Giro d'Italia*, where each rider must conquer their own climbs, but even the strongest riders cannot do their best without the support of their team. Stefan, you are a source of strength, inspiration, optimism, fun, and unconditional love. You have a rare ability to combine rigor with lightness. Every climb (and every view) is more beautiful when riding with you, every descent feels safer when I can follow your line, every flat route is faster when I can ride in your *Windschatten*. Thank you for cheering me on, for sharing the road with me, and for riding through the challenges of life by my side.

Contents

Preface

Governments play a central role in shaping both social and economic outcomes through regulation and public spending. In particular, government actions can affect outcomes by altering the incentives of economic actors. In this dissertation, I analyze the strategic responses of households and firms to these incentives. To identify causal relationships, I combine empirical reduced-form methods with micro-level administrative and survey data from the United States and Italy. This dissertation comprises three self-contained chapters that empirically examine research questions at the intersection of industrial organization and public economics.

Chapter 1, titled *Uncovering the Role of Moral Hazard in Health Insurance*, studies whether there is evidence of ex-ante moral hazard in health insurance—i.e., whether lower out-of-pocket costs lead individuals to engage more in risky behaviors. Due to aging populations that increase healthcare demands and costs, governments face the challenge of balancing patients' access to care with the financial sustainability of the healthcare system. Therefore, understanding the role of ex-ante moral hazard is important, given that it can lead to higher healthcare spending through the development of comorbidities and preventable health issues.

To provide causal evidence, I exploit the staggered implementation of state-level policies in the U.S. that lower the out-of-pocket costs for insulin. In line with the scope of the policies, I focus on privately insured households that have diagnosed diabetes. I use household-level grocery purchase data combined with a survey that provides health-related information. The results show that reduced insulin out-of-pocket costs lead to a sustained increase in purchases of sugar. The latter is a nutrient that is closely linked to insulin needs and can contribute to the development of long-term health risks. I document that the results are driven by those groups that have more scope to engage in ex-ante moral hazard—e.g., those with a diabetes type related to lifestyle management and those who report giving little importance to a low-sugar diet. Additionally, I find an increase in sales of diabetes devices. Combined, these patterns indicate a shift from lifestyle management to treatment that is consistent with an ex-ante moral hazard interpretation of the results.

Chapter 2, titled *Targeted Bidders in Government Tenders*, is joint work with Leonardo Giuffrida. This chapter examines the impact of affirmative action programs in U.S. federal procurement. In particular, we focus on set-aside programs, which limit competition to small businesses or to small and disadvantaged businesses. These programs aim to support the growth of smaller firms and promote greater equity in the procurement process. While policymakers generally favor full and open competition to maximize value for taxpayers, such competition can disproportionately benefit large firms, leaving smaller businesses at a disadvantage and limiting their opportunities for growth.

Combining machine-learning-based propensity scores with inverse probability weighting, we estimate the impact of these programs on both competition and contract efficiency. Using over a decade of contract-level procurement data, we find that while set-aside programs increase bidder participation, they also result in higher extra costs and delays. We provide suggestive evidence that the observed outcomes stem from different mechanisms depending on the type of set-aside: adverse selection appears to be at play in small business set-asides, where less experienced firms are more likely to win contracts, while moral hazard plays a larger role in small and disadvantaged business set-asides, with firms exerting lower post-award effort. To evaluate whether long-term improvements in firm performance offset these short-term inefficiencies, we exploit a large, unexpected increase in program spending. Our findings reveal mixed evidence: although firm growth following a win is modest, winning firms tend to become increasingly dependent on set-aside opportunities. Overall, the results suggest that short-term inefficiencies are not clearly compensated by long-term gains in firm growth, at least when considering the procurement market.

Chapter 3, titled *Procuring Survival*, is joint work with Leonardo Giuffrida and Gabriele Rovigatti. In this chapter, we examine the impact of public procurement contracts on firm survival. Public procurement can help relax firms' demand constraints, and it may do so differently than market-based private demand. While both public and private demand generate revenues, they can differ in profitability, which is an important predictor of firm survival. Public contracts may impose higher administrative burdens, yet they may also improve firms' access to credit.

We construct a novel dataset for Italy by integrating multiple sources: balance sheet and income statement data for the universe of limited liability companies, administrative records on firm entry and exit, and the quasi-universe of public procurement contracts. We exploit close-call auctions in Italy's construction sector and implement a regression discontinuity design that compares firms narrowly winning a government contract to those that narrowly lose. Our results show that winning a contract significantly increases firm survival well beyond the median duration of the contract. This effect is not driven by a boost in earnings, but rather by a higher share of revenue from public customers. We find no evidence of productivity gains; instead, the results point to improved access to credit, particularly for financially constrained firms.

Chapter 1

Uncovering the Role of Moral Hazard in Health Insurance

*Matilde Cappelletti**

1.1 Introduction

Healthcare spending in the U.S. has risen steadily over recent decades, averaging 17.5% of GDP between 2016 and 2019 (CMS, 2023), with similar trends in other OECD countries (OECD, 2019). Due to aging populations, both care demands and costs are expected to rise even further. Therefore, policymakers face the challenge of balancing the financial sustainability of the healthcare system and access to care. Policies that reduce out-of-pocket costs for patients, such as the U.S. Inflation Reduction Act, aim to improve access and affordability. However, their effect on patient behavior is ambiguous. While such policies may lead to greater healthcare utilization (Brot-Goldberg et al., 2017), they may also increase incentives to engage in risky behaviors due to reduced financial exposure to medical costs (Ehrlich and Becker, 1972).

In this paper, I evaluate the effects of reducing out-of-pocket costs on households' healthcare utilization and risky behaviors, particularly diet choices. I focus primarily on risky behaviors as they can provide evidence of ex-ante moral hazard in health insurance. In the health economics literature, ex-ante moral hazard refers to the idea that insurance, by lowering the out-of-pocket price of care, may reduce incentives to maintain a healthy lifestyle (Ehrlich and Becker, 1972).[1] Understanding whether ex-ante moral hazard affects patient behavior is crucial, as it might lead to preventable health issues and, consequently, to higher healthcare spending. Causal evidence of the effect on changes in out-of-pocket costs on ex-

* Researcher's own analyses calculated (or derived) based in part on data from Nielsen Consumer LLC and marketing databases provided through the NielsenIQ Datasets at the Kilts Center for Marketing Data Center at the University of Chicago Booth School of Business. The conclusions drawn from the NielsenIQ data are those of the researcher and do not reflect the views of NielsenIQ. NielsenIQ is not responsible for, had no role in, and was not involved in analyzing and preparing the results reported herein.

[1] In the literature, ex-ante moral hazard also refers to preventive measures, such as doctor check-ups. In this paper, I focus mainly on lifestyle-related behaviors. Note that ex-post moral hazard refers to the notion that insurance coverage may lead to increased healthcare use (Pauly, 1968).

ante moral hazard is lacking, while the effect on healthcare use has been extensively studied (Einav and Finkelstein, 2018).

To provide causal evidence, I exploit a natural experiment provided by the staggered implementation of U.S. state-level policies. Between 2020 and 2022, 23 states reduced the out-of-pocket costs for insulin for insured individuals. The policies set a monthly cap on out-of-pocket costs, ranging from $25 to $100, with any excess covered by insurance. In a staggered difference-in-differences approach à la Callaway and Sant'Anna (2021), I compare the grocery purchasing behavior of households with diabetes and private health insurance in states with the cap to those in states without the cap. I find that lowering insulin costs leads to riskier nutritional behaviors, consistent with ex-ante moral hazard. Households increase their sugar purchases, particularly among those who have more control over their condition through lifestyle. The results also show a sustained increase in the sales of devices essential for daily diabetes management. Taken together, the findings highlight a shift from lifestyle management to treatment.

These results provide new evidence for understanding the role of ex-ante moral hazard in health insurance along three dimensions. First, by analyzing the nutritional content of purchased food—rather than relying on product categories—I reduce reliance on subjective definitions of risky behavior. Nutritional content allows me to precisely quantify changes in the purchase of health-relevant nutrients (e.g., sugar) and avoid product misclassification. Second, I focus on individuals with chronic conditions, such as diabetes, whose higher healthcare requirements may cause them to react differently to financial incentives than people in better health. Third, while the existing literature primarily examines the extensive margin of insurance (i.e., obtaining insurance coverage) in relation to risky behaviors, this paper focuses on the intensive margin. Empirical evidence at the intensive margin is limited due to the challenges of finding a plausibly exogenous identification strategy and observing risky behaviors, health status, and insurance plan information simultaneously. In this paper, I overcome these empirical challenges by exploiting recent policies that reduce insulin costs and constructing a dataset with the necessary information.

I use a rich dataset constructed by combining several data sources. The primary source is the 2019–2022 NielsenIQ Consumer Panel provided by the Kilts Center. This is a representative panel data on grocery purchases for over 83,000 unique households. I complement this dataset with extensive information on the nutritional content of the products purchased. To identify households potentially affected by the policy, I match the Consumer Panel with the NielsenIQ Annual Ailments, Health, and Wellness Survey, which provides health-related information, including diabetes diagnoses, diabetes types, and insurance coverage.

In line with the scope of the policy, I focus on individuals with diagnosed diabetes. This group provides a relevant setting for answering my research questions for three main reasons. First, diabetes affects 11.3% of the U.S. population and is expected to rise due to demographic trends (Ong et al., 2023). It accounts for 25% of U.S. healthcare spending (Hirsch, 2016; ADA, 2023) and is the seventh leading cause of death, mainly due to its association with multiple comorbidities (ADA, 2018). Second, diabetes management depends heavily on both prescription drugs and diet. Tracking carbohydrate intake, especially sugar, is essential for regulating blood glucose and determining insulin dosage, which increases with higher carbohydrate consumption (CDC, 2023). Third, the distinction between diabetes types provides important insight into ex-ante moral hazard. While type I diabetes is genetic and requires insulin use regardless of an individual's lifestyle, type II diabetes—the most prevalent form that affects 95% of individuals with diabetes—can often be managed through changes in diet and exercise. Consequently, individuals with type II diabetes have greater scope for behavioral responses to financial incentives, making them more susceptible to ex-ante moral hazard.

I find that the reduction in insulin out-of-pocket costs leads to a 4.8% increase in carbohydrate purchases in the first quarter following the policy implementation. While one possible alternative explanation is that households are correcting previously suboptimal consumption constrained by limited insulin access, the composition of carbohydrate purchases suggests otherwise. Specifically, I document a sustained increase in sugar purchases (9.3%), a carbohydrate subtype that can increase the long-term risk of comorbidities, such as cardiovascular and kidney disease (Holman et al., 2008). These patterns are consistent with ex-ante moral hazard, whereby reduced short-term financial costs weaken incentives for preventive behavior. Supporting this interpretation, the increase in sugar purchases is primarily driven by households with type II diabetes, a condition often manageable through lifestyle changes. Heterogeneity analysis further supports the ex-ante moral hazard interpretation, as households placing low importance on a low-carbohydrate diet exhibit the most significant increases in sugar purchases.

I examine the extent of ex-ante moral hazard by studying the effect of the policy on other dietary components that could lead to comorbidities. I observe short-term increases in nutrients, such as fat, cholesterol, and calorie purchases. The results further highlight the shift toward less healthy nutritional behavior. However, the results are temporary, unlike sugar, showing that households return to the pre-policy level in the second quarter. Hence, they suggest that ex-ante moral hazard is primarily focused on behaviors most closely related to the prescription drug affected by the cost change. Finally, I can exclude an alternative explanation for the observed increase in risky behaviors, namely that an income effect is at play rather than ex-ante moral hazard. I find no evidence of an income effect, as households do not increase their grocery expenditures.

Additional analyses show that the policy leads to an increase in medication utilization and sales of testing devices. For these analyses, due to limitations in Consumer Panel data, I utilize diabetes device sales data from over 30,000 stores in the NielsenIQ Retail Scanner Data. The findings are in line with the ex-ante moral hazard interpretation. In particular, they suggest that individuals rely more on medication and monitoring to manage the effects of their dietary choices. While this shift from diet-based to medication-based disease management may be effective in the short term, it raises concerns about long-term health risks, such as the development of diabetic kidney disease (Kitada, Kanasaki, and Koya, 2014; Oliveira et al., 2024). To assess whether this behavioral shift has adverse consequences, I examine ketone strip sales—a proxy for disease management—and find a decline following the policy implementation. This result suggests that, at least in the short term, sugar levels are well-managed despite changes in diet.

These results suggest that improving access to insulin might lead to behavioral distortions that need to be considered to improve long-term health outcomes. Indeed, reduced costs lead to lower incentives for a healthy lifestyle, as shown by the rise in sugar consumption. To mitigate the unintended consequence of ex-ante moral hazard and to fully achieve the health benefit potentially due to improved insulin access, complementary behavioral interventions may be necessary. For instance, providing nutritional counseling to groups that are identified as being prone to engaging in risky behaviors could be beneficial. These findings might also inform decisions in states planning to introduce the cap. They might also extend to other diseases requiring a combination of diet and medication for effective management, such as high cholesterol, which is typically managed with statins alongside lifestyle adjustments. Moreover, these results might also be relevant for setting the out-of-pocket costs for new drugs, such as semaglutides (e.g., Wegovy and Ozempic).

Related Literature. This paper contributes to five strands of literature. First, I relate to papers examining whether insurance coverage increases risky behaviors due to ex-ante moral hazard.[2] Existing papers investigate the effect of insurance coverage (i.e., the extensive margin) on risky behaviors—such as alcohol and snack food consumption—and find mixed evidence (Dave and Kaestner, 2009; Simon, Soni, and Cawley, 2017; Cotti, Nesson, and Tefft, 2019; Chen et al., 2023). My contribution to this literature is twofold. First, I study the intensive margin, specifically how individuals who are already insured respond to a reduction in out-of-pocket costs. Empirical evidence at this margin remains limited, with notable exceptions including Brook et al. (1983) and Newhouse (1993). Second, to the best of my knowledge, I am the first to use nutritional content to measure risky behaviors in the moral hazard literature. This allows me to capture specific nutrients such as sugar, fiber, and saturated fat, and rely less on researcher-driven classifications (e.g., "snack food"). I am thus able to quantify the risky behavior more precisely and link it to disease mechanisms.

Second, I relate to the large literature investigating the impact of changes in out-of-pocket costs for insulin. Recent medical studies examine the effects of insulin out-of-pocket caps, finding consistent reductions in patient costs but mixed evidence on usage behavior (Anderson et al., 2024; Garabedian et al., 2024; Giannouchos, Ukert, and Buchmueller, 2024; Ukert, Giannouchos, and Buchmueller, 2024). Research from middle-income countries indicates that expanding coverage or subsidizing diabetes care can improve health outcomes (Americo and Rocha, 2020; McNamara and Serna, 2022; Barkley, 2023). Following a decrease (increase) in out-of-pocket costs for other diseases, the existing evidence mainly suggests an increase (decrease) in the use of prescription drugs (Gaynor, Li, and Vogt, 2007; Fiorio and Siciliani, 2010; Puig-Junoy, García-Gómez, and Casado-Marín, 2016) and healthcare (Chandra, Gruber, and McKnight, 2010; Ziebarth, 2010; Chandra, Gruber, and McKnight, 2014; Shigeoka, 2014). However, some scholars find no effect on prescription drug

[2] The ex-ante moral hazard literature also studies preventive care, e.g., Barros, Machado, and Sanz-de-Galdeano (2008) and Spenkuch (2012).

and healthcare consumption (Puig-Junoy, García-Gómez, and Casado-Marín, 2016; Park, Kim, and Choi, 2019; Martínez-Jiménez, García-Gómez, and Puig-Junoy, 2021).[3] Contributing to this literature, I provide new evidence on how reduced out-of-pocket costs affect not only medication use but also non-prescription, disease-related purchases, and dietary behavior, a key component of chronic disease management.

Third, I contribute to the literature linking food purchasing behavior to health outcomes. Prior work shows that lower prices for sugar-rich foods increase the risk of obesity and related conditions (Gračner, 2021) and that individuals with obesity are more responsive to food promotions (Bao et al., 2020). My paper is most closely related to Oster (2018) and Hut and Oster (2022), who find that households adjust their diets following a diabetes diagnosis. However, these changes are short-lived and limited to specific food categories. In this paper, I examine a policy setting that alters financial incentives—rather than health information—and demonstrate how reductions in prescription drug out-of-pocket costs affect dietary behavior. This offers new evidence on how insurance design can shape health-related consumption, particularly for individuals with diet-related chronic diseases.

A fourth relevant strand of research focuses on the factors influencing dietary choices. Prior work highlights the persistence of dietary habits over time (Hut, 2020), the importance of individual preferences beyond prices (Dubois, Griffith, and Nevo, 2014), and the role of information—individuals often lack accurate knowledge about nutrition or have no information at all (Belot, James, and Spiteri, 2020; Vitt et al., 2021). Through this work, I show that out-of-pocket costs for prescription drugs are also a factor that can influence dietary choices. This result is relevant for further understanding which factors influence diet, especially for individuals with nutrition-related diseases.

Fifth, I contribute to the literature on how government-led interventions affect unhealthy food consumption. One common intervention is a sin tax on sugary drinks, which aims to reduce sugar intake. Evidence shows mixed results: while such taxes reduce soda consumption on average (Dubois, Griffith, and O'Connell, 2020), individuals may compensate by substituting toward other sugary foods (Lozano-Rojas and Carlin, 2022) or by purchasing taxed products in nearby, untaxed areas (Cawley et al., 2019). Other interventions, such as food labeling, have been found to lower sugar and calorie consumption by improving awareness (Barahona, Otero, and Otero, 2023). While vouchers for healthy food also show improvements in diet, their effects may diminish over time (Griffith, Hinke, and Smith, 2018; Hinnosaar, 2023). I extend this literature by showing that indirect incentives can also influence dietary behavior. My findings suggest that government policies can have unintended consequences on unhealthy eating among households with chronic conditions.

This dissertation chapter is organized as follows. Section 1.2 reports background information on diabetes and the insulin out-of-pocket cost policy. Section 1.3 describes the data.

[3] In the broader literature on out-of-pocket costs, most scholars find that a decrease (increase) in out-of-pocket costs leads to positive (adverse) health effects (Huh and Reif, 2017; Dunn and Shapiro, 2019; McNamara and Serna, 2022; Barkley, 2023; Chandra, Flack, and Obermeyer, 2024). In contrast, others find negligible effects (e.g., Chandra, Gruber, and McKnight (2014)).

Section 1.4 illustrates the empirical strategy, and Section 1.5 presents and discusses the results. Section 1.6 concludes.

1.2 Background

In this section, I discuss the clinical background of diabetes, its management, and the policy intervention analyzed in the paper.

1.2.1 Diabetes: Definition and Management

In healthy individuals, the pancreas produces insulin, which regulates blood glucose levels. When this process is impaired, the blood glucose levels chronically rise, leading to diabetes. There are two main types of diabetes: type I and type II.[4] In type I diabetes, the pancreas produces either little insulin or no insulin (Khin, Lee, and Jun, 2023). In type II diabetes, the body does not process insulin correctly and fails to remove sugar from the bloodstream. Type I is genetic and affects only 5% of individuals with diabetes. The majority (95%) has type II diabetes, for which advanced age and obesity are strong predictors. Nearly 80% of individuals with type II are overweight or obese (Iglay et al., 2016). Thus, weight management plays a crucial role in managing type II diabetes and its related comorbidities. A 10% weight loss can improve diabetes management and lower blood pressure (Look AHEAD Research Group, 2007).

Overall, 23% of patients with diabetes use insulin (ADA, 2022). Type II diabetes may require insulin for effective management if blood sugar levels are not well-controlled, i.e., if lifestyle changes and oral medications are insufficient. Unlike type II diabetes, which can be improved or even reversed through sustained lifestyle changes, type I diabetes cannot be cured, and all individuals with type I diabetes must rely on insulin therapy. This difference between type I and type II diabetes has important implications for ex-ante moral hazard. Since type I patients must use insulin and cannot modify their condition through behavioral changes, they face less scope for engaging in risky behaviors. In contrast, individuals with type II diabetes—whose condition can be managed or improved through diet and exercise—may have stronger incentives to adjust their behavior when financial constraints on insulin use are relaxed.

To avoid sharp changes in blood sugar, insulin dosing must accompanied by regular blood sugar monitoring and carbohydrate counting. Carbohydrate counting is needed since insulin dosage is based on carbohydrate intake.[5] The type of the carbohydrate is also crucial for managing blood sugar levels and the disease (Ley et al., 2014). While fiber-rich carbohydrates have a slow impact on glucose levels, foods high in sugars can quickly elevate blood

[4] Other, less common forms of diabetes are excluded in the analyses, both because they are not observed in the data and account for a small share of cases.

[5] The exact insulin dosage is individual-specific, as it depends on the insulin-to-carbohydrate ratio. If an individual consumes 45 grams of carbohydrates and has a 1:15 insulin-to-carb ratio, they would require three units of insulin. In addition, the individual must also consider their current blood glucose level relative to their target.

sugar levels. Frequent blood sugar spikes can lead to the development of comorbidities for diabetic patients (Holman et al., 2008).

In addition to limiting sugars and including fibers, non-saturated fat and protein intake is recommended (Ley et al., 2014). Conversely, saturated fats and cholesterol are associated with an increased risk of cardiovascular disease and hyperlipidemia (Fernandez and Andersen, 2014; Imamura et al., 2016), while excessive sodium intake contributes to hypertension (Grillo et al., 2019). In Section 1.5, I investigate whether and to what extent ex-ante moral hazard is at play. To do so, I examine the effect of the policy on nutrients correlated with comorbidities.

Comorbidities, such as hypertension, hyperlipidemia, and cardiovascular diseases, are common for individuals with diabetes. In the U.S., nearly one-third of individuals with type I diabetes have hypertension. The prevalence is higher among those with type II, with four out of five individuals being affected (Landsberg and Molitch, 2004). Hyperlipidemia, i.e., abnormally high levels of fats in the blood, affects 77% of diabetic patients (Iglay et al., 2016). Hypertension and hyperlipidemia are also risk factors for cardiovascular diseases, such as heart failure (Fuchs and Whelton, 2020). Heart failures can be deadly: a fifth of patients with diabetes die from stroke-related complications (Phipps et al., 2012). These risks highlight the role of preventive behavior and adherence to dietary recommendations, which may be influenced by the changes in out-of-pocket insulin costs.

1.2.2 The Policy

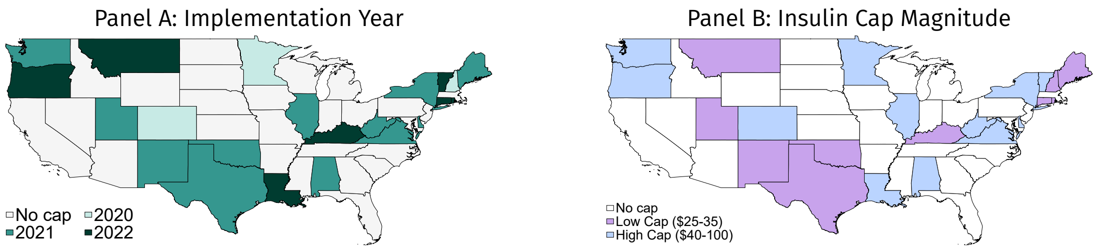
The Cap on Insulin Out-of-Pocket Costs. Starting in 2020, several states in the U.S. passed laws to reduce insulin out-of-pocket costs for individuals with private insurance, who made up the majority (68%) of the population in 2019 (U.S. Census Bureau, 2020). Colorado was the first U.S. state to pass such legislation, which became effective in January 2020 and imposed a $100 cap. Between September 2020 and January 2022, 22 other states adopted similar legislation, establishing caps ranging from $25 to $100.[6] The policy caps out-of-pocket spending for privately insured individuals at $25 to $100 per 30-day insulin supply, depending on the state, with insurers covering any costs above the cap. On average, the policy led to a $20–27 decrease in out-of-pocket costs per month (Giannouchos, Ukert, and Buchmueller, 2024; Ukert, Giannouchos, and Buchmueller, 2024).

In 2019, before the policy was implemented, the average monthly out-of-pocket costs for insulin was $82, according to a report by the Health Care Cost Institute (2024), which analyzed prescription drug claims data from privately insured patients. The report also estimates that 70% of individuals would benefit from a $35/month cap, while 25% would benefit from a $100 cap, as shown in Figure 1.A.1 in Appendix 1.A. Moreover, Bakkila, Basu, and Lipska (2022) find that in 2017–2018, 15% of diabetic patients spent nearly half of their income on insulin prescriptions. These high costs led to decreased insulin use, as well

[6] To date, 26 states have passed legislation that caps insulin out-of-pocket costs. In my analysis, however, I exclude some of the states, as I discuss in Section 1.4. Medicare also introduced a $35 cap on out-of-pocket costs starting from January 2023. As the data extends until December 2022, I am not able to analyze the effect of this policy.

as poor blood sugar control. Similarly, a survey by Herkert et al. (2019) shows that 25.5% of patients with diabetes underutilized insulin due to its high cost.

Figure 1.1. Insulin Out-of-Pocket Cost Cap Policy



Notes: Own data collection from state-level bills. Panel A reports the year each state implemented its cap on insulin out-of-pocket costs. Panel B indicates whether the cap set by each state falls below (low) or above (high) the median cap level.

Figure 1.1 illustrates the timing of the policy implementation (Panel A) and the size of the cap (Panel B).[7] In Panel A, the white-colored states did not implement an insulin cap policy or did so after 2022. The shades of green indicate when the policy was implemented, ranging from light green for states that implemented the policy in 2020 to darker green for those that implemented it in 2022. In Panel B, the map uses purple to highlight states that implemented a “low cap" (i.e., a below-median cap, ranging from $25 to 35) and blue for those with “high cap" (i.e., an above-median cap, ranging from $40 to 100). The maps highlight geographical variation in policy implementation, as states enacted the cap at different times and with caps of varying levels.

Potential Policy Effects and Ex-Ante Moral Hazard. Reducing insulin out-of-pocket costs might affect diet, as insulin dosing is closely tied to carbohydrate intake. To illustrate the potential effects of the cap, consider two scenarios in which, holding income, preferences, and insurance constant, the amount of insulin an individual can afford per day varies. In the first scenario, the individual can buy 20 units per day. In this case, the individual may need to limit carbohydrate consumption to avoid risking elevated blood glucose levels. In the second scenario, the same individual can afford 100 units of insulin per day. In this scenario, the higher insulin access could affect nutritional choices because the additional insulin can be used to compensate for higher carbohydrate intake. The two scenarios mimic the pre- and post-period of the policy, which increases insulin affordability through the insulin cap. While the cap directly benefits those who previously paid above the threshold, it may also affect individuals who were already paying at or below it by allowing them to purchase more insulin at no additional cost.

The scenarios are stylized to highlight a potential shift in individual incentives, which is the central question addressed in this paper. They also underscore that some individuals may have been consuming suboptimally low levels of insulin prior to the policy, not out of preference but due to affordability constraints. Nevertheless, if individuals respond to lower insulin costs by increasing carbohydrate consumption in ways that raise long-term health

[7] Table 1.A.1 reports, for each state, the cap amount, the date the bill was signed, and the date it became effective.

risks, such behavior would be consistent with ex-ante moral hazard. In this context, ex-ante moral hazard refers to a reduction in preventive effort through dietary choices because the immediate financial consequences of risky behavior are lowered. Although the long-term costs of diabetes-related complications remain unchanged, the policy lowers the short-term financial cost of managing blood glucose. This reduction may alter incentives for nutritional choices and increase the risk of developing comorbidities over time. Besides carbohydrate purchases, the policy may also affect dietary decisions more broadly, including total caloric intake. Among others, increased consumption of calories can lead to weight gain, which may exacerbate diabetes-related symptoms and comorbidities.

The reduction in insulin's out-of-pocket cost might also lead to an income effect. The magnitude of this effect depends on pre-policy spending and the magnitude of the cap in the household's state of residence. Households already paying below the cap would observe no savings. However, those spending, for example, $200 per month in a $35-cap state could save approximately $2,000 annually. These savings correspond to 2.9% of the average gross household income in the dataset ($67,000) and could influence grocery purchases and spending on diabetes-related devices. In addition, lower out-of-pocket costs may increase medication utilization, consistent with prior evidence that patients are price sensitive (Goldman et al., 2004; Shankaran and Ramsey, 2015; McAdam-Marx et al., 2024).

1.3 Data

This section first describes the food purchase data, then presents the health and insurance information, and concludes with details on the retailer-level data on diabetes device sales.

1.3.1 Household Data

Food Purchase Data. I utilize the NielsenIQ Consumer Panel (henceforth NCP) data from 2019 to 2022, which is available through the Kilts Center at the University of Chicago Booth School of Business. The NCP contains data on consumer purchases, as panelists scan their purchased items with at-home scanner technology upon returning home from a shopping trip. Shopping trips included are those from supermarkets, drug stores, large and small retailers, and online outlets. The data contain individual demographic information, such as household size, structure, income, education level of the household heads, and ages of all household members. The NCP also includes information on the state of residence, which is relevant for the identification strategy.

For each purchase, the NCP reports the Universal Product Code (UPC)—a unique 12-digit identifier for each product—along with the quantity purchased, store information, and pricing details. Prices are either recorded by panelists or, when available, retrieved from NielsenIQ's store-level data. For a subset of products, NielsenIQ provides additional detailed information on nutritional content.[8] The matched data allow me to construct di-

[8] NielsenIQ started providing this dataset in 2021. Using the UPC, I can assign this information to purchases made in 2019 and 2020. For all years 2019–2022, I have the information for 60% of the transactions, amounting

etary outcomes, such as total grams of carbohydrates purchased per quarter, normalized by household size.

To assess the overall healthiness of household diets, I employ the NCP to construct three diet score measures. Following Oster (2015) and Hut and Oster (2022), I use ratings derived from a survey of 17 doctors who evaluated 59 food categories as a "good" (+1), "neutral" (0), or "bad" (-1) source of calories for diabetic patients. Using product-level purchase data, I categorize items based on product descriptions and assign an average rating to each category, informed by the doctors' assessments.[9] I then compute a household-level diet score by weighting each product's rating by its share of the household's total grocery spending. The spending diet score offers the advantage of complete price coverage, capturing the full composition of the shopping basket. In contrast, nutritional content and serving size data are only available for a subset of purchases. However, the spending die score approach may be sensitive to price variation, as unhealthy foods may be cheaper than healthier options. To address this concern, I construct two alternative diet scores: one based on the share of calories and another on the share of serving sizes relative to the total basket of groceries.

Previous studies support the validity of the NCP data. Einav, Leibtag, and Nevo (2010) found that information on shopping trips, products, and quantities in the NielsenIQ Consumer Panel is generally accurate. For example, they report that approximately 20% of shopping trips are not captured in the data. Most discrepancies are related to prices, which are not central to my analysis, as I primarily rely on data concerning trips, products, and quantities. Moreover, Oster (2018) finds that NCP is representative of the individuals' calorie consumption. She reports that the calorie levels in the NCP correspond to around 80% of the calories reported in the National Health and Nutrition Examination Survey, where panelists keep food diaries. Finally, note that NielsenIQ incentivizes survey participation through monthly prize drawings, redeemable gift points, and gift cards.

Health and Insurance Status Data. The Annual Ailments, Health and Wellness Survey (henceforth, the Ailment Survey) is administered to NCP panelists in the first quarter of the corresponding year. The data are detailed as panelists are asked several questions about their ailments, general health, and diet. Instrumental for my analysis, the survey records whether participants have type I or II diabetes or prediabetes.[10] I include the first two in my analysis and exclude the latter, yielding 9,598 unique households where at least one member has diabetes. The survey also captures information on households' comorbidities, average exercise frequency, and the importance they give to a low-carbohydrate diet. Starting in 2020, I also observe whether a household member has private health insurance.

While the data offer rich details on household behaviors, a key limitation is the partial coverage of health and insurance information due to voluntary participation in the Ailment

---

to 42% of the household spending on groceries. Prior work by Felipe Lozano-Rojas informed the construction of nutritional outcomes.

[9] For example, ice cream received a score of –1, juice scored –0.64, eggs scored 0.7. Full category scores are reported in Table 1.A.2 in Appendix 1.A.

[10] An individual with prediabetes is in a stage where medication is not yet required, but it may become necessary if the condition is not reversed through lifestyle changes, such as improved diet and increased physical activity.

Survey. Only a subset of households (64%) ever participate. To address this limitation, I utilize information from participating households to impute the diabetes status of the remaining panelists based on their purchasing behavior. Specifically, I employ variables observed for all NCP households. I implement a rule-based classification approach using (i) the annual amount in USD spent on diabetic products, (ii) the number of trips made to purchase diabetic products each year, and (iii) the number of packages of diabetic products purchased. Conditional on positive values, I compute quartiles for each variable and classify a household as having diabetes if it falls in the top two quartiles on at least one measure.[11] As shown in Table 1.B.1 in Appendix 1.B, this method correctly identifies 67% of households with diabetes and allows me to recover diabetes status for an additional 1,970 households. Approximately 10% of the households have diabetes, which is similar to the share of diabetes in the U.S. population (11.3%).

Additionally, I impute the insurance status for non-respondents using a logit model with random effects. I use household-level information, such as the head of household age, household size, household income, education in years, and hours worked per week. The imputation results, reported in Table 1.B.2, show that 85% of households with private insurance are correctly identified. In Appendix 1.B, I provide a detailed explanation of the two imputations, and I show that respondents and non-respondents are largely comparable. In Section 1.5.3, I also show that the results are robust to the exclusion of households with imputed insurance and diabetes status.

Final Sample and Descriptives. I construct the final sample, considering that the policy is effective for households with diabetes who hold private insurance. Therefore, I include in the sample households with at least one member with diabetes and at least one member covered by private health insurance. From the sample, I exclude 100 households that move to a different state between 2019 and 2022. Alaska and Hawaii are not included in the dataset, and I exclude two additional states: Louisiana because the law became effective in August 2022, and I only observe a few months in the post-period; Oregon because it passed various insulin-related laws during the period, which were different from the policies implemented in other states. The final sample comprises 69,969 observations and 8,651 unique households, all of which have diabetes, private insurance, and reside in one of the considered states.

Table 1.1 reports households and shopping behavior descriptives.[12] I show the descriptives for households in control states (in columns 2 and 3) and in treated states (columns 4 and 5). States that did not implement the policy are classified as the control group, while those that did are classified as the treated group.[13] Panel A summarizes household characteristics from the first year each household appears in the panel. On average, the heads of

[11] I use this method for its strong performance. See Appendix 1.B for alternatives and comparisons.

[12] Table 1.A.3 in Appendix 1.A reports similar statistics comparing the whole panel to the selected sample.

[13] This classification helps mitigate concerns related to the imputation of diabetes status and insurance coverage. Defining the control group as households without diabetes would increase the risk of bias, as the imputation model produces a relatively high number of false negatives—households that have diabetes but are not identified as such. As a result, treated households (i.e., with diabetes) could be mistakenly included in the control group (i.e., without diabetes), introducing a downward bias in the estimated treatment effect.

household are nearly 61 years old, and households consist of 2.5 individuals. The average years of education is 14.5. In terms of income, households in treated states earn an average of $66,900, while those in control states earn slightly less, at $66,162. The share of marital status (72% married, 7% single), presence of children under six (1%), and racial composition (78% white) are similar across the treatment and control groups.

Panel B reports purchasing patterns for households with diabetes. Quarterly food expenditures are similar across treatment and control states, averaging around $1,300 per household, which corresponds to $200 per person per month. This aligns with the $154 per-person monthly food spending reported by the Bureau of Labor Statistics (2020). The table also presents nutritional content descriptives, including average per-person quarterly quantities. On average, each quarter, households with diabetes buy food containing ∼8,500 grams of carbohydrates, including ∼500 grams of fiber and ∼1,400 grams of sugar. The food purchases contain ∼3,300 grams of fats and ∼1,700 grams of proteins. These quantities correspond to roughly 70,000 total calories per quarter, which represents about 37% of the recommended daily caloric intake of 2,100 calories for a moderately active 61-year-old, i.e., 1,800 for women and 2,400 for men (Cleveland Clinic, 2023).

I also report descriptives for household diet scores, which range from –1 (least healthy) to 1 (most healthy). Households diet score average –0.06 based on spending, –0.25 based on calories, and –0.23 based on serving size. These averages suggest that households tend to eat unhealthily, though not severely so. Notably, the diet score based on spending tends to be higher, reflecting the fact that healthier products often cost more and, therefore, account for a larger share of expenditure.[14]

1.3.2 Retailer Data

I utilize sales data from the NielsenIQ Retail Scanner Data to examine the impact of the policy on sales of diabetes devices. The data contain the weekly units sold, prices, and discounts (if any) for the products. The data also include a unique store identifier, the type of store (e.g., supermarket or drugstore), and the state in which the store operates. The data includes only devices that are sold without a prescription. Because insulin requires a prescription, it is not observed in this dataset. However, I do observe several non-prescription diabetes devices—such as glucose meters and testing strips—that are essential for managing the disease. These devices also represent a substantial share of monthly out-of-pocket expenses (Chua, Lee, and Conti, 2020).

The sample is restricted to diabetes-related devices sold by drug stores. I construct two primary outcome variables: the average sale price and the number of units sold. To account for differences in package sizes, the unit measure reflects the number of individual product units contained in a package (e.g., whether an insulin syringe package contains 10 or 100 syringes). I construct outcomes for four categories of diabetes devices, identified using

[14] Figure 1.A.2 in Appendix 1.A shows the distribution of these scores.

Table 1.1. Panel and Retailer Data Descriptives

	Control States		Treated States	
	Mean	S.D.	Mean	S.D.
Panel A: Household (HH) Characteristics				
HH Heads Age	60.95	11.28	60.85	11.50
HH Size	2.32	1.12	2.37	1.18
HH Heads Education (years)	14.39	2.19	14.43	2.20
HH Income	66,162	25,872	66,900	26,023
HH Income (per person)	33,414	18,172	33,411	18,345
Married	0.72	.	0.71	.
Single HH	0.18	.	0.18	.
Presence of Children <6 yrs	0.01	.	0.01	.
White	0.78	.	0.79	.
Observations	5,517		3,134	
Panel B: Purchasing Behavior				
General Behavior				
Total Spent $/Quarter	1,261	781	1,245	770
Total Spent $/Quarter (per person)	623	425	606	417
Nutrients and Calories (per person)				
Carbohydrate (g)	8,534	6,943	8,472	7,299
Fibers (g)	498	393	499	392
Sugar (g)	1,433	1,917	1,426	1,809
Proteins (g)	1,729	1,219	1,726	1,194
Fats (g)	3,340	2,941	3,306	2,845
Cholesterol (mg)	23	97	20	91
Sodium (mg)	175	333	174	329
Calories	71,115	51,780	70,550	51,996
Diet Score (Share Spending)	-0.06	0.10	-0.06	0.11
Observations	44,841		25,128	
Panel C: Diabetes Device Sales				
Insulin Syringes				
Units	215.28	435.67	223.46	447.20
Price	0.62	0.28	0.61	0.29
Glucose Monitors				
Units	8.95	19.09	9.08	19.52
Price	11.52	7.90	11.43	8.03
Glucose Test Stripcs				
Units	387.31	494.84	388.81	490.51
Price	2.18	2.01	2.24	2.05
Ketone Test Strips				
Units	26.17	54.39	28.61	56.19
Price	0.24	0.04	0.24	0.04
Observations	894,624		528,240	

Notes: Own calculations on 2019–2022 NielsenIQ Panel and Retailer Data. Treated states implemented a cap on out-of-pocket costs for insulin between 2020 and 2022, while control states did not. In Panel A, the observation is at the household level using information from the first year the household enters the panel. In Panel B, the observation is at the household-quarter level, and values are in grams (g) or milligrams (mg). In Panel C, the observation is at the retailer month-level. Prices are unit prices in USD.

NielsenIQ product descriptions.[15] First, I examine insulin syringes, which are used to inject insulin.[16] Second, I study the purchase behavior around glucose monitors, which are small devices that are used to measure blood glucose levels several times a day. Third, I assess the effect of the policy on glucose testing strips, which are used together with glucose monitors. Together, these two categories provide insight into testing behavior. Fourth, I examine ketone strip sales, which are used to monitor high levels of ketones, typically during illness or prolonged hyperglycemia.

[15] Up to 2020, NielsenIQ reports two broad diabetes-related categories: insulin syringes and testing products. Glucose monitors can be identified using keywords such as “BG K" and “MNSYS." From 2021 onward, more detailed product classifications allow for the identification of additional products in earlier years.

[16] Although insulin pumps and pens are alternatives to syringes, 65% of individuals with diabetes in the U.S. do not use them (Endocrine News, 2014).

Panel C of Table 1.1 reports descriptive statistics on diabetes supplies sales from the monthly retailer-level data from over 30,000 drug stores for control states (in columns 2 and 3) and in treated states (columns 4 and 5). The values are similar across treated and control states, especially in terms of prices. In treated states, around 223 syringes, nine glucose monitors, 389 glucose test strips, and 29 ketone test strips are sold per month in each store. Prices per unit range from an average of \$0.24 for ketone test strips to \$11 for glucose monitors.

1.4 Estimation and Identification

Staggered Difference-in-Differences Approach. The implementation of the policy is staggered and is rolled out at nine different points in time. To address the “forbidden comparison" problem that arises when using a two-way fixed effect in such a setting, I employ the staggered difference-in-differences approach à la Callaway and Sant’Anna (2021).[17] This approach is useful in my setting since the short-run treatment effects may be stronger than the long-run effects, and there are potentially heterogeneous treatments across cohorts that might vary over time. Given that the policy was implemented over two years, its effects may change over time.

To describe the research design, I employ the potential outcome framework à la Robins (1986) and define D_{it} as an indicator equal to one if household i resides in a state that has implemented the policy cap in quarter t. The outcome of interest Y_{it} can be written as $Y_{it} = Y_{it}(1)D_{it} - Y_{it}(0)(1 - D_{it})$, where $Y_{it}(1)$ and $Y_{it}(0)$ are the potential outcomes with and without the policy, respectively. The typical challenge in this setting is the inability to observe a unit with and without the treatment simultaneously, i.e., it is not possible to observe the counterfactual. To overcome this challenge, difference-in-differences approaches rely on comparing individuals in the treated states to those in the control states.

Callaway and Sant’Anna (2021) approach this comparison by estimating group-time average treatment effects. This group-time effect is defined as the average treatment effect (ATT) for group g at time t, and a group is defined by the period in which units are first treated. If an observation is never treated, then $G = \infty$. Formally,

$$ATT(g,t) \;=\; \mathbb{E}[Y_t(g) - Y_t(0)|G_g \;=\; 1] \tag{1.1}$$

where G_g is an indicator equal to one if an observation is treated for the first time in period g. For each $g \in \mathcal{G}$ and $t \in \{1\ldots,\mathcal{T}\}$,[18] the regression estimated to obtain $ATT(g,t)$ is

[17] This problem was highlighted in the recent staggered difference-in-differences literature. When estimating a two-way fixed effect such $y_{it} = a_i + b_t + \theta(treat_i \cdot post_t) + e_{it}$, the main coefficient of interest is θ, being the effect for treated individuals in the post period (i.e., after the policy implementation). However, θ is estimated by comparing the same unit across time and comparing different units with and without the treatment at the same time t. This second difference is problematic because one of the comparisons involves later-treated units versus earlier-treated units. This difference is a “forbidden" comparison, especially if the treatment effects are expected to be heterogeneous, as the parallel assumption no longer holds.

[18] Note that $\bar{g} = \max_{i=1,\ldots,n} G_i$ is the maximum G observed in the data and that $\mathcal{G} = \mathrm{supp}(G) \setminus \{\bar{g}\} \subseteq \{2,3,\ldots,\mathcal{T}\}$ denotes the support of G excluding $\bar{g}$. $\mathcal{T}$ is the maximum number of period observed.

equivalent to:

$$Y_{it} = \alpha_1^{g,t} + \alpha_2^{g,t} \cdot G_{g,i} + \alpha_3^{g,t} \cdot I\{T = t\} + \beta^{g,t}(G_g \times I\{T = t\}) + \gamma^{g,t} \cdot X_i + \varepsilon^{g,t} \tag{1.2}$$

where Y_{it} is an outcome of interest for household i at quarter t. $\alpha_1^{g,t}$ is the intercept, while $\alpha_2^{g,t}$ is the group fixed-effect for group g and $\alpha_3^{g,t}$ the time t fixed effect. $\beta^{g,t}$ is equal to $ATT(g,t)$ and thus is the coefficient of interest, capturing the treatment effect of the policy implementation. X_i controls for the state-level share of Democratic votes in the 2016 presidential election. This control is included to account for the probability that a state implements the insulin cap or other health-related policies that might influence the outcome.[19] $\varepsilon^{g,t}$ is the normally-distributed error term. As several states never implement the policy, I employ the households in never-treated states as the control group, defined as $C = 1$. The $ATT(g,t)$ for each $t > g$ is then given by:[20]

$$\begin{aligned} \text{ATT}(g,t) = \mathbb{E}\Bigg[& \left(\frac{G_g}{\mathbb{E}[G_g]} - \frac{\frac{p_g(X)C}{1-p_g(X)}}{\mathbb{E}\left[\frac{p_g(X)C}{1-p_g(X)}\right]} \right) \\ & \times \left(Y_t - Y_{g-1} - \mathbb{E}\left[Y_t - Y_{g-1} \mid X, C = 1\right]\right)\Bigg] \end{aligned} \tag{1.3}$$

Equation (1.3) provides a doubly robust method to estimate the ATT, building on Sant'Anna and Zhao (2020). The method is doubly robust since the ATT is identified even if either (but not both) the propensity score model or the outcome regression models are misspecified. This doubly robust method has two building blocks, namely the inverse probability weighting and the outcome regression. The first block adjusts the distribution of covariates between the treated and untreated groups. It involves estimating the propensity score $p_g(X) = p_{g,T}(X) = P(G_g = 1|X, G_g + C = 1)$, which represents the probability of being first treated in period g, conditional on covariates and on either being a member of group g or not participating in the treatment at any time. The propensity score is estimated via logit using the share of Democratic votes. The term $\frac{G_g}{\mathbb{E}[G_g]}$ represents the resulting weight for the treated, while $\frac{\frac{p_g(X)C}{1-p_g(X)}}{\mathbb{E}\left[\frac{p_g(X)C}{1-p_g(X)}\right]}$ is the weight for the control group. The second block, instead, represents the difference in outcomes adjusted by the expected outcome difference for the comparison group.[21]

To illustrate key characteristics relevant to this approach, I present the staggered policy rollout in Figure 1.2. I report the two-letter abbreviation of the state on the y-axis and the quarter of implementation on the x-axis. Hence, each square represents a state-quarter combination. The light blue indicates no treatment, while the dark blue indicates the treatment. I observe households over 16 quarters, and I consider four quarters before the policy

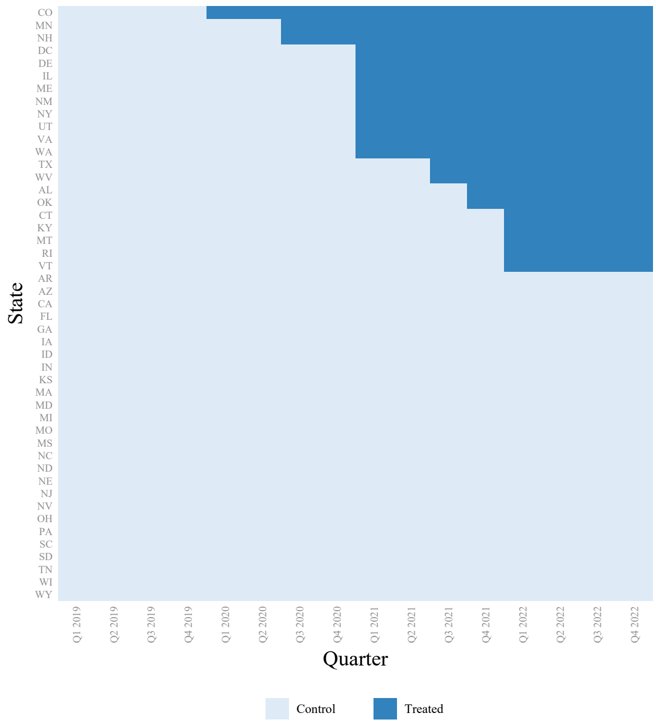
[19] Given that the treatment itself might influence the post-treatment covariates, I use the pre-treatment values from the 2016 presidential elections. I obtain the state-level Democratic vote shares from https://www.fec.gov/resources/cms-content/documents/federalelections2016.pdf.

[20] When the pre-treatment covariates play no role in identification, the ATT can be rewritten as $ATT(g,t) = \mathbb{E}[Y_t - Y_{g-1}|G_g = 1] - \mathbb{E}[Y_t - Y_{g-1}|C = 1]$.

[21] I implement the method using the `csdid` command in Stata (Rios-Avila, Sant'Anna, and Callaway, 2023).

implementation (pre-period) for the first state and four quarters after implementation (post-period) for the last state. The figure shows that some states belong to the same group g as they implement the policy simultaneously, with nine states enacting it in the first quarter of 2021. The figure also highlights that several states never implement the policy within the considered period.

Figure 1.2. Policy Implementation, Staggered Timing



Notes: This figure shows the staggered rollout of the insulin out-of-pocket cost cap by state and quarter.

Identification Assumptions. For the identification strategy to be valid and for a causal interpretation of the results, three assumptions need to be satisfied.[22] First, the approach relies on the *irreversibility of treatment assumption*. Formally, this is defined as follows: $D_1 = 0$ almost surely (a.s.), and for $t = 2, \ldots, \mathcal{T}, \quad D_{t-1} = 1$ implies that $D_t = 1$ *a.s.* This means that no unit is treated in the first period of the data and that once a unit becomes treated, it stays treated. This assumption is satisfied in my setting, as it is also represented in Figure 1.2.

Second, the method relies on a (conditional) *no-anticipation* assumption. Formally, for all $g \in \mathcal{G}, t \in \{1 \ldots, \mathcal{T}\}$ such that $t < g$:

$$\mathbb{E}[Y_t(g)|X, G_g = 1] = \mathbb{E}[Y_t(0)|X, G_g = 1] \quad a.s. \tag{1.4}$$

[22] An additional assumption is having panel data, which is the case in my setting. The case with cross-section data is an extension in Callaway and Sant'Anna (2021).

This assumption requires that the household does not actively choose the treatment status.[23] Given that the bill was approved a few months before being enacted, there could be potential anticipation effects, which might lead households to modify their behavior in three possible ways.[24] First, households could move to take advantage of the policy. However, in my data, only around 100 households moved during the analysis period. Hence, this seems to be a minor concern and I exclude those households from the analysis. Second, people may switch insurance status to take advantage of the cap. I do not find any evidence that a significantly larger share of the population takes up insurance after the policy is implemented.[25] Moreover, in my analysis, I limit my sample to individuals who already have insurance before the reform, a selection that partially addresses these concerns. Third, households may change their behavior in ways that directly affect the outcomes of interest. For example, households could ration insulin a few months before the policy is implemented to save money, knowing that they will soon be able to buy more insulin. Consequently, they may also reduce their carbohydrate purchases. The event study estimates, both on nutritional outcomes and on diabetes device sales, show no evidence of this.

Third, the method relies on the (conditional) *parallel trends assumption*, based on a never-treated group:

$$\mathbb{E}[Y_t(0) - Y_{t-1}(0)|X, G_g = 1] = \mathbb{E}[Y_t(0) - Y_{t-1}(0)|X, C = 1] \quad a.s. \tag{1.5}$$

for all $g \in \mathcal{G}, t \in \{2 \ldots, \mathcal{T}\}$ s.t. $t > g$. This assumption states that households' dietary outcomes in treated states would have evolved parallel to those in control states, had they not been treated. Given that parallel trends in the *post* period cannot be directly tested due to the absence of a counterfactual, a common approach to evaluating the credibility of this assumption is to check whether they hold in the pre-period. I show that they do in Section 1.5. Nevertheless, the COVID-19 pandemic could be a threat to this assumption. On the one hand, some shocks likely affected both treated and control states similarly, and so the time-fixed effects partially account for them. On the other hand, in certain cases, treated states reacted differently than untreated states. Therefore, in Section 1.5.3, I present various robustness checks to address such concerns.[26]

Finally, a potential threat to the identification strategy is the possibility of spillover effects, such as individuals crossing state borders to benefit from the insulin price cap implemented in a neighboring state. However, this is unlikely, as the cap applies only to individuals who are enrolled in insurance plans in the states enacting the law.

[23] Under this assumption, it follows that $ATT(g, t) = 0$ for all pre-treatment periods $t < g$.

[24] In Colorado, the law was passed in May 2019 and was enacted in January 2020.

[25] I use yearly data on insured individuals by state to check whether individuals take up more insurance in the first and second years after the policy is implemented. In Figure 1.A.3 in Appendix 1.A, I plot the share and confidence intervals of insurance coverage from two years before the policy implementation to two years after.

[26] While individuals with diabetes may have been more affected by COVID-19, this concern is mitigated in my context, as the analysis is restricted to diabetic households.

1.5 Results

In this section, I begin by examining the effects of the policy on the purchases of carbohydrates and two key subtypes: fiber and sugar. I then assess the scope of ex-ante moral hazard by analyzing additional dietary outcomes and exploring the heterogeneity of the effects across groups. Additionally, I study the impacts on medication use, diabetes management, and short-term disease management. Finally, I demonstrate the robustness of my findings across alternative specifications.

1.5.1 Nutritional Choices

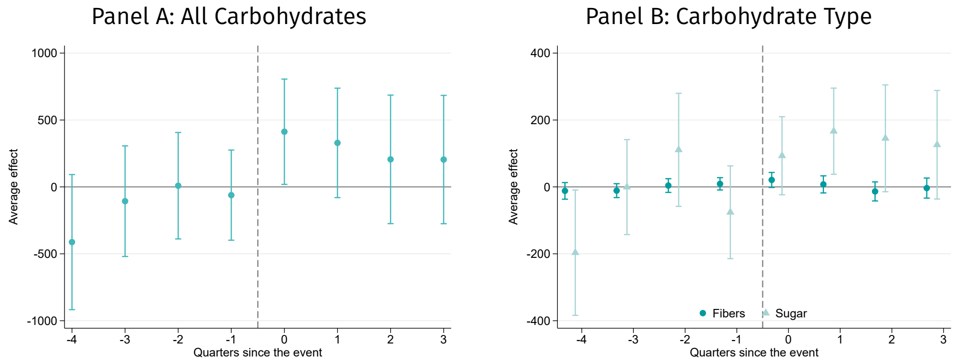
Ex-Ante Moral Hazard: Effects on Carbohydrate, Fiber, and Sugar Purchases. As highlighted in Section 1.2, carbohydrates and insulin consumption are tightly linked. Therefore, I first investigate the effect of the policy on carbohydrate purchases. Figure 1.3 shows the dynamics of the treatment effect for all types of carbohydrates (Panel A) and two subtypes of carbohydrates: fiber and sugar (Panel B). In particular, Panel A plots the event study for grams of carbohydrates purchased per quarter, adjusted by household size. It reports the coefficients for four quarters before and after the policy implementation. Considering both panels, all but one of the coefficients of the pre-period are close to zero and are not significant. As shown in Table 1.A.4 in Appendix 1.A, the average coefficient in the pre-treatment periods is not statistically significant. Moreover, I test the parallel trend hypothesis in the pre-period, which cannot be rejected. This suggests that there are no pre-trends prior to the policy implementation.

The interpretation of the coefficients in Figure 1.3 deviates slightly from that of a standard two-way fixed effects (TWFE) event study. The key difference lies in the coefficient for period $t-1$, which is typically omitted in TWFE specifications and serves as the reference period. This difference is due to the identification proposed by Callaway and Sant'Anna (2021). In the post-period, as in the TWFE, the coefficient for each period $t+k$ is interpreted relative to $g-1$, the last period before treatment for group g. The same logic applies to the control group. This is analogous to the TWFE framework, where $t-1$ is the baseline period. Thus, the coefficient at $t+1$ captures the difference between $t+1$ and $g-1$. In contrast, pre-treatment coefficients are identified using adjacent-period differences, such as between t and $t-1$. Table 1.A.4 in Appendix 1.A reports the results with coefficients normalized at $t-1$, which is comparable to the TWFE approach.

Panel A of Figure 1.3 reports the effect of the policy on total carbohydrate purchases. In the first quarter following implementation, households increased their carbohydrate purchases by 412.5 grams. This gram amount corresponds to the carbohydrate content of 11 cups of cooked white rice. Given a baseline average of 8,472 grams per quarter in the pre-period, this represents a 4.8% increase. This increase is consistent with the scenarios discussed in Section 1.2.2. This change might be due to lower incentives in healthy nutritional choices, consistent with an ex-ante moral hazard interpretation. However, an alternative explanation for this effect could be that households have been consuming suboptimal levels of carbohydrates prior to the policy due to limited access to insulin. If so, the observed increase could reflect optimal consumption rather than a behavioral distortion.

To understand whether the change in consumption reflects ex-ante moral hazard or improved carbohydrate purchase, I examine whether there are changes in the composition of carbohydrates purchased. An increase in carbohydrates, such as fiber, could suggest that the policy leads to optimal consumption, while a rise in sugar purchases could indicate evidence of ex-ante moral hazard. Dietary fiber is generally considered a beneficial carbohydrate, as it does not cause glycemic spikes and is associated with improved glycemic control, reduced body weight, and decreased risk of comorbidities (McRae, 2018; Reynolds, Akerman, and Mann, 2020). These positive effects are observed for both types of diabetes and across a range of fiber intakes. In contrast to fiber, excessive sugar consumption can lead to blood sugar spikes, which require additional insulin to regulate. In particular, repeated blood sugar spikes can lead to both short-term adverse health effects and long-term development of comorbidities, such as heart disease (Holman et al., 2008; Malik et al., 2010). Therefore, sugar purchases capture a risky behavior that increases the long-run probability of developing comorbidities, consistent with the definition of ex-ante moral hazard.

Figure 1.3. Carbohydrates, Purchased Grams



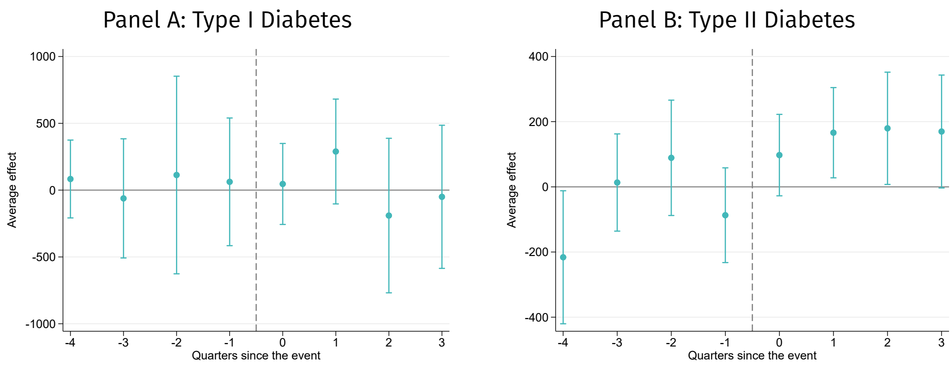
Notes: The figure reports the event study estimates and the 95 percent confidence intervals using the Callaway and Sant'Anna (2021) estimator. I use NielsenIQ Panel Data and the observation is at the household-quarter level.

Panel B of Figure 1.3 plots the estimates for two carbohydrate subtypes: dietary fiber (dark green) and sugar (light green). The dark-green circle estimates indicate that fiber purchase increases by 20.7 grams per person (4.1%). The light green triangles of Panel B in Figure 1.3 show that there is a statistically significant increase in sugar purchases in the post-period. The average effect in the entire post-period corresponds to an increase of 132.8 grams (9.3%), as shown in Table 1.A.4 in Appendix 1.A.

All in all, the results from Panels A and B of Figure 1.3 suggest that a shift in incentives following the policy drives the observed changes in household purchasing behavior. The alternative explanation—that households are reaching an optimal level of carbohydrate consumption—is not supported, given the absence of a significant effect on fiber purchases. Instead, the sugar purchase findings provide suggestive evidence that households substitute towards products with a higher sugar content, particularly as overall carbohydrate purchases return to pre-policy levels after one period. These results are consistent with an ex-ante moral hazard interpretation, in which individuals engage more in risky behavior after a decrease in out-of-pocket costs.

To further investigate whether ex-ante moral hazard is driving the observed behavioral changes, I examine whether there are heterogeneous effects by diabetes type. As discussed in Section 1.2.1, there are strong differences between the two main types of diabetes. Type I is genetic and requires insulin regardless of nutritional choices. Instead, type II can often be managed through diet and exercise. Therefore, individuals with type II diabetes may have greater scope for ex-ante moral hazard, as their disease management depends more heavily on lifestyle choices. If ex-ante moral hazard is driving the behavioral changes observed, one would expect a stronger response among those with type II diabetes, especially on sugar.

Figure 1.4. Sugar, Purchased Grams, by Diabetes Type



Notes: The figure reports the event study estimates and the 95 percent confidence intervals using the Callaway and Sant'Anna (2021) estimator. I use NielsenIQ Panel Data and the observation is at the household-quarter level.

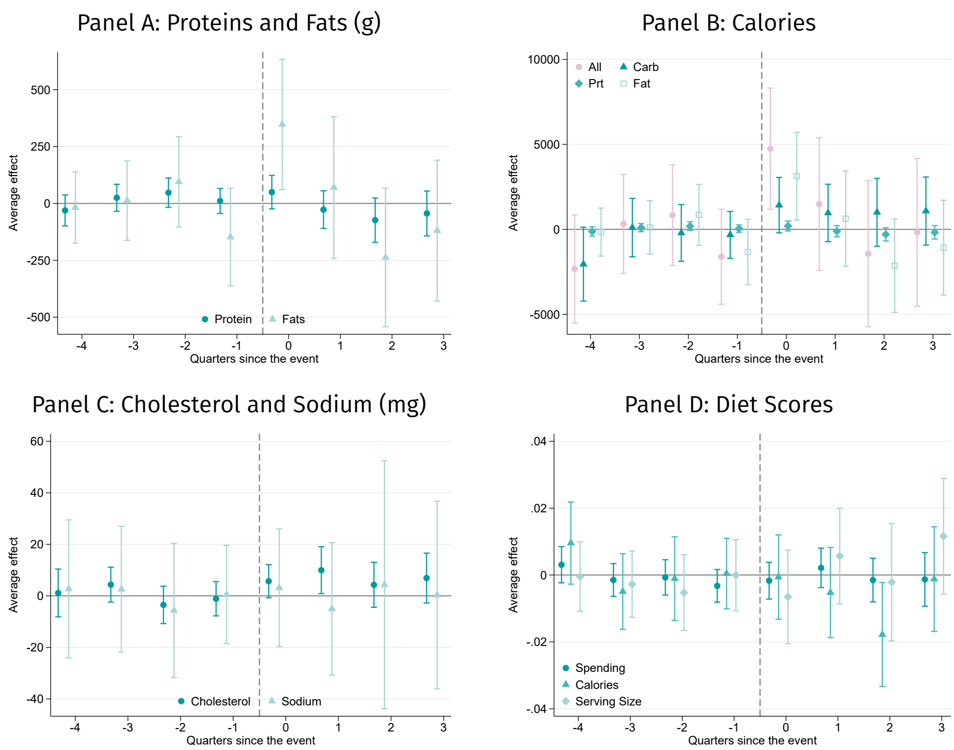
Figure 1.4 shows the estimated effects on sugar purchases by diabetes type.[27] Panel A shows no statistically significant change in sugar purchases among households with Type I diabetes. While these null effects partially reflect the relatively small share of Type I households in the sample (around 10%), they are consistent with the idea that this group has limited scope to adjust their dietary choices in response to changes in insulin out-of-pocket costs. In contrast, Panel B reports the estimates for households with type II diabetes. For this group, sugar purchases increase by 11% in the post-policy period, with the effect both sustained over the four periods and statistically significant. Figure 1.A.4 in Appendix 1.A reports the results for carbohydrates and fiber. Households with type II diabetes increase their overall carbohydrate purchases by 4.3% in the first quarter of the post-period (statistically significant at the 10% level). Instead, households with type I diabetes exhibit a non-statistically significant upward trend in fiber purchases. These results highlight differences in trends between type I and type II diabetes. The different responses in sugar purchases support the interpretation that the policy led to higher engagement in risky behaviors, which

[27] Note that I do not impute the diabetes type. Therefore, this sample includes only households that responded to both surveys. In Section 1.5.3, I show that results based on this subsample are comparable to the baseline results.

is consistent with an ex-ante moral hazard interpretation. Indeed, the effects are driven by those for whom lifestyle plays a greater role in disease management. Since households with type II diabetes are driving the results, I restrict the analysis to this subgroup in the rest of this section.

Extent of Ex-Ante Moral Hazard: Evidence from Additional Nutrients. In this subsection, I investigate whether reducing the out-of-pocket expenses of insulin affects dietary habits that extend beyond increased consumption of sugar and carbohydrates. These findings provide more information on the extent of the policy's impacts. In particular, it helps assess whether behavioral responses extend beyond sugar—which is directly linked to insulin use—to dietary choices less closely associated with the prescription drug.

Figure 1.5. Other Nutrients, Calories, and Diet Scores



Notes: The figure reports the event study estimates and the 95 percent confidence intervals using the Callaway and Sant'Anna (2021) estimator. The y-axis in Panel A is measured in grams (g), while the y-axis in Panel C is in milligrams (mg). The source of the data is in the NielsenIQ Panel data. The observation is at the household-quarter level. The sample is restricted to households in which at least one member has type II diabetes.

Panel A of Figure 1.5 plots the coefficients for proteins (dark green) and fats (light green) purchased. Alongside carbohydrates, proteins and fats are the other two macronutrients, namely essential nutrients that the body requires in large quantities to produce energy. Protein purchases exhibit a 49.7 gram (3.0%) increase in the first quarter after the policy is implemented. However, these results are not statistically significant. Although protein consumption is encouraged, high-protein diets are not found to be a better disease management method than other diets (Larsen et al., 2011). The light green triangle reports the coefficients for grams of fat, showing a significant increase of 347.2 grams (11.6%) in the first quarter of the post-period. There are, however, different types of fats: saturated fats should be avoided (Ley et al., 2014). In meta-analyses, a reduction in dietary saturated fats was associated with improvements in glucose management and a decrease in cardiovascular risk (Imamura et al., 2016; Schwab et al., 2021). Figure 1.A.5 in Appendix 1.A shows that there is no change in saturated fat purchases after the policy is implemented.

In Panel B of Figure 1.5, I plot the coefficients both for the overall calories (pink circle) and the calories from the three macronutrients, i.e., carbohydrates, proteins, and fats. Calories serve as a comprehensive measure of dietary intake, as they represent a weighted sum of the three macronutrients: carbohydrates, proteins, and fats. One gram of fat corresponds to nine calories, while one gram of carbohydrates and one gram of protein each correspond to four calories. Therefore, small changes in each of these nutrients could lead to more substantial differences in the overall caloric intake. Importantly, weight loss accompanied by substantial calorie restriction can even reverse the progression of type II diabetes (Wilding, 2014). Calorie purchases increase by 7.1% in the first quarter, exhibiting only a temporary effect. The overall calorie coefficients follow a pattern similar to that of fats, which is consistent with fats exerting a higher impact on calories.

I also investigate the policy's effect on essential micronutrients, which can provide additional information on households' responses to the policy. I plot the coefficients for cholesterol (dark green) and sodium (light green) in Panel C of Figure 1.5. Cholesterol can contribute to cardiovascular diseases, and excessive sodium intake can lead to hypertension. While I find no significant changes in sodium purchases, I observe a 6.7 mg increase in cholesterol in the post-period, which is significant at the 10% level. This represents a substantial increase of 39.4% in percentage terms. The effect is temporary and returns to pre-policy levels in subsequent periods.

To assess the overall effect of the policy on diet, I employ three diet scores, which are composite measures that capture the overall healthiness of a household's dietary choices. Panel D of Figure 1.5 reports the event study coefficients for three different diet scores, from darkest to lightest green, for spending, calories, and serving size, respectively. The three diet scores yield mixed results, with the findings varying depending on the specific definition employed. First, the spending diet score exhibits no effect after the policy implementation. Among the diet scores, the spending diet score appears to be the most precisely estimated, as its coefficient is close to zero, and standard errors are smaller. Second, there is some evidence of a worsening in the households' diet using the share of calories diet score, which becomes statistically significant two quarters after the policy implementation. Third, the diet score, calculated using the share of serving size, shows a slight upward trend in $t+2$ and $t+3$. Overall, the mixed results suggest that the policy does not lead to substantial changes in overall dietary quality. However, the diet score analysis also appears to mask important behavioral responses, which are shown in the nutrient-level analysis.

Heterogeneous Effects on Nutritional Choices. To further rationalize these findings, I investigate whether there are heterogeneous effects across six households' characteristics.[28] For five characteristics, I split the sample based on whether the value falls below ("low") or above ("high") the median. For one variable, i.e., comorbidities, I use a binary indicator that takes the value of one if the household reports any comorbidity. In Figure 1.6, I report the average coefficient from the first quarter of the event study for sugar purchases. Additionally, in Figure 1.A.6 and 1.A.7, I report the coefficients for the other macro- and micronutrients considered, as well as calories and diet scores. Overall, although not all of the estimated effects are statistically significant, I find that all but one group increases their sugar purchases. Across outcomes, there are no statistically significant differences between groups. However, there is variation when examining specific household characteristics and dietary outcomes.

The first three dimensions of heterogeneity I consider are demographic characteristics, namely head age, head education, and household income. These are reported in the first three rows of Figures 1.6, 1.A.6 and 1.A.7. First, I find that households with older heads exhibit a statistically significant coefficient (at the 10% level) for sugar purchases, while those with younger heads do not. Older households also appear to buy slightly more fat and cholesterol. These effects result in higher calorie purchases for both groups. However, the overall healthiness of the diet remains unchanged. In contrast, differences across education and income levels are more pronounced. Among higher-educated households, I observe minor, non-statistically significant increases mostly in "good" nutrients—i.e., fiber—but not in "bad" ones—i.e., fats and calories. For lower-educated households, instead, I observe an increase in the purchase of sugar and sodium, which is usually found in processed and packaged foods (Quader et al., 2017). Lower-income households exhibit a significant increase in sugar purchases and also raise their fat consumption, contributing to a small decline in their diet score. Conversely, higher-income households do not significantly change the purchase of any other nutrients. These different reactions likely reflect varying food choices driven by differences in income and budget constraints.

In addition, I examine heterogeneity along dimensions of health status and health consciousness, using self-reported information from the Ailment Survey. First, I consider the importance of the household following a low-sugar diet. Survey responses are on a Likert scale from one, "Not important at all," to five, "Very important," with values above four classified as above the median. Second, I use information on exercise frequency to classify households relative to the median. Households that exercise three times per week or less are categorized as below the median. Physical activity is broadly defined to include yoga, walking, and gardening, among other activities. Panelists report their exercise frequency by selecting from eight response options, ranging from "never" to "every day." Third, I construct an indicator equal to one if the household has other comorbidities related to diabetes and diet, such as hypertension, heart disease, and high cholesterol. These heterogeneity dimensions allow further to corroborate the ex-ante moral hazard interpretation of the findings.

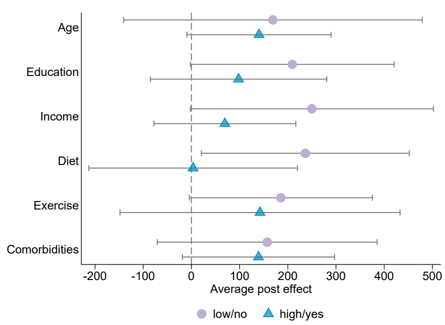
[28] For treated households, I take the value of the year prior to the policy implementation. For other households, I use the value from 2019, as the first states were treated in 2020.

First, to support an interpretation consistent with ex-ante moral hazard, one would expect households that place less importance on a low-sugar diet to respond more strongly to the policy. These households already appear to place lower value on preventive health behavior and dietary disease management. Conversely, households that place greater importance on sugar reduction should exhibit little or no change. The results align with the ex-ante moral hazard interpretation and provide evidence against the alternative explanation of optimal consumption after the policy implementation. I find that households below the median increase their purchases of sugar and carbohydrates more than those above the median. Notably, households that report assigning high importance to a low-sugar diet do not change their purchasing behavior in response to the policy.

A similar hypothesis applies to exercise frequency. If ex-ante moral hazard is the driving mechanism, households with lower exercise frequency, i.e., those less engaged in preventive health behaviors, should respond more strongly to the policy. In line with this hypothesis, I find that households below the median in exercise frequency increase their purchases of carbohydrates, fiber, fats, and calories. They also increase their consumption of saturated fat and cholesterol, both of which are associated with adverse long-term health outcomes. These patterns suggest that less active households, similar to those placing lower importance on diet, may rely more heavily on insulin and less on lifestyle-based disease management.

Finally, households affected by comorbidities might be more cautious in their dietary choices. Conversely, households without comorbidities might have a larger scope for ex-ante moral hazard and risky behaviors, as they are healthier. I find that households without comorbidities increase their purchases of carbohydrates, calories, and cholesterol. For this group, all three diet score measures show a deterioration in dietary quality following the policy implementation. Conversely, I find no increase in carbohydrate purchases among households with comorbidities, though they do increase purchases of sugar and fat. Importantly, these households do not increase cholesterol consumption, which could further worsen their health.

Figure 1.6. Sugar - Heterogeneity

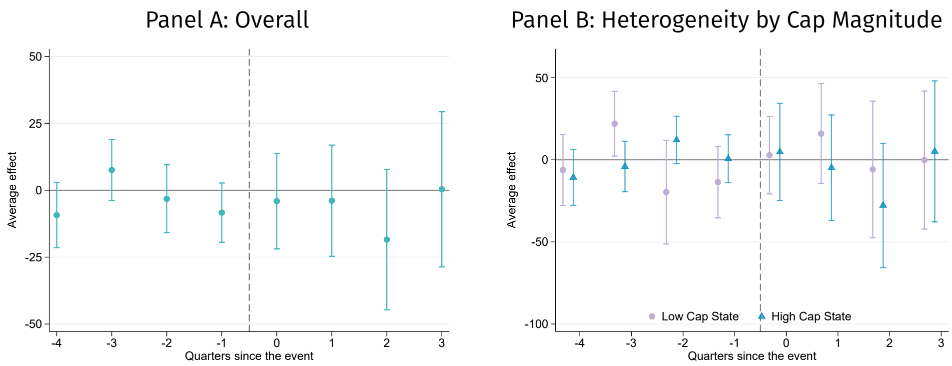


Notes: The figure reports the first-quarter event study estimates and the 95 percent confidence intervals using the Callaway and Sant'Anna (2021) estimator. Source of the data is the NielsenIQ Panel. The sample is restricted to households in which at least one member has type II diabetes.

Income Effect. An alternative explanation for the results presented in the previous subsections is the presence of an income effect. Specifically, the reduction in out-of-pocket insulin costs may increase households' disposable income, which could lead to higher overall grocery spending and, consequently, changes in the nutritional content of food purchases. To exclude this alternative explanation, I construct an outcome measuring per capita household grocery expenditures.

As shown in Panel A of Figure 1.7, I find no statistically significant change in grocery spending following the policy implementation, suggesting that an income effect does not seem to be at play. Given that the size of the income effect may vary with the size of the cap, Panel B reports results separately for states with caps below the median (purple circle) and above the median (blue triangle). The differences between the two groups are minor and do not suggest the presence of income effects that are different across states. Figure 1.A.8 in Appendix 1.A shows heterogeneity analyses by household characteristics, as in the previous subsection. I plot the average post-period coefficient over the four quarters after the policy implementations. Overall, I find no significant differences in grocery spending across groups, including when comparing households with below- and above-median income levels.

Figure 1.7. Total Grocery Spending



Notes: The figure reports the event study estimates and the 95 percent confidence intervals using the Callaway and Sant'Anna (2021) estimator. I use NielsenIQ Panel Data and the observation is at the household-quarter level. I only consider households with type II diabetes.

Taken together, the results presented in this section suggest that a decrease in out-of-pocket costs leads to an increase in risky behaviors. The policy has temporary effects on several dietary outcomes and a sustained effect on sugar purchases. Sugar is an important driver of insulin needs, as its consumption leads to short-term blood glucose spikes that may require more frequent insulin use. The effect is particularly pronounced for households with type II diabetes, who can often manage the disease through diet and exercise. These patterns are consistent with ex-ante moral hazard for households with type II diabetes. By reducing out-of-pocket insulin costs, the policy lowers the financial cost of sugar consumption. However, increased sugar consumption increases the risk of developing comorbidities in the future. The absence of persistent changes in other nutritional choices suggests that the policy does not lead to a widespread decline in preventive dietary behavior. Instead, the response appears narrowly focused on sugar, which is most directly linked to insulin use and, thus, to the incentives directly affected by the policy.

1.5.2 Diabetes Devices for Disease Management

In this subsection, I investigate whether the policy leads to changes in sales of medication and testing devices. These outcomes contribute to a broader assessment of its impact by offering insight into how the policy influences healthcare utilization and disease management. To do so, rather than using household-level data as in the previous section, I use store-month-level data because diabetes devices are purchased infrequently at the household level, which results in few observations. Although I run the analysis at the store-month level, the identification strategy discussed in Section 1.4 remains unchanged.

I study the effect of the insulin policy on the sales of diabetes devices. Even though the insulin cap policy does not cover these devices, their usage may still be indirectly affected by the policy, as they are complements to insulin usage.[29] The four panels of Figure 1.8 report the results on the inverse hyperbolic sine transformation of units sold for the considered devices. I report the coefficients for the twelve months preceding and following the policy implementation.

The first outcome of interest for evaluating medication utilization is insulin consumption, which is not observed in the data. In the absence of these data, purchases of insulin syringes serve as a proxy for medication utilization. Syringes are necessary for injecting insulin, so an increase in syringe purchases can plausibly reflect higher insulin consumption, as well as increased injection frequency. If anything, using syringe sales as a proxy may introduce a downward bias in the results. This bias can occur if individuals increase their insulin dosage without increasing the number of syringes used—for example, by using a larger dose per injection. Panel A shows that, in treated states, there is a sustained increase in insulin syringe sales following the implementation of the policy. While the upward trend begins in the first months of the post-policy period, the effects become statistically significant starting in the fifth month. One possible explanation for this delay is a gradual adjustment in insulin consumption. Individuals may initially make only marginal increases in usage, which may not immediately translate into detectable increases in syringe sales.

Additionally, I investigate the policy's effect on diabetes testing behavior by analyzing purchases of glucose monitors and glucose testing strips. The latter are inserted into a glucose monitor with a drop of blood. Glucose monitors display blood glucose levels and, in some models, provide additional indicators—such as color-coded alerts—to signal whether the values are within a normal range for the individual.[30] Panel B shows no change in glucose monitor sales, while Panel C shows an increase in testing strips already in the first period of the policy implementation. Because glucose monitors are durable devices that can last up to five years, individuals are unlikely to purchase them frequently, which may explain

[29] In Connecticut, some devices are covered. Results remain robust when excluding the state (see Section 1.5.3).

[30] The NielsenIQ datasets do not include purchases of continuous glucose monitors (CGMs), as these devices require a prescription. However, their adoption remains limited, as recent estimates suggest that only 13% of individuals with type II diabetes use them (Mayberry et al., 2023). CGMs are monitors that allow patients to continuously track glucose levels by inserting a needle with a sensor in the body, usually in the upper arm. This sensor is connected to an external device or an app on the phone, which can send a notification in case of abnormal blood glucose values.

the absence of a detectable policy effect on this outcome. In contrast, glucose testing strips are disposable and intended for single use. The observed increase in purchases of testing strips, therefore, suggests that individuals are testing their glucose levels more frequently.

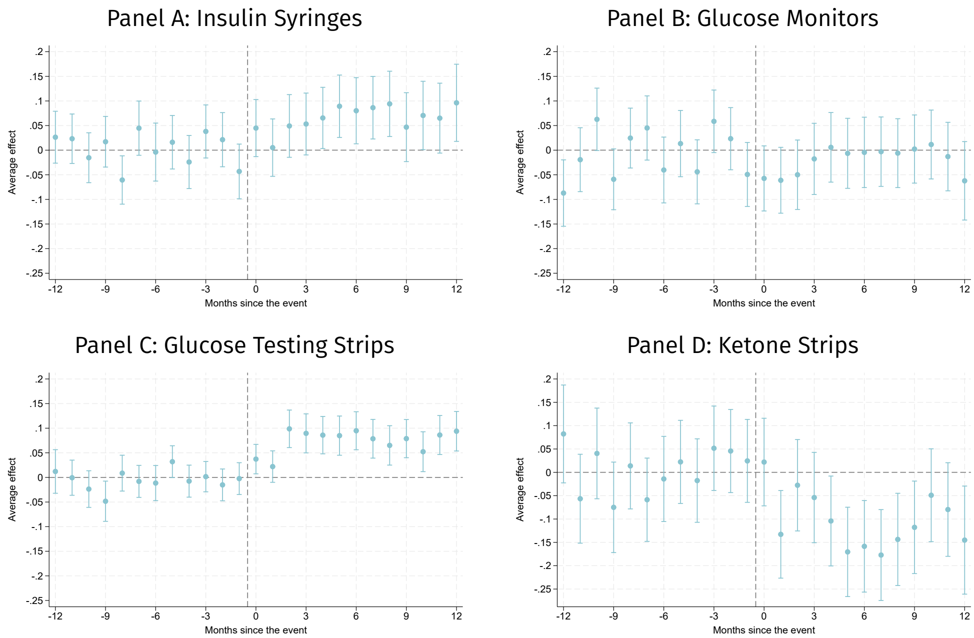
Taken together, the results show an increase in insulin syringe sales—suggesting either an increase in consumption or frequency of injections, or both—, as well as an increase in testing device sales. These findings are consistent with the evidence of ex-ante moral hazard presented in the previous section: lower out-of-pocket insulin costs appear to incentivize greater sugar consumption, with patients managing the short-term consequences by increasing their use of diabetes-related devices. One might argue that, although individuals appear to be shifting from diet-based to medication-based disease management, they might be able to manage well the diseases, through increased injections and blood glucose testing. Although the data do not allow me to rule out this concern fully, I investigate the effect of the policy on a proxy for disease management.

In particular, I study whether there is a decrease in the frequency of spikes in glucose levels for individuals with diabetes. I proxy for this by examining ketone strip sales, which are used to monitor the level of ketones in the blood or urine. High ketone levels can occur when the body burns fat instead of carbohydrates, often due to insufficient insulin. For individuals with diabetes, elevated ketones can signal an increased risk of diabetic ketoacidosis, a potentially life-threatening complication (Puttanna and Padinjakara, 2014; Misra and Oliver, 2015). Medical guidelines recommend ketone testing when blood glucose is high and accompanied by symptoms such as nausea, vomiting, or fever (Albanese-O'Neill et al., 2017).

In Panel D of Figure 1.8, I plot the sales of ketone strips. During the post-period, I observe a decreasing trend, which becomes significant four months after the end of the period. The effect is stable and quite persistent over time. The decrease in sales suggests that individuals are testing their ketone levels less frequently. The delayed effect may stem from a period of adjustment as individuals adjust to their changes in behaviors. These results suggest that, at least in the short-run, the disease is well managed. A limitation is that ketone strips sales capture only very extreme events. Therefore, other outcomes should be consider for a more precise measure of diseases management. In particular, there may be long-term health consequences of repeated blood sugar spikes, which can result from high sugar intake. Such spikes raise hemoglobin A1C levels, which is a key marker of glucose management. High level of hemoglobin A1C are associated with an increased risk of diabetic kidney disease and other complications (Kitada, Kanasaki, and Koya, 2014; Oliveira et al., 2024).

Given that the magnitude of the cap varies across states, the responses may differ depending on the size of the cap. To examine the heterogeneous effects of the insulin out-of-pocket cap, I split the sample into two groups: states with above-median caps and those with below-median caps. For these two groups, Figure 1.A.9 in Appendix 1.A reports the average treatment effect in the post-period for all four devices considered in Figure 1.8. Overall, I find that the policy's effects are stronger in low-cap states, where insulin cost reductions are greater and more individuals are affected. These states exhibit larger increases in insulin syringe, glucose testing strip, and ketone strip purchases, indicating more frequent insulin use and monitoring. As in the main analysis, glucose monitor sales remain unchanged.

Figure 1.8. Units Sold of Diabetes Devices



Notes: The figure reports the event study estimates and the 95 percent confidence intervals using the Callaway and Sant'Anna (2021) estimator. The outcome variables are the inverse hyperbolic sine transformation of monthly units sold of insulin syringes (Panel A), glucose monitors (Panel B), glucose testing strips (Panel C), and ketone testing strips (Panel D). The observation is at the store-month level. Source of the data in the NielsenIQ retail-level scanner data.

Finally, one potential concern is that, for these analyses, I use the overall sales, which include all consumers. However, the products considered—such as insulin syringes, glucose monitors, and testing strips—are specific to diabetes management and are not typically used for other conditions, which mitigates this concern. A related limitation is that I cannot distinguish privately insured individuals (who are directly affected by the policy) from those with other forms of insurance. This limitation raises concerns that changes in diabetes device sales may reflect unrelated price changes rather than being driven by the policy. To address this, Figure 1.A.10 in Appendix 1.A reports the results on prices. The results show no statistically significant post-policy price changes, except for glucose testing strips and ketone strips, where prices begin to decline in the fifth month after the policy implementation. However, the increase in testing strip sales is evident from the first month after implementation and grows further by the third month, suggesting that price changes are unlikely to be the primary driver. If the price were driving the results of ketone strips, a lower price should increase the demand for strips. Therefore, I can exclude that a change in price drives the sales effect.

1.5.3 Robustness Checks

In this subsection, I test and show that my analysis is robust to different specifications. I provide four sets of robustness checks, namely on sample section, confounders from pandemic conditions and policy rollout differences, a placebo test, and alternative methodology.

First of all, I consider different sample selections. I start by addressing a limitation of the NCP data, namely that food purchases are observed at the household level, not at the individual one. Thus, I may underestimate the effect if only the household member with diabetes changes their behavior. To address this concern, I restrict the analysis to single-person households. As shown in Panel A of Figures 1.A.11 and 1.A.12, the estimated effects for this sample (dark green circles) are larger than the baseline (gray triangles), suggesting that household-level analysis likely understates the impact. However, this approach reduces the sample size, which leads to wider confidence intervals. Additionally, I show that the results are robust to restricting the sample to households that participate in the Ailment Survey, for whom diabetes and insurance status are reported rather than imputed. Panel B of Figures 1.A.11 and 1.A.13 shows that there are no differences. The significance and magnitude are almost unchanged, as can be seen by comparing ailment only (dark green circles) to the baseline results (gray triangles).

Second, I address concerns regarding the COVID-19 pandemic and differences in policy rollout. One potential concern is that treated states may differ systematically from control states, particularly in their responses to COVID-19, which could confound the estimated policy effects. I observe that states in the control group were less likely to implement stay-at-home policies. Such policies, in turn, could affect food purchasing behavior. Therefore, as a robustness check, I exclude the states without stay-at-home policies.[31] In Panel A of Figure 1.A.14 and in Figure 1.A.15, I show that the results are robust to this exclusion. In addition, the pandemic caused temporary layoffs and stimulus checks between 2020 and 2021. I report the timing of the shocks in Table 1.A.5. To account for these shocks, I exclude the three states—i.e., Colorado, Minnesota, and New Hampshire—that reduced insulin out-of-pocket costs in 2020. I find similar results, which I report in Panel B of Figure 1.A.14 and in Figure 1.A.16. Finally, I also show that the analysis is robust to the selection of included states. In Panel C of Figure 1.A.14, I exclude Connecticut, given that this state's legislation also includes diabetes devices for the cap calculation. In Panel D, I exclude Maryland, North Dakota, and Nebraska from the analysis, as they implemented the policy in January 2023 and January 2024, respectively. In these robustness checks, the results are virtually unchanged. This is also the case for fibers and sugar, as shown in Figures 1.A.17, and 1.A.18.

Third, I also run a placebo analysis to show that the results are explicitly driven by the policy and not by other confounding factors, some of which are discussed in the previous paragraph. I exploit the characteristics of the policy to perform the placebo test. I run the analysis on households with diabetes and Medicare. These households serve as an appropriate placebo group, as they were similarly exposed to other potential confounders, such as the COVID-19 pandemic, but were not subject to the insulin cap. I replicate the main

[31] I exclude Utah, Wyoming, North Dakota, South Dakota, Nebraska, Oklahoma, Arkansas, and Iowa.

analysis on this subgroup and report the results in Figure 1.A.19 in Appendix 1.A. I find no policy effect among Medicare households, supporting the identification strategy.

Fourth, I show that my results are robust to alternative difference-in-differences specifications. In particular, I employ the standard two-way fixed effects approach and the approach proposed by Sun and Abraham (2021). The latter method and Callaway and Sant'Anna (2021) are similar in accounting for differential timing in treatment. There is one main difference that could be marginally relevant in my context. Callaway and Sant'Anna (2021) can be used if the parallel trends assumption only holds after conditioning on observables. Even if, in my main specifications, I do control for the share of Democratic votes at the state level, the parallel trends also hold without the covariates. Moreover, the individual fixed effects capture most of the yearly demographics I observe. Therefore, the two methods should yield similar results. Figure 1.A.20 in Appendix 1.A replicates the carbohydrate purchase results and shows similar results for all three methods considered.

1.6 Conclusion

This paper studies the role of ex-ante moral hazard in health insurance at the intensive margin, i.e., when the out-of-pocket cost changes. In a staggered difference-in-differences approach à la Callaway and Sant'Anna (2021), I leverage the exogenous variation in out-of-pocket costs provided by the staggered implementation of policies by 23 U.S. states. These policies reduced the out-of-pocket costs for insulin for insured individuals with diabetes, the focus of this study. I use grocery purchase data from 2019 and 2022, combined with the nutritional content and households' health status information.

My findings address important gaps in the economics literature on health insurance and risky behaviors in three ways. First, I can evaluate changes in health-related nutrient purchases to assess the degree of risky behaviors. Second, I focus my analysis on individuals who have chronic conditions, in particular diabetes. Given the central role that both lifestyle and medication play in disease management, this group is particularly relevant for studying moral hazard. Moreover, the distinction between type I and type II diabetes further allows for heterogeneity analysis, which shows that those with more behavioral discretion (type II) are more responsive to cost changes. Third, I contribute to the limited literature on ex-ante moral hazard at the intensive margin, moving beyond the commonly studied extensive margin of insurance coverage. Together, these contributions offer a more nuanced and comprehensive understanding of how cost-sharing reforms affect risky behaviors.

I find that lowering out-of-pocket costs can lead to an increase in carbohydrate (4.8%) and sugar (9.3%) purchases. This result is primarily driven by households with type II diabetes, who have greater discretion over their lifestyle choices and disease management. I also find evidence that certain groups—such as those giving lower importance to low-carbohydrate diets and those with lower incomes—are also more responsive to the policy. These results are consistent with ex-ante moral hazard. Particularly, when financial exposure to medical costs declines, individuals may reduce their effort in maintaining a healthy diet. The increase in sugar consumption and greater reliance on medical supplies for diabetes management indicate a behavioral shift from lifestyle management toward treatment. Importantly, I rule out alternative explanations, such as income effects, which support the interpretation of a moral hazard response.

In recent years, policies in the U.S. have been implemented to reduce out-of-pocket costs for patients, such as the U.S. Inflation Reduction Act. At a time of high healthcare spending, these policies aim to improve affordability. However, the findings from this work emphasize the importance of understanding and considering behavioral responses to the policy. Indeed, to achieve the full health benefits that may be attained through improved access, it may be necessary to implement complementary policies, such as nutritional education, for those who are more likely to engage in risky behaviors. These findings are relevant for chronic diseases, such as diabetes or high cholesterol. For these diseases, a combination of medication and lifestyle changes is necessary for effectively managing the disease. Therefore, understanding how price changes impact incentives is crucial in order to guarantee both access and the long-term financial sustainability of healthcare systems as new drugs continue to emerge on the market.

Finally, because of the nature of the available data, this paper has two limitations. First, I cannot directly link individual food purchases to diabetes device usage, as the latter data are frequent only at the retailer level. This data limitation restricts my ability to establish a clear connection between dietary behavior and disease management practices. Second, I am unable to assess the long-term health outcomes of the policy, which prevents a comprehensive welfare analysis from interpreting the adverse effects of ex-ante moral hazard. Nevertheless, this work provides an initial step in documenting the role of ex-ante moral hazard through behavioral reactions to lower insulin expenses. Future studies should better link dietary choices and disease management, as well as quantify the welfare implications of the policy.

References

ADA. 2018. “Economic costs of diabetes in the US in 2017.” *Diabetes Care* 41 (5): 917–28. [4]

ADA. 2022. *Support for INSULIN Act at Senate Press Conference.* https://diabetes.org/newsroom/american-diabetes-association-announces-support-for-insulin-act-at-senate-press-conference. Accessed: February 27, 2024. [8]

ADA. 2023. *About Diabetes.* https://diabetes.org/about-diabetes/statistics/about-diabetes. ADA. Accessed: November 26, 2023. [4]

Albanese-O’Neill, Anastasia, Mengdi Wu, Kellee M Miller, Laura Jacobsen, Michael J Haller, and Desmond Schatz. 2017. “Poor adherence to ketone testing in patients with type 1 diabetes.” *Diabetes Care* 40 (4): e38–e39. [29]

Americo, Pedro, and Rudi Rocha. 2020. “Subsidizing access to prescription drugs and health outcomes: The case of diabetes.” *Journal of Health Economics* 72: 102347. [6]

Anderson, Kelly E, Nathorn Chaiyakunapruk, Eric J Gutierrez, H Weston Schmutz, Michael R Rose, Diana Brixner, and R Brett McQueen. 2024. “State Out-Of-Pocket Caps On Insulin Costs: No Significant Increase In Claims Or Utilization: Article examines impact of state out-of-pocket caps on insulin costs.” *Health Affairs* 43 (8): 1137–46. [6]

Bakkila, Baylee F, Sanjay Basu, and Kasia J Lipska. 2022. “Catastrophic Spending On Insulin In The United States, 2017–18.” *Health Affairs* 41 (7): 1053–60. [9]

Bao, Ying, Matthew Osborne, Emily Yucai Wang, and Edward C Jaenicke. 2020. “BMI, Food Purchase, and Promotional Sensitivity.” *Available at SSRN 3260896.* [7]

Barahona, Nano, Cristóbal Otero, and Sebastián Otero. 2023. “Equilibrium effects of food labeling policies.” *Econometrica* 91 (3): 839–68. [7]

Barkley, Aaron. 2023. “The human cost of collusion: Health effects of a Mexican insulin cartel.” *Journal of the European Economic Association.* [6, 7]

Barros, Pedro Pita, Matilde P Machado, and Anna Sanz-de-Galdeano. 2008. “Moral hazard and the demand for health services: a matching estimator approach.” *Journal of Health Economics* 27 (4): 1006–25. [6]

Belot, Michèle, Jonathan James, and Jonathan Spiteri. 2020. “Facilitating healthy dietary habits: An experiment with a low income population.” *European Economic Review* 129: 103550. [7]

Brook, Robert H, John E Ware Jr, William H Rogers, Emmett B Keeler, Allyson R Davies, Cathy A Donald, George A Goldberg, Kathleen N Lohr, Patricia C Masthay, and Joseph P Newhouse. 1983. “Does free care improve adults’ health? Results from a randomized controlled trial.” *New England Journal of Medicine* 309 (23): 1426–34. [6]

Brot-Goldberg, Zarek C, Amitabh Chandra, Benjamin R Handel, and Jonathan T Kolstad. 2017. “What does a deductible do? The impact of cost-sharing on health care prices, quantities, and spending dynamics.” *Quarterly Journal of Economics* 132 (3): 1261–318. [3]

Bureau of Labor Statistics. 2020. *Consumer Expenditures — 2019.* https://www.bls.gov/opub/reports/consumer-expenditures/2019/. Accessed: 2023-11-04. [14]

Callaway, Brantly, and Pedro HC Sant’Anna. 2021. “Difference-in-differences with multiple time periods.” *Journal of Econometrics* 225 (2): 200–230. [4, 16, 18, 20–23, 26, 27, 30, 32, 39–48, 52]

Cawley, John, David Frisvold, Anna Hill, and David Jones. 2019. “The impact of the Philadelphia beverage tax on purchases and consumption by adults and children.” *Journal of Health Economics* 67: 102225. [7]

CDC. 2023. *Carbohydrate Counting to Manage Blood Sugar.* https://www.cdc.gov/diabetes/healthy-eating/carb-counting-manage-blood-sugar.html. Accessed: 2024-09-12. [5]

Chandra, Amitabh, Evan Flack, and Ziad Obermeyer. 2024. “The health costs of cost sharing.” *Quarterly Journal of Economics* 139 (4): 2037–82. [7]

Chandra, Amitabh, Jonathan Gruber, and Robin McKnight. 2010. “Patient cost sharing in low income populations.” *American Economic Review* 100 (2): 303–8. [6]

Chandra, Amitabh, Jonathan Gruber, and Robin McKnight. 2014. “The impact of patient cost-sharing on low-income populations: evidence from Massachusetts.” *Journal of Health Economics* 33: 57–66. [6, 7]

Chen, Chen, Gordon Guoen Liu, Tangxin Wang, and Jialong Tan. 2023. “Ex-ante moral hazard and health insurance: Evidence from China’s urban residence basic medical insurance scheme.” *Health Economics* 32 (11): 2516–34. [6]

Chua, Kao-Ping, Joyce M Lee, and Rena M Conti. 2020. “Out-of-pocket spending for insulin, diabetes-related supplies, and other health care services among privately insured US patients with type 1 diabetes.” *JAMA Internal Medicine* 180 (7): 1012–14. [14]

Cleveland Clinic. 2023. *How Many Calories a Day Should I Eat?* https://health.clevelandclinic.org/how-many-calories-a-day-should-i-eat. Accessed: 2023-11-04. [14]

CMS. 2023. *National Health Expenditure Projections 2022-2031: Forecast Summary and Overview.* https://www.cms.gov/files/document/highlights.pdf. Accessed: 2024-10-06. [3]

Cotti, Chad, Erik Nesson, and Nathan Tefft. 2019. “Impacts of the ACA Medicaid expansion on health behaviors: evidence from household panel data.” *Health Economics* 28 (2): 219–44. [6]

Dave, Dhaval, and Robert Kaestner. 2009. “Health insurance and ex ante moral hazard: evidence from Medicare.” *International Journal of Health Care Finance and Economics* 9: 367–90. [6]

Dubois, Pierre, Rachel Griffith, and Aviv Nevo. 2014. "Do prices and attributes explain international differences in food purchases?" *American Economic Review* 104 (3): 832–67. [7]

Dubois, Pierre, Rachel Griffith, and Martin O'Connell. 2020. "How well targeted are soda taxes?" *American Economic Review* 110 (11): 3661–704. [7]

Dunn, Abe, and Adam Hale Shapiro. 2019. "Does medicare part D save lives?" *American Journal of Health Economics* 5 (1): 126–64. [7]

Ehrlich, Isaac, and Gary S Becker. 1972. "Market insurance, self-insurance, and self-protection." *Journal of Political Economy* 80 (4): 623–48. [3]

Einav, Liran, and Amy Finkelstein. 2018. "Moral hazard in health insurance: what we know and how we know it." *Journal of the European Economic Association* 16 (4): 957–82. [4]

Einav, Liran, Ephraim Leibtag, and Aviv Nevo. 2010. "Recording discrepancies in Nielsen Homescan data: Are they present and do they matter?" *Quantitative Marketing and Economics* 8: 207–39. [12]

Endocrine News. 2014. *Grudge Match: Pens vs. Syringes.* https://endocrinenews.endocrine.org/april-2014-grudge-match-pens-vs-syringes/. Accessed: 2024-05-29. [15]

Fernandez, Maria Luz, and Catherine J Andersen. 2014. "Effects of dietary cholesterol in diabetes and cardiovascular disease." *Clinical Lipidology and Metabolic Disorders* 9 (6): 607. [9]

Fiorio, Carlo V, and Luigi Siciliani. 2010. "Co-payments and the demand for pharmaceuticals: evidence from Italy." *Economic Modelling* 27 (4): 835–41. [6]

Fuchs, Flávio D, and Paul K Whelton. 2020. "High blood pressure and cardiovascular disease." *Hypertension* 75 (2): 285–92. [9]

Garabedian, Laura F, Fang Zhang, Rebecca Costa, Stephanie Argetsinger, Dennis Ross-Degnan, and J Frank Wharam. 2024. "Association of State Insulin Out-of-Pocket Caps With Insulin Cost-Sharing and Use Among Commercially Insured Patients With Diabetes: A Pre–Post Study With a Control Group." *Annals of Internal Medicine* 177 (4): 439–48. [6]

Gaynor, Martin, Jian Li, and William B Vogt. 2007. "Substitution, spending offsets, and prescription drug benefit design." In *Forum for Health Economics & Policy,* vol. 10. 2. De Gruyter. [6]

Giannouchos, Theodoros V, Benjamin Ukert, and Thomas Buchmueller. 2024. "Health outcome changes in individuals with type 1 diabetes after a state-level insulin copayment cap." *JAMA Network Open* 7 (8): e2425280–e2425280. [6, 9]

Goldman, Dana P, Geoffrey F Joyce, Jose J Escarce, Jennifer E Pace, Matthew D Solomon, Marianne Laouri, Pamela B Landsman, and Steven M Teutsch. 2004. "Pharmacy benefits and the use of drugs by the chronically ill." *JAMA* 291 (19): 2344–50. [11]

Gračner, Tadeja. 2021. "Bittersweet: How prices of sugar-rich foods contribute to the diet-related disease epidemic in Mexico." *Journal of Health Economics* 80: 102506. [7]

Griffith, Rachel, Stephanie von Hinke, and Sarah Smith. 2018. "Getting a healthy start: The effectiveness of targeted benefits for improving dietary choices." *Journal of Health Economics* 58: 176–87. [7]

Grillo, Andrea, Lucia Salvi, Paolo Coruzzi, Paolo Salvi, and Gianfranco Parati. 2019. "Sodium intake and hypertension." *Nutrients* 11 (9): 1970. [9]

Health Care Cost Institute. 2024. *Capping Out-of-Pocket Spending on Insulin Would Lower Costs for a Substantial Proportion of Commercially Insured Individuals.* https://healthcostinstitute.org/hcci-originals-dropdown/all-hcci-reports/capping-out-of-pocket-spending-on-insulin-would-lower-costs-for-a-substantial-proportion-of-commercially-insured-individuals-1. Accessed: 2024-05-29. [9, 39]

Herkert, Darby, Pavithra Vijayakumar, Jing Luo, Jeremy I Schwartz, Tracy L Rabin, Eunice DeFilippo, and Kasia J Lipska. 2019. "Cost-related insulin underuse among patients with diabetes." *JAMA Internal Medicine* 179 (1): 112–14. [10]

Hinnosaar, Marit. 2023. "The persistence of healthy behaviors in food purchasing." *Marketing Science* 42 (3): 521–37. [7]

Hirsch, Irl B. 2016. "Insulin in America: A right or a privilege?" *Diabetes Spectrum* 29(3). [4]

Holman, Rury R, Sanjoy K Paul, M Angelyn Bethel, David R Matthews, and H Andrew W Neil. 2008. "10-year follow-up of intensive glucose control in type 2 diabetes." *New England Journal of Medicine* 359 (15): 1577–89. [5, 9, 21]

Huh, Jason, and Julian Reif. 2017. "Did medicare part d reduce mortality?" *Journal of Health Economics* 53: 17–37. [7]

Hut, Stefan. 2020. "Determinants of dietary choice in the US: Evidence from consumer migration." *Journal of Health Economics* 72: 102327. [7]

Hut, Stefan, and Emily Oster. 2022. "Changes in household diet: Determinants and predictability." *Journal of Public Economics* 208: 104620. [7, 12, 39, 50]

Iglay, Kristy, Hakima Hannachi, Patrick Joseph Howie, Jinfei Xu, Xueying Li, Samuel S Engel, Lori M Moore, and Swapnil Rajpathak. 2016. "Prevalence and co-prevalence of comorbidities among patients with type 2 diabetes mellitus." *Current Medical Research and Opinion* 32 (7): 1243–52. [8, 9]

Imamura, Fumiaki, Renata Micha, Jason HY Wu, Marcia C de Oliveira Otto, Fadar O Otite, Ajibola I Abioye, and Dariush Mozaffarian. 2016. "Effects of saturated fat, polyunsaturated fat, monounsaturated fat, and carbohydrate on glucose-insulin homeostasis: a systematic review and meta-analysis of randomised controlled feeding trials." *PLoS Medicine* 13 (7): e1002087. [9, 24]

Imbens, Guido W, and Donald B Rubin. 2015. *Causal inference in statistics, social, and biomedical sciences.* Cambridge university press. [55]

Khin, Phyu Phyu, Jong Han Lee, and Hee-Sook Jun. 2023. "Pancreatic beta-cell dysfunction in type 2 diabetes." *European Journal of Inflammation* 21: 1721727X231154152. [8]

Kitada, Munehiro, Keizo Kanasaki, and Daisuke Koya. 2014. "Clinical therapeutic strategies for early stage of diabetic kidney disease." *World Journal of Diabetes* 5 (3): 342. [5, 29]

Landsberg, Lewis, and Mark Molitch. 2004. "Diabetes and hypertension: pathogenesis, prevention and treatment." *Clinical and Experimental Hypertension* 26 (7-8): 621–28. [9]

Larsen, RN, NJ Mann, E Maclean, and JE Shaw. 2011. "The effect of high-protein, low-carbohydrate diets in the treatment of type 2 diabetes: a 12 month randomised controlled trial." *Diabetologia* 54: 731–40. [23]

Ley, Sylvia H, Osama Hamdy, Viswanathan Mohan, and Frank B Hu. 2014. "Prevention and management of type 2 diabetes: dietary components and nutritional strategies." *Lancet* 383 (9933): 1999–2007. [8, 9, 24]

Look AHEAD Research Group. 2007. "Reduction in weight and cardiovascular disease risk factors in individuals with type 2 diabetes: One-year results of the Look AHEAD trial." *Diabetes Care* 30 (6): 1374–83. https://doi.org/10.2337/dc07-0048. [8]

Lozano-Rojas, Felipe, and Patrick Carlin. 2022. "The effect of soda taxes beyond beverages in Philadelphia." *Health Economics* 31 (11): 2381–410. [7]

Malik, Vasanti S, Barry M Popkin, George A Bray, Jean-Pierre Després, Walter C Willett, and Frank B Hu. 2010. "Sugar-sweetened beverages and risk of metabolic syndrome and type 2 diabetes: a meta-analysis." *Diabetes Care* 33 (11): 2477–83. [21]

Martínez-Jiménez, Mario, Pilar García-Gómez, and Jaume Puig-Junoy. 2021. "The Effect of Changes in Cost Sharing on the Consumption of Prescription and Over-the-Counter Medicines in Catalonia." *International Journal of Environmental Research and Public Health* 18 (5): 2562. [7]

Mayberry, Lindsay S, Charmin Guy, Chase D Hendrickson, Allison B McCoy, and Tom Elasy. 2023. "Rates and correlates of uptake of continuous glucose monitors among adults with type 2 diabetes in primary care and endocrinology settings." *Journal of General Internal Medicine* 38 (11): 2546–52. [28]

McAdam-Marx, Carrie, Natalia Ruiz-Negron, Jane M Sullivan, and Jamie M Tucker. 2024. "The effects of patient out-of-pocket costs on insulin use among people with type 1 and type 2 diabetes with Medicare Advantage insurance—2014–2018." *Health Services Research* 59 (1): e14152. [11]

McNamara, Cici, and Natalia Serna. 2022. "The impact of a national formulary expansion on diabetics." *Health Economics* 31 (11): 2311–32. [6, 7]

McRae, Marc P. 2018. "Dietary fiber intake and type 2 diabetes mellitus: an umbrella review of meta-analyses." *Journal of Chiropractic Medicine* 17 (1): 44–53. [21]

Misra, S, and NS Oliver. 2015. "Utility of ketone measurement in the prevention, diagnosis and management of diabetic ketoacidosis." *Diabetic Medicine* 32 (1): 14–23. [29]

Newhouse, Joseph P. 1993. *Free for all?: lessons from the RAND health insurance experiment.* Harvard University Press. [6]

OECD. 2019. *Health at a Glance 2019: OECD Indicators.* Paris: OECD Publishing. https://doi.org/10.1787/4dd50c09-en. [3]

Oliveira, Carolina Piras de, Mary Anne Dellva, Juliana Bue-Valleskey, Annette M Chang, and Birong Liao. 2024. "Fasting and postprandial plasma glucose contributions to hemoglobin A1c and time in range in people with diabetes on multiple daily injection insulin therapy: Results from the PRONTO-T1D and PRONTO-T2D clinical trials." *Journal of Diabetes and its Complications* 38 (1): 108648. [5, 29]

Ong, Kanyin Liane, Lauryn K Stafford, Susan A McLaughlin, Edward J Boyko, Stein Emil Vollset, Amanda E Smith, Bronte E Dalton, Joe Duprey, Jessica A Cruz, Hailey Hagins, et al. 2023. "Global, regional, and national burden of diabetes from 1990 to 2021, with projections of prevalence to 2050: a systematic analysis for the Global Burden of Disease Study 2021." *Lancet* 402 (10397): 203–34. [4]

Oster, Emily. 2015. "Diabetes and diet: Behavioral response and the value of health." Working paper. National Bureau of Economic Research. [12, 39, 50]

Oster, Emily. 2018. "Diabetes and diet: Purchasing behavior change in response to health information." *American Economic Journal: Applied Economics* 10 (4): 308–48. [7, 12]

Park, Eunja, Daeeun Kim, and Sookja Choi. 2019. "The impact of differential cost sharing of prescription drugs on the use of primary care clinics among patients with hypertension or diabetes." *Public Health* 173: 105–11. [7]

Pauly, Mark V. 1968. "The economics of moral hazard: comment." *American Economic Review,* 531–37. [3]

Phipps, Michael S, Ania M Jastreboff, Karen Furie, and Walter N Kernan. 2012. "The diagnosis and management of cerebrovascular disease in diabetes." *Current Diabetes Reports* 12: 314–23. [9]

Puig-Junoy, Jaume, Pilar García-Gómez, and David Casado-Marín. 2016. "Free medicines thanks to retirement: impact of coinsurance exemption on pharmaceutical expenditures and hospitalization offsets in a national health service." *Health Economics* 25 (6): 750–67. [6, 7]

Puttanna, Amar, and RNK Padinjakara. 2014. "Diabetic ketoacidosis in type 2 diabetes mellitus." *Practical Diabetes* 31 (4): 155–58. [29]

Quader, Zerleen S, Lixia Zhao, Cathleen Gillespie, Mary E Cogswell, Ana L Terry, Alanna Moshfegh, and Donna Rhodes. 2017. "Sodium Intake Among Persons Aged above 2 Years — United States, 2013–2014." *MMWR. Morbidity And Mortality Weekly Report* 66. [25]

Reynolds, Andrew N, Ashley P Akerman, and Jim Mann. 2020. "Dietary fibre and whole grains in diabetes management: Systematic review and meta-analyses." *PLoS Medicine* 17 (3): e1003053. [21]

Rios-Avila, Fernando, Pedro Sant'Anna, and Brantly Callaway. 2023. "CSDID: Stata module for the estimation of Difference-in-Difference models with multiple time periods." [17]

Robins, James. 1986. "A new approach to causal inference in mortality studies with a sustained exposure period—application to control of the healthy worker survivor effect." *Mathematical Modelling* 7 (9-12): 1393–512. [16]

Sant'Anna, Pedro HC, and Jun Zhao. 2020. "Doubly robust difference-in-differences estimators." *Journal of Econometrics* 219 (1): 101–22. [17]

Schwab, Ursula, Andrew N Reynolds, Taisa Sallinen, Angela Albarosa Rivellese, and Ulf Risérus. 2021. "Dietary fat intakes and cardiovascular disease risk in adults with type 2 diabetes: a systematic review and meta-analysis." *European Journal of Nutrition* 60: 3355–63. [24]

Shankaran, Veena, and Scott Ramsey. 2015. “Addressing the financial burden of cancer treatment: from copay to can’t pay.” *JAMA Oncology* 1 (3): 273–74. [11]

Shigeoka, Hitoshi. 2014. “The effect of patient cost sharing on utilization, health, and risk protection.” *American Economic Review* 104 (7): 2152–84. [6]

Simon, Kosali, Aparna Soni, and John Cawley. 2017. “The impact of health insurance on preventive care and health behaviors: evidence from the first two years of the ACA Medicaid expansions.” *Journal of Policy Analysis and Management* 36 (2): 390–417. [6]

Spenkuch, Jörg L. 2012. “Moral hazard and selection among the poor: Evidence from a randomized experiment.” *Journal of Health Economics* 31 (1): 72–85. [6]

Sun, Liyang, and Sarah Abraham. 2021. “Estimating dynamic treatment effects in event studies with heterogeneous treatment effects.” *Journal of Econometrics* 225 (2): 175–99. [32, 48]

U.S. Census Bureau. 2020. “Health Insurance Coverage in the United States: 2019.” Report P60-271. U.S. Government Publishing Office. Accessed: 2024-11-04. [9]

Ukert, Benjamin D, Theodoros V Giannouchos, and Thomas C Buchmueller. 2024. “Colorado Insulin Copay Cap: Lower Out-Of-Pocket Payments, Increased Prescription Volume And Days’ Supply: Article examines Colorado’s insulin copay cap law.” *Health Affairs* 43 (8): 1147–55. [6, 9]

Vitt, Nicolai, Jonathan James, Michèle Belot, and Martina Vecchi. 2021. “Daily stressors and food choices: A lab experiment with low-SES mothers.” *European Economic Review* 136: 103754. [7]

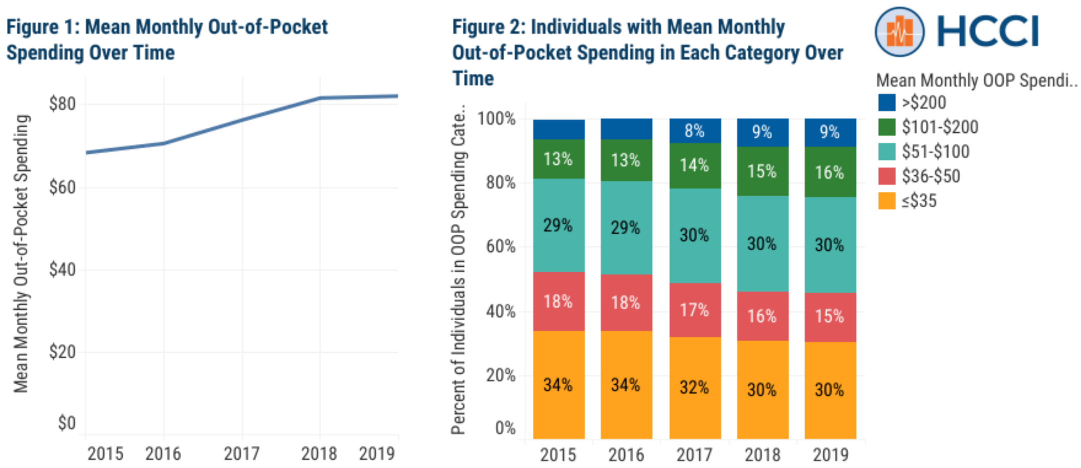
Wilding, JPH4238418. 2014. “The importance of weight management in type 2 diabetes mellitus.” *International Journal of Clinical Practice* 68 (6): 682–91. [24]

Ziebarth, Nicolas R. 2010. “Estimating price elasticities of convalescent care programmes.” *Economic Journal* 120 (545): 816–44. [6]

Appendices to Chapter 1

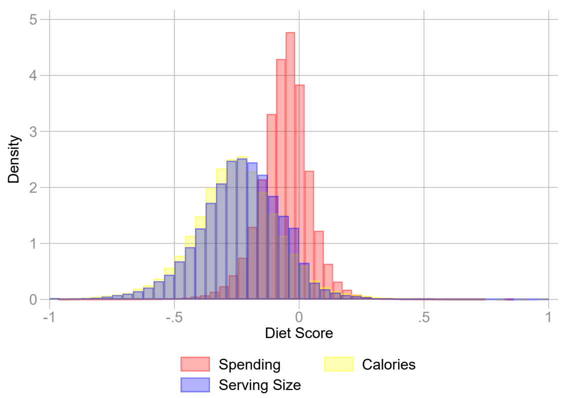
Appendix 1.A Additional Figures and Tables

Figure 1.A.1. Out-of-pocket Spending over the Years



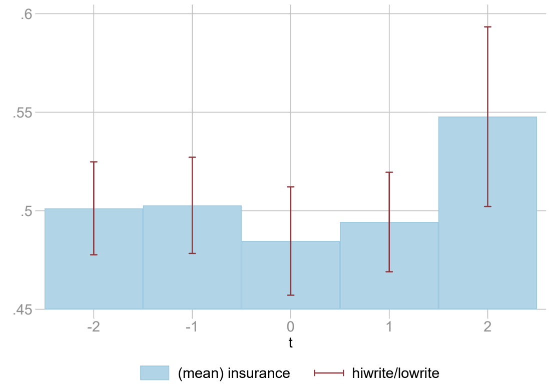
Source: HCCI report, see Health Care Cost Institute (2024).

Figure 1.A.2. Diet Score Density



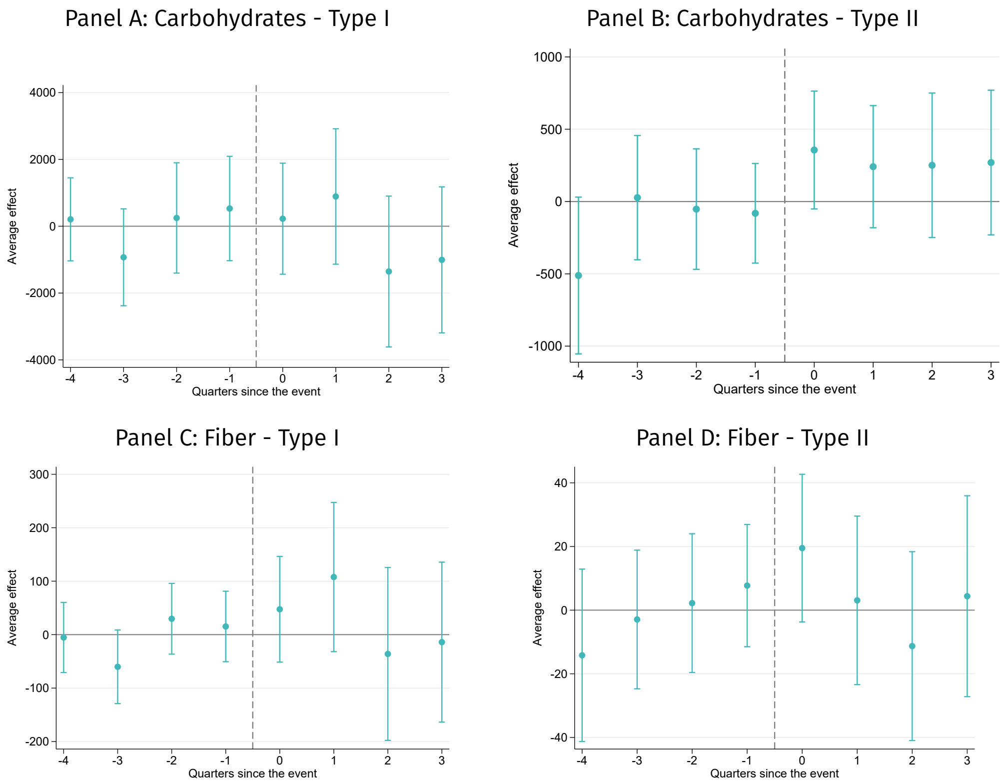
Notes: The figure reports the average post-period event study estimates and the 95 percent confidence intervals using the Callaway and Sant'Anna (2021) estimator. The observation is at the household-quarter level. Source of the data is the NielsenIQ Panel, Oster (2015), and Hut and Oster (2022).

Figure 1.A.3. Employer-sponsored Insurance Share over the Years



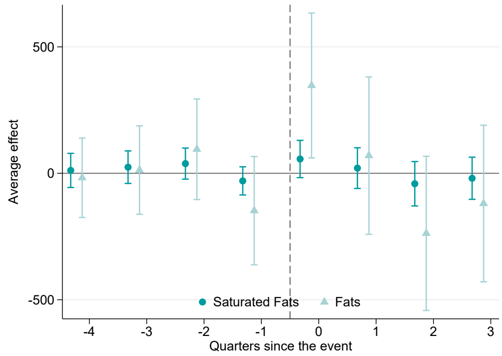
Notes: Own calculation on 2018–2022 KFF insurance count data at the state-year level. This histogram reports the share of individuals having an employer-sponsored insurance (a type of private insurance). The states are stacked around the year zero, i.e., the year in which they implemented the insulin out-of-pocket cost cap. Note that at $t = 2$, only CO, MN and NH.

Figure 1.A.4. Carbohydrate (All and Fibers): Heterogeneity by Diabetes Type



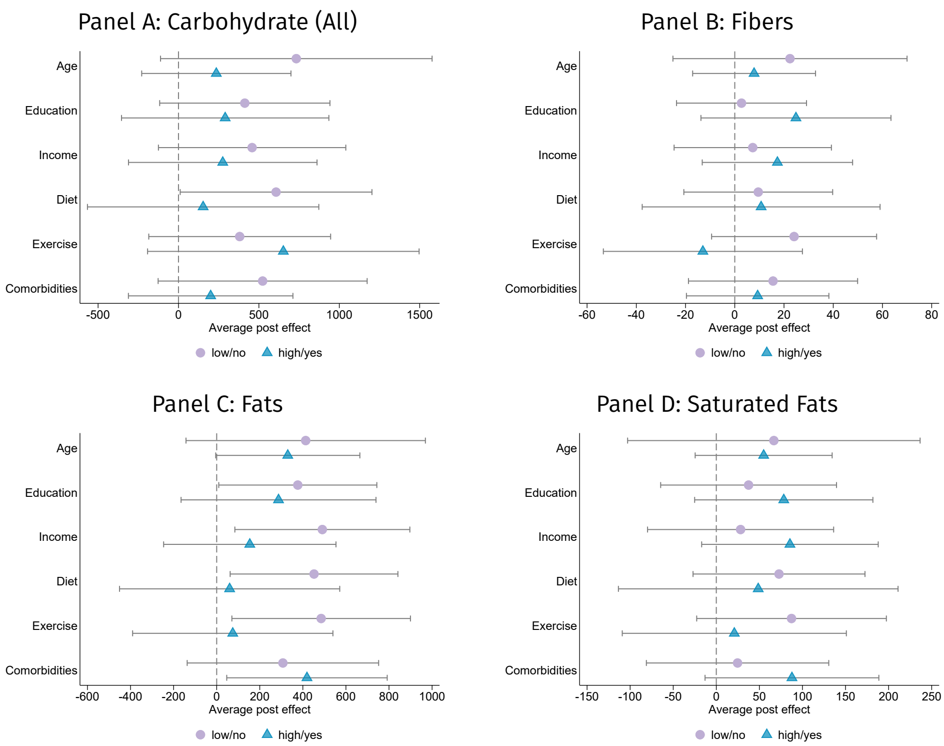
Notes: The figure reports the average post-period event study estimates and the 95 percent confidence intervals using the Callaway and Sant'Anna (2021) estimator. The observation is at the household-quarter level. Source of the data is the NielsenIQ Panel and the Ailment Survey.

Figure 1.A.5. Effect by Fat Types, Purchased Grams



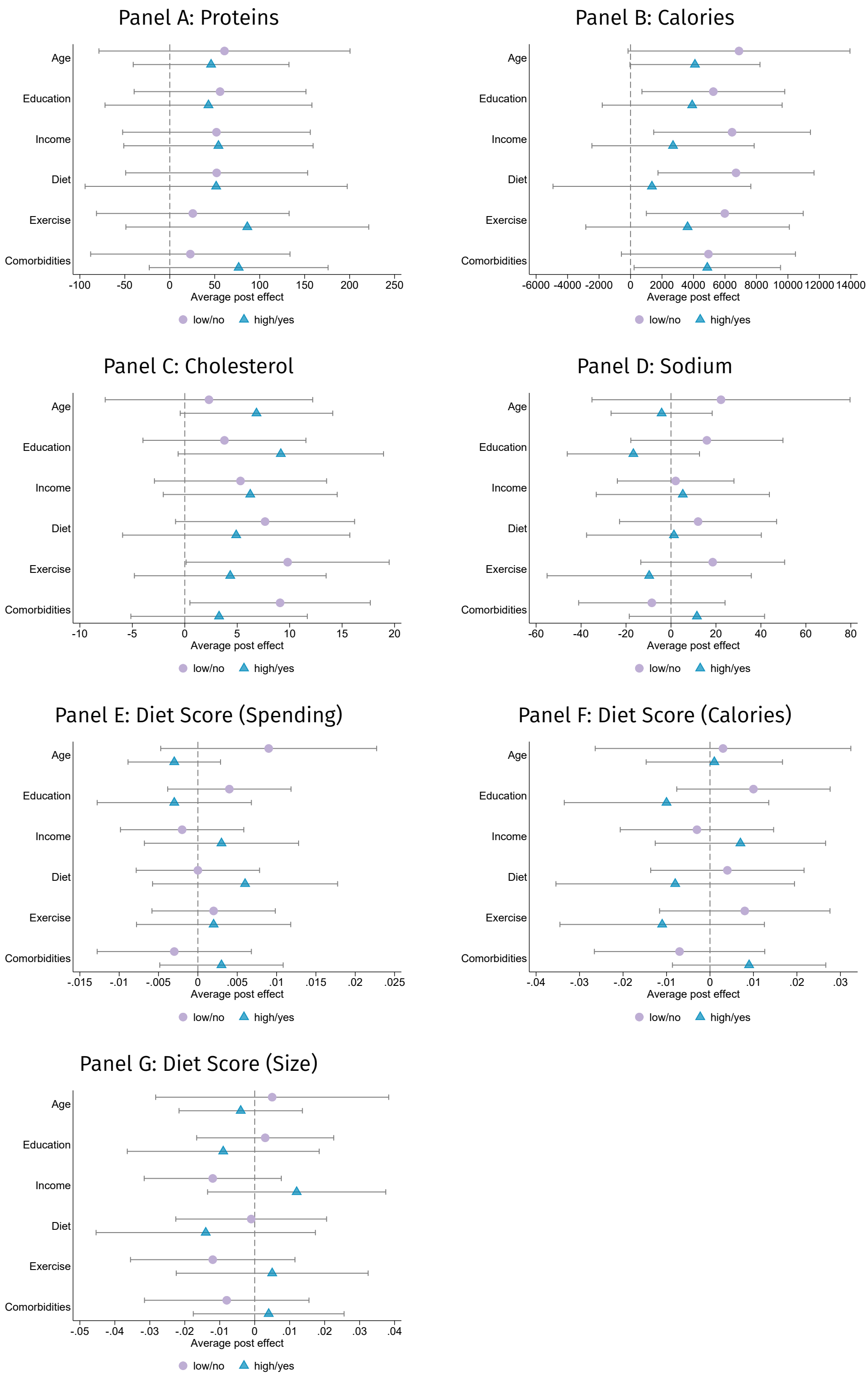
Notes: The figure reports the event study estimates and the 95 percent confidence intervals using the Callaway and Sant'Anna (2021) estimator. Source of the data is the NielsenIQ Panel and the Ailment Survey. The observation is at the household-quarter level. The sample is restricted to households in which at least one member has type II diabetes.

Figure 1.A.6. Nutrients - Heterogeneity



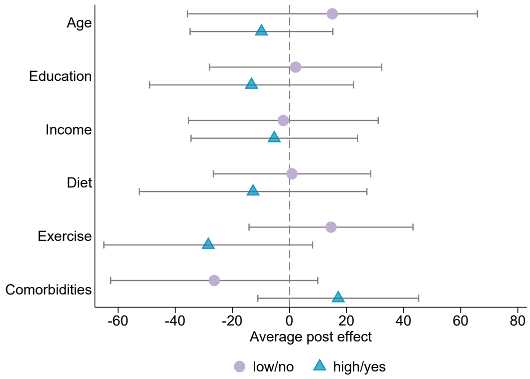
Notes: The figure reports the average post-period event study estimates and the 95 percent confidence intervals using the Callaway and Sant'Anna (2021) estimator. The observation is at the household-quarter level. Source of the data is the NielsenIQ Panel and the Ailment Survey. The sample is restricted to households in which at least one member has type II diabetes.

Figure 1.A.7. Nutrients - Heterogeneity (II)



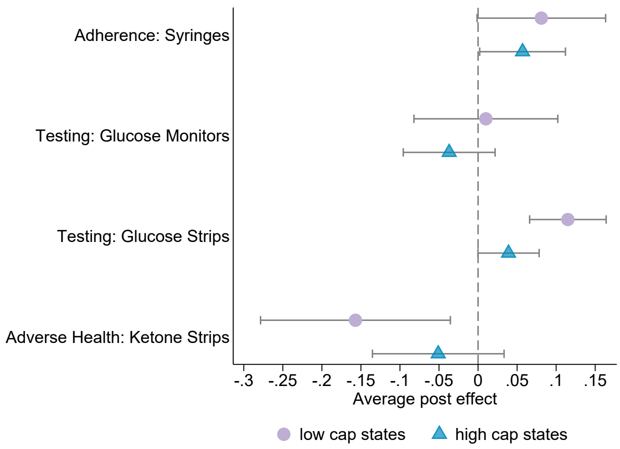
Notes: The figure reports the average post-period event study estimates and the 95 percent confidence intervals using the Callaway and Sant'Anna (2021) estimator. The observation is at the household-quarter level. Source of the data is the NielsenIQ Panel and the Ailment Survey. The sample is restricted to households in which at least one member has type II diabetes.

Figure 1.A.8. Income - Heterogeneity



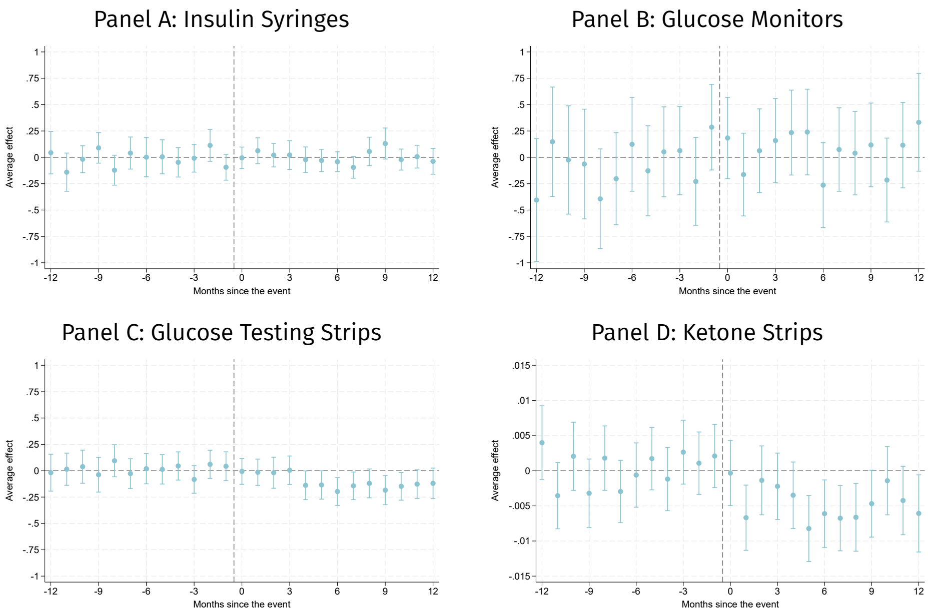
Notes: The figure reports the average post-period event study estimates and the 95 percent confidence intervals using the Callaway and Sant'Anna (2021) estimator. The observation is at the household-quarter level. Source of the data is the NielsenIQ Panel and the Ailment Survey. The sample is restricted to households in which at least one member has type II diabetes.

Figure 1.A.9. Heterogeneity by Cap Size



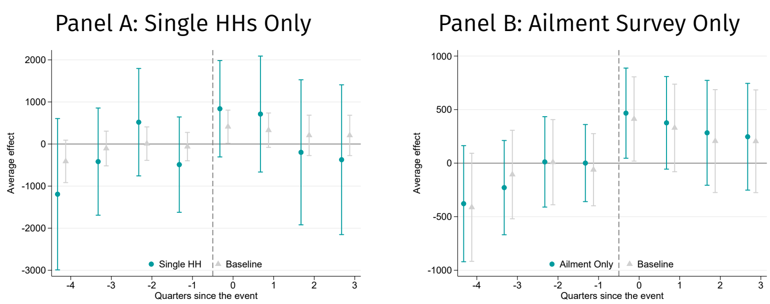
Notes: The figure reports the average post-period event study estimates and the 95 percent confidence intervals using the Callaway and Sant'Anna (2021) estimator. The observation is at the store-month level (drug stores only). Source of the data is the NielsenIQ retail-level scanner data.

Figure 1.A.10. Unit Prices of Diabetes Devices



Notes: The figure reports the event study estimates and the 95 percent confidence bands using the Callaway and Sant'Anna (2021) estimator. The outcome variables are unit prices. The observation is at the store-month level (drug stores only). Source of the data in the NielsenIQ retail-level scanner data.

Figure 1.A.11. Carbohydrate - Household Selection Robustness Checks

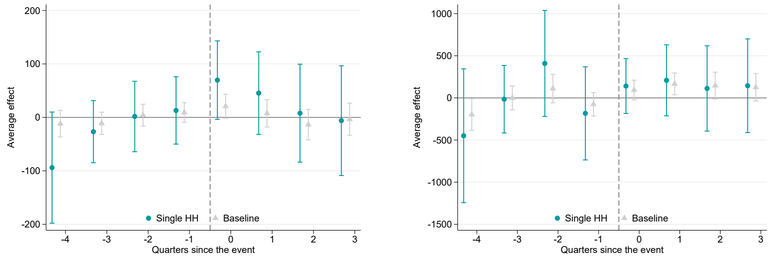


Notes: The figure reports the event study estimates and the 95 percent confidence bands using the Callaway and Sant'Anna (2021) estimator. Source of the data is the NielsenIQ Panel. Panel A restricts the sample to single households only, while Panel B restricts the sample to households both in the NCP and the Ailment Survey.

Figure 1.A.12. Fibers and Sugar - Single HHs Only

Panel A: Fibers

Panel B: Sugar

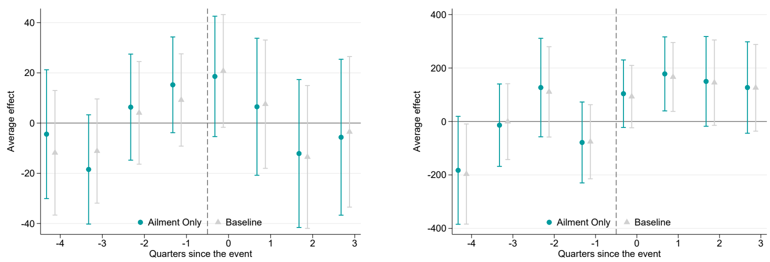


Notes: The figure reports the event study estimates and the 95 percent confidence intervals using the Callaway and Sant'Anna (2021) estimator. The outcome variables the grams purchased on the different nutrients by the household. The observation is at the household-quarter level. Source of the data is the NielsenIQ Panel. I restrict the sample to single households only.

Figure 1.A.13. Fibers and Sugar - Ailment Survey Only

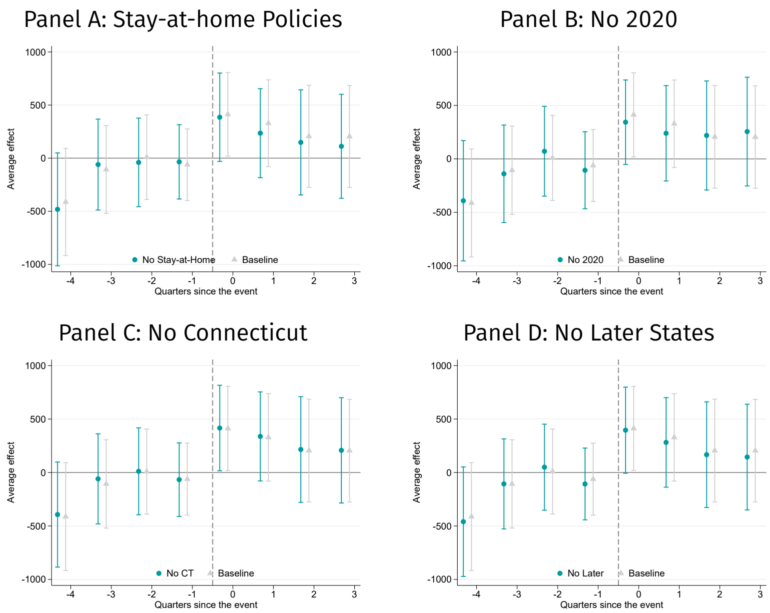
Panel A: Fibers

Panel B: Sugar



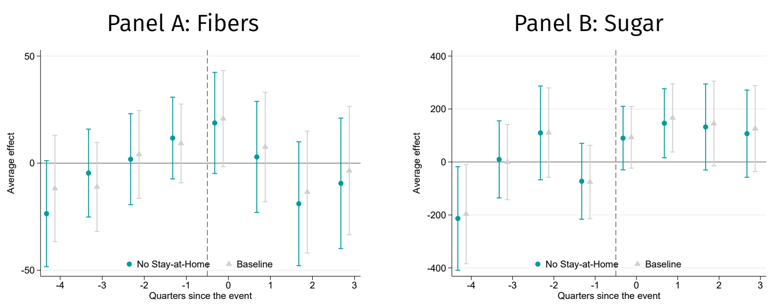
Notes: The figure reports the event study estimates and the 95 percent confidence intervals using the Callaway and Sant'Anna (2021) estimator. The outcome variables the grams purchased on the different nutrients by the household. The observation is at the household-quarter level. Source of the data is the NielsenIQ Panel. I restrict the sample to households both in the NCP and the Ailment Survey.

Figure 1.A.14. Carbohydrate – COVID-19 and Other Confounders Robustness Checks



Notes: The figure reports the event study estimates and the 95 percent confidence intervals using the Callaway and Sant'Anna (2021) estimator. The outcome variables are the grams of carbohydrates per person per quarter. The observation is at the household-quarter level. The source of the data is in the NielsenIQ Panel data. In Panel A, I exclude the states without stay-at-home policies. In Panel B, I exclude the states that implemented the policy in 2020 (i.e., Colorado, Minnesota, and New Hampshire). In Panel C, I exclude Connecticut given the different legislation and in Panel D I exclude the states that were later treated, i.e., Maryland, North Dakota and Nebraska.

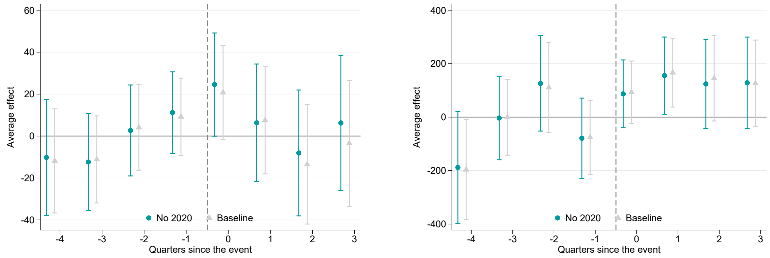
Figure 1.A.15. Fibers and Sugar – Stay-at-home Policies



Notes: The figure reports the event study estimates and the 95 percent confidence intervals using the Callaway and Sant'Anna (2021) estimator. The outcome variables are the grams of carbohydrates per person per quarter. The observation is at the household-quarter level. Source of the data is the NielsenIQ Panel. I exclude the states without stay-at-home policies.

Figure 1.A.16. Fibers and Sugar - No 2020

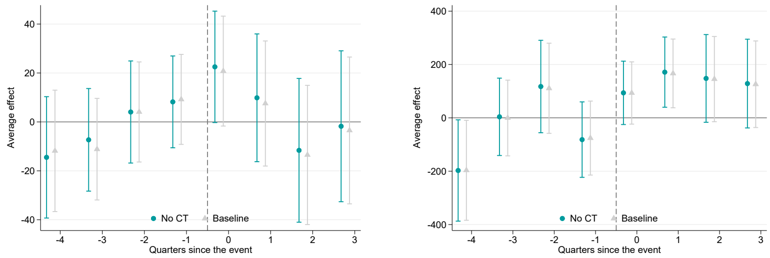
Panel A: Fibers | Panel B: Sugar



Notes: The figure reports the event study estimates and the 95 percent confidence intervals using the Callaway and Sant'Anna (2021) estimator. The outcome variables are the grams of carbohydrates per person per quarter. The observation is at the household-quarter level. Source of the data is the NielsenIQ Panel. I exclude states that implemented the policy in 2020.

Figure 1.A.17. Fibers and Sugar - No Connecticut

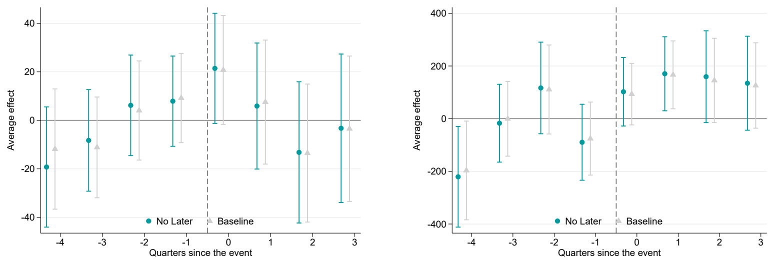
Panel A: Fibers | Panel B: Sugar



Notes: The figure reports the event study estimates and the 95 percent confidence intervals using the Callaway and Sant'Anna (2021) estimator. The outcome variables are the grams of carbohydrates per person per quarter. The observation is at the household-quarter level. The source of the data is in the NielsenIQ Panel data. I exclude Connecticut.

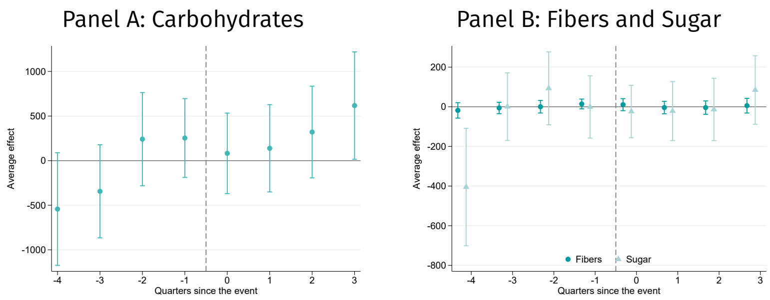
Figure 1.A.18. Fibers and Sugar - No Later States

Panel A: Fibers | Panel B: Sugar



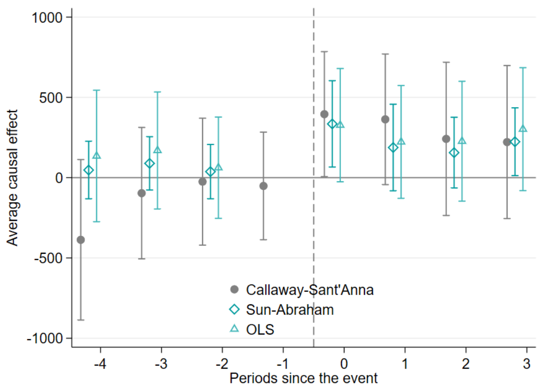
Notes: The figure reports the event study estimates and the 95 percent confidence intervals using the Callaway and Sant'Anna (2021) estimator. The outcome variables are the grams of carbohydrates per person per quarter. The observation is at the household-quarter level. The source of the data is in the NielsenIQ Panel data. I exclude states that were later treated, i.e., Maryland, North Dakota and Nebraska.

Figure 1.A.19. Medicare Placebo



Notes: The figure reports the event study estimates and the 95 percent confidence bands using the Callaway and Sant'Anna (2021) estimator. Source of the data in the NielsenIQ Panel. The figures report falsification test on households with diabetes and Medicare, which were not affected by the insulin cap policy.

Figure 1.A.20. Carbohydrate: Other Methods



Notes: The figure reports the event study estimates and the 95 percent confidence intervals using the Callaway and Sant'Anna (2021) estimator, Sun and Abraham (2021) estimator and the two-way fixed effect (OLS). The outcome variables the grams purchased on the different nutrients by the household. The observation is at the household-quarter level. Source of the data is the NielsenIQ Panel.

Table 1.A.1. State-Level Out-of-pocket Cost Legislation

State	State Code	Diabetes %	Legislation	Date Effective	Date Signed
Alabama	AL	12.9	$100	October 1, 2021	April 13, 2021
Colorado	CO	7.1	$100	January 1, 2020	May 22, 2019
Connecticut	CT	8.2	$25	January 1, 2022	July 1, 2020
Delaware	DE	10.5	$100	January 1, 2021	July 16, 2020
Delaware	DE	10.5	$35	January 1, 2023	October 26, 2022
Illinois	IL	9.3	$100	January 1, 2021	January 24, 2020
Kentucky	KY	11.4	$30	January 1, 2022	March 22, 2021
Louisiana	LA	12.7	$75	August 1, 2022	June 18, 2022
Maine	ME	8.3	$35	January 1, 2021	March 31, 2020
Maryland	MD	9.1	$30	January 1, 2023	May 16, 2022
Minnesota	MN	7.8	$50 (90-day supply)	July 1, 2020	April 15, 2020
Montana	MT	7.8	$35	January 1, 2022	April 29, 2021
Nebraska	NE	8.9	$35	January 1, 2024	September 1, 2023
New Hampshire	NH	7.5	$30	September 14, 2020	July 16, 2020
New Jersey	NJ	8.6	$30	January 1, 2023	June 29, 2022
New Mexico	NM	11.0	$25	January 1, 2021	March 4, 2020
New York	NY	9.1	$100	January 1, 2021	April 3, 2020
North Dakota	ND	9.3	$25	July 1, 2023	April 10, 2022
Oklahoma	OK	11.9	$30	November 1, 2021	April 20, 2021
Oregon	OR	8.4	$75	January 1, 2022	March 31, 2021
Rhode Island	RI	9.0	$40	January 1, 2022	July 7, 2021
Texas	TX	12.1	$25	September 1, 2021	June 14, 2022
Utah	UT	8.5	$30	January 1, 2021	March 30, 2020
Vermont	VT	6.7	$100	January 1, 2022	October 2, 2020
Virginia	VA	9.8	$50	January 1, 2021	April 8, 2020
Washington	WA	7.9	$100	January 1, 2021	March 31, 2020
Washington	WA	7.9	$35	January 1, 2023	March 4, 2022
Washington D.C.	DC	8.3	$30	January 1, 2021	March 31, 2020
West Virginia	WV	13.1	$100	July 1, 2021	March 6, 2020

Notes: Column 3 report diabetes state-level prevalence percentages. Source is the KFF and the Behavioral Risk Factor Surveillance System (BRFS). Columns 4-6 report details on the state-level out-of-pocket cost legislation. In particular, column 4 reports the cap on the out-of-pocket cost for insulin, column 5 when such cap became effective and column 6 when the bill was signed. Columns 4-6 are based on own data collection from each state House Bills.

Table 1.A.2. Food Categories and Their Healthiness

Food Category	"Good"	"Bad"	Food Category	"Good"	"Bad"
Panel A: Unhealthier Food Categories					
Ice Cream	0	17	Potato Chips	1	16
Flavored Syrup	0	17	Jam	1	16
Cake Mix	0	17	Salad Dressing	1	15
Cookies	0	17	Pasta Dinner	1	15
Soda	0	17	Snack Crackers	1	14
Frozen Biscuits	0	17	Margarine	1	12
Frozen Pizza	0	17	Bread	1	12
Cookie mix	0	17	Juice	2	13
Slice-n-Bake Cookies	0	17	Flour	2	8
Chocolate Chips	0	17	Regular Milk	3	11
Sugar	0	17	Potatoes	3	9
Mayonnaise	0	16	Applesauce	3	8
Butter	0	14	Pasta	4	9
Spam	0	14	Rice	4	9
Creamer	0	12	Pretzels	4	9
Panel B: Healthier Food Categories					
Pickles	4	3	Lite Dressing	4	3
Cold Cereal	7	4	Olives	7	4
Canned Vegetables	8	3	Ground Beef	9	6
Canned Beans	9	5	Soup	9	5
Frozen Fruit	9	5	Natural Cheese	10	5
Breakfast Bars	10	4	Salsa	10	3
Olive Oil	11	0	Peanut Butter	12	3
Dried Fruit	12	3	Tuna	12	1
Cottage cheese	13	2	Eggs	13	1
Frozen Vegetables	14	1	Yogurt	14	0
Shrimp	15	1	Hot Cereal	15	0
Fresh Fruit	15	0	Chicken	16	0
Fish	17	0	Nuts	17	0
Low Fat Milk	17	0	Dried Beans	17	0
Vegetables	17	0			

Notes: Source of the data: Oster (2015) and Hut and Oster (2022). For each food category, the number of doctors who rated the product as a good or bad source of calories for diabetic patients is reported. For readability, Panel A reports unhealthier foods, i.e., the majority of doctors defined the food as a bad source of calories. For Panel B, healthier foods are reported. Here, the major of doctor rated the product as a "good source of calories".

Table 1.A.3. NCP Descriptives

	All HH		HH with Diab. & Ins.	
	Mean	Median	Mean	Median
Panel A: Household (HH) Characteristics				
HH Heads Age Heads	53.97	55.00	60.87	62.00
HH Size	2.47	2.00	2.36	2.00
HH Heads Education (years)	14.50	14.00	14.41	14.00
HH Income	61,467	65,000	66,409	65,000
HH Income (per person)	30,060	25,000	33,064	28,333
Married [N/Y]	0.63	.	0.72	.
Single HH [N/Y]	0.24	.	0.17	.
Presence of Children <6 yrs [N/Y]	0.04	.	0.01	.
White [N/Y]	0.78	.	0.78	.
Diabetes Share [N/Y]	0.10	.	1.00	.
N	83,401		8,651	
Panel B: Purchasing Behavior				
General Behavior				
Total Spent $/Quarter	1,057	900	1,256	1,089
Total Spent $/Quarter (per person)	533	441	617	525
Nutrients and Calories				
Carbohydrate	2,972	2,399	3,146	2,612
Fibers	194	141	205	153
Sugar	494	307	510	331
Fats	1,527	1,005	1,657	1,152
Proteins	831	660	899	732
Calories	28,596	22,175	30,668	24,587
Diet Score	-0.06	-0.05	-0.06	-0.05
N	944,593		69,937	

Notes: Own calculations on NielsenIQ Panel Data. Observation is at the household (HH) level using information from the first year in which the household enters the panel. *HH with Diab. & Ins.* are households with diabetes and private insurance (both declared and predicted).

Table 1.A.4. Event Studies

	(1) Carbs	(2) Fib	(3) Sgr	(4) Prt	(5) Fat	(6) Sat Fat	(7) Chol	(8) Sod	(9) Cals
Average Estimates									
Pre Average	142.4	-6.867	24.56	-41.74	80.34	-1.850	3.039	3.872	1125.7
	(167.6)	(9.604)	(52.03)	(28.01)	(84.40)	(24.32)	(2.573)	(8.743)	(1275.3)
Post Average	287.9	2.829	132.8**	-26.88	6.652	3.155	6.080*	4.615	1104.0
	(196.9)	(11.27)	(66.32)	(36.84)	(111.1)	(31.41)	(3.275)	(12.95)	(1578.5)
Leads and Lags									
$q-4$	261.7	2.526	8.544	-53.50	53.15	-19.74	1.057	10.67	1310.9
	(243.1)	(13.71)	(75.52)	(39.62)	(99.14)	(36.19)	(3.236)	(13.57)	(1698.0)
$q-3$	104.1	-13.92	-10.60	-50.83	48.28	-3.551	6.836**	4.122	647.5
	(212.1)	(11.85)	(65.29)	(34.42)	(104.5)	(29.75)	(3.413)	(11.44)	(1590.3)
$q-2$	61.42	-9.204	75.74	-20.88	139.6	17.74	1.223	-3.176	1418.6
	(171.9)	(9.383)	(70.73)	(26.96)	(107.5)	(27.29)	(3.164)	(9.288)	(1392.3)
q	412.5**	20.77*	93.16	50.14	329.9**	52.99	5.697*	6.065	4819.1**
	(200.8)	(11.46)	(59.52)	(35.97)	(137.0)	(35.40)	(3.057)	(10.87)	(1710.9)
$q+1$	328.9	7.529	166.6**	-15.58	67.10	27.32	9.476**	-1.294	1857.3
	(208.7)	(13.04)	(65.72)	(40.43)	(147.6)	(38.39)	(4.289)	(12.02)	(1878.0)
$q+2$	205.9	-13.50	145.3*	-79.29*	-246.7*	-46.55	3.391	7.273	-1714.0
	(245.0)	(14.52)	(81.48)	(47.15)	(144.4)	(42.52)	(4.149)	(22.28)	(2063.5)
$q+3$	204.4	-3.487	126.0	-62.77	-123.6	-21.15	5.755	6.416	-546.3
	(244.7)	(15.31)	(82.77)	(47.92)	(147.3)	(39.89)	(4.483)	(17.19)	(2089.2)

t statistics in parentheses
* $p < 0.10$, ** $p < 0.05$

Notes: N = 55,534. The table reports the event study estimates and the standard errors intervals using the Callaway and Sant'Anna (2021) estimator. All specifications have household and quarter fixed effects. $q-1$ is the reference period. The outcome variables are in grams columns 1–6, and columns 7–8 in milligrams: (1) Carbohydrates (2) Fibers, (3) Sugar, (4) Protein, (5) Fat, (6) Saturated Fats, (7) Cholesterol, (8) Sodium, and (9) Calories. The observation is at the household-quarter level. Source of the data is the NielsenIQ Panel.

Table 1.A.5. Timeline - COVID-19 Measures and Stimuli

March 13, 2020	COVID-19: National Emergency
April 2020	COVID-19 Stimulus Checks for Individuals ($1,200)
April 2020	Temporary layoff soared from 2 million to 18 million
December 2020	COVID-19 Stimulus Checks for Individuals ($600)
January 2021	The cap becomes effective in several other sates
March 2021	COVID-19 Stimulus Checks for Individuals ($1,400)

Notes: The table reports the timeline of the different COVID-19 measures and stimuli that were implemented between 2020 and 2021.

Appendix 1.B Diabetes and Insurance Status Imputation

In this appendix I provide additional details on the diabetes and insurance status imputation. As a first step, I impute diabetes status by using past survey responses in the survey, as households do not participate every year. Given that diabetes is a chronic disease, this is a reasonable assumption. In a second step, I impute the diabetes diagnosis for the individuals who do not reply to the Ailment Survey using information available for all NCP panelists. I explore several methods for the imputation, i.e., logit, probit, and random forest, accounting for the panel data structure with either random effects or time fixed effects. In addition, I also consider a simple rule based on observed diabetes-related purchases. To verify the accuracy of these imputations, I estimate the model on 80% of the sample and impute the diabetes status for the remaining 20% of the sample. The sample is composed only by observations that appear both in the NCP and in the Ailment Survey. For the imputation, I use demographic information—such as age, education, and race—and diabetes device purchases. The logit, probit, and random forest models perform well in correctly determining true positives and true negatives. However, these methods fail to capture a significant number of true positives and, thus, to significantly increase the sample size. For example, a random-effect logit model would increase the number of unique households with diabetes by 4.4%.[32]

Instead, using a simple rule, I identify the status of 1,970 additional observations, corresponding to around 1,261 unique households, leading to a 13.1% increase in unique households observed. This rule considers three aspects of purchasing insulin syringes and testing products. I employ (i) the annual amount in USD spent on diabetic products, (ii) the number of trips made to purchase diabetic products each year, and (iii) the number of packages of diabetic products purchased. I calculate the quartiles for these three variables using only positive values. I do this for insulin and testing products separately, using six different variables in total. From the quartile calculation, values equal to zero are excluded. Hence, if the variable equals zero, I assign a quartile value of zero. If at least one of these variables is in the third or fourth quartile, I categorize that household as having diabetes. I report the results of the rule imputation in Table 1.B.1. I show that 67% of households with diabetes are correctly identified.

[32] The main limiting factor is that I have so-called imbalanced classes since I impute the diabetes status, which affects only 12% of the U.S. population.

Table 1.B.1. Diabetes Status Imputation

		Diabetes		
		No	Yes	Total
Diabetes Imputation	No	110,314	23,041	133,355
	Yes	1,564	3,169	4,733
		111,878	26,210	138,088

Notes: This table reports the confusion matrix for the diabetes status imputation rule I perform. Observation is at the household-year level. The tables has two rows—indicating whether the imputation rule identifies the household as having diabetes—and two columns—indicating whether the household actually has diabetes. It reports the number of true negative (top left), false negatives (top right), false positives (bottom left), and true positives (bottom right). The last columns and rows report the total number of observations per group.

Similarly, I impute the insurance status for non-respondents. I employ a logit with random effects using the following information: the head of household age (a categorical variable with four different levels: below 40, between 40 and 60, between 60 and 65, between 65 and 70, and above 70), household size, household income, education in years, hours worked per week, a binary indicator equal to one for married couples, a binary indicator equal to one for being white, the type of residence (e.g., multi-family house or trailer), an indicator if the household is below the poverty line, and an indicator for the presence of children below 18. Before imputing the out-of-sample (i.e., for the non-respondents) of insurance status, I estimate the model on 80% of the Ailment Survey sample—which I randomly select—and impute the insurance status out-of-sample on the remaining 20%. The imputation results, reported in Table 1.B.2 show that 85% of households with private insurance are correctly identified.

Table 1.B.2. Private Insurance Imputation

		Insurance		
		No	Yes	Total
Insurance Imputation	No	3,848	11,141	14,989
	Yes	1,244	7,294	8,538
		5,092	18,435	23,527

Notes: This table reports the confusion matrix for the insurance status imputation I perform. I estimate a fixed effect logit on 80% of the sample and test the imputation on the remaining 20%, reported here. The tables has main two rows—indicating whether the imputation rule identifies the household as having private insurance—and two main columns—indicating whether the household has private insurance. It reports the number of true negative (top left), false negatives (top right), false positives (bottom left), and true positives (bottom right). The last columns and rows report the total number of observations.

Finally, I show that respondents and non-respondents are comparable. By construction, I can verify the validity of the imputation only for the Ailment Survey respondents. However, given that I impute the diabetes and insurance status of non-respondents, one might be concerned that the imputation for non-respondents might yield different results. As I cannot verify whether this is the case, I show that individuals participating in the Ailment Survey are similar to those only in the NCP by looking at the standardized difference between the two groups for the variables used for the imputation.

A standardized difference is defined as $\frac{\mu_a - \mu_b}{\sigma}$, where μ_a is the average for group a, μ_b for group b and σ is the common standard deviation. Standardized differences have the advantage of not being affected by different sample sizes across groups. For interpretation, a standardized difference equal to 0.5 indicates that the mean difference is half the standard deviation. In the context of covariate balance among two groups, while there is no consensus on the threshold, 0.25 is usually considered a small difference (Imbens and Rubin, 2015). I report the standardized differences for the variables used for the diabetes and insurance imputation in Panel A and B of Table 1.B.3, respectively. The two groups seem comparable as most standardized differences are well below 0.25 (in absolute terms). While three differences are above this value, they are only marginally so and they are below 0.5, which is considered a medium difference. To further mitigate any related concerns, in Section 1.5.3, I check that the results are robust to the exclusion of the imputed observations.

Table 1.B.3. *NielsenIQ Consumer Panel (NCP)* vs. *Ailment Survey*

	NCP Only		NCP & Ailment Survey		
	Mean	S.D.	Mean	S.D.	St. Diff.
Panel A: Variables Used for Prediction of Diabetes					
Different products - insulin syringes	0.09	0.79	0.15	0.99	-0.060
Total spent - insulin syringes	3.03	43.61	6.29	62.95	-0.059
Quantity - insulin syringes	0.12	1.16	0.20	1.46	-0.060
Different products - testing	0.02	0.36	0.04	0.62	-0.029
Total spent - testing	0.64	12.61	1.37	27.93	-0.032
Quantity - testing	0.03	0.57	0.05	1.00	-0.026
Panel B: Variables Used for Prediction of Insurance Status					
HH Heads Age	54.87	13.10	58.61	13.31	-0.283
HH Size	2.54	1.35	2.17	1.18	0.290
HH Income	59833.55	28,788.63	55435.14	29,108.44	0.152
HH Heads Education (years)	14.50	2.19	14.56	2.20	-0.029
Employment Head	4.56	2.89	5.36	3.18	-0.264
Married	0.65	.	0.60	.	0.120
White	0.76	.	0.81	.	-0.134
Living in Trailer	0.04	.	0.04	.	-0.031
Below Poverty Line	0.09	.	0.10	.	-0.048
Presence of Children <18 yrs	0.23	.	0.14	.	0.247
Observations	116,118		120,487		
Unique HHs	51,409		53,596		

Notes: Own calculations on NielsenIQ Panel Data. Observation is at the household-year level. "NCP Only" refers to households found only in the NielsenIQ Consumer Panel, while "NCP & Ailment Survey" includes households in both the NCP and the Ailment Survey. For all variables used to impute diabetes (Panel A) and insurance status (Panel B), I report the mean, standard deviation, and standardized difference (Cohen's d) between the two groups. Differences below 0.25 are considered small (Imbens and Rubin, 2015).

Chapter 2

Targeted Bidders in Government Tenders

Matilde Cappelletti and Leonardo Giuffrida[*]

2.1 Introduction

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

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

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

However, fostering equity through set-asides entails trade-offs on actual competition and efficiency in terms of contract outcomes and firm dynamics (see Section 2.3). The em-

[*] The authors gratefully acknowledge financial support from the Leibniz SAW project "Market Design by Public Authorities." Support by the German Research Foundation (DFG) through CRC TR 224 (Project A02) is gratefully acknowledged.

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

pirical existing literature has yet to conclusively assess these dimensions (e.g., Denes, 1997; Marion, 2007, 2009; Athey, Coey, and Levin, 2013; Nakabayashi, 2013; Shagbazian et al., 2025). In particular, unresolved issues include the generalizability of findings across heterogeneous procurements and set-aside types, as well as the implications for the execution stage of contracts and the long-term dynamics of winners. This paper aims to fill these gaps by analyzing short- and long-run effects of set-aside programs, concerning procurement and firm-related outcomes, respectively.

Our setting is U.S. federal procurement from 2008 to 2018, where the government allocates more than a quarter of its budget through set-asides, ranging from \$50.1 billion to \$81.2 billion yearly over the period of our analysis. This setting is ideal for shedding light on the unanswered issues for three reasons. First, it encompasses different types of set-asides. Second, the database contains detailed information on procurement categories and contract execution stage, and tracks supplier firms over time. Third, it allows us to leverage a set-aside spending shock.

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

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

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

Although the Federal Acquisition Regulation (FAR) §19.502 mandates the set-aside of contracts below a certain dollar threshold (the "simplified acquisition threshold"), it also makes clear that the set-aside is ultimately at the discretion of the agency.[2] Indeed, the agency's decision to set aside ultimately hinges on two key factors. First, the agency will set-aside a contract upon a positive assessment of at least two potential "competitive bidders" among the targeted firm set (the so-called Rule of Two). The regulation requires contracting personnel to explicitly report the evaluation of market research by filling out *ad hoc* forms, which are typically reviewed by the internal small business office, which must approve them prior to acquisition.[3] Thus, the role of private or soft information in the decision is minimal. Second, the agency's decision is influenced on the stringency of its annual set-aside spending target. The attainment of this target is evaluated annually by the Small Business Administration, and each agency receives scorecards on the goal achievement, implying greater future oversight over the agency's procurement activities in case of non-compliance (Dilger, 2024). Thus, agencies are incentivized to comply. We employ a propensity score-based method that, exploiting the wealth of our data, approximates the agency's regulatory constraints and information set to predict each set-aside decision.

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

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

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

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

[4] Specifically, in the first stage, we employ a machine learning approach for prediction of propensity scores (Goller et al., 2020; Lippmann, 2021). Among the different machine learning techniques, we choose the random forest for three main reasons. First, it requires only two tuning parameters and it is not sensitive to their specification. Second, it outperforms the traditional method for estimating the propensity score (i.e., the logit) in predicting the out-of-sample treatment (e.g., Lee, Lessler, and Stuart, 2010). Third, it enables us to include a wealth of predictors, thereby fully exploiting the richness of our available data and achieving covariate balance. On this dimension, the random forest is again superior to the logit. We provide a battery of statistical tests corroborating these arguments, which confirm the validity of our identification approach. In Section 2.5.1, we present recent examples in the literature of the adoption of random forest for propensity score estimation and further predictive purposes.

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

We lay out two non-competing asymmetric-information channels through which set-aside tendering negatively affects contract performance. First, set-asides select targeted suppliers that may lack quality or experience (adverse selection). Second, set-asides may disincentivize suppliers to perform optimally, knowing they have repeated opportunities within the program (moral hazard). We provide suggestive evidence of adverse selection for SBSA and moral hazard for DBSA through additional exercises based on winner fixed effects regressions, alternative winner-level outcomes (e.g., incumbency status), and effect heterogeneity on procurements differentially exposed to screening and monitoring.

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

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

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

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

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

Second, our paper relates to studies that examine policies mechanically altering competition in public tenders. Such policies are grouped into two categories: bid preferences and set-asides.[6] Krasnokutskaya and Seim (2011) show that preference programs increase procurement costs. Nonetheless, tailoring the program to different project categories marginally increases costs since the effect is extremely heterogeneous. Marion (2007) and Marion (2009) also find a cost increase from such programs. Rosa (2019) finds that, under such programs, affiliation—i.e., the dependence between the bidder's cost and that of its competitors—makes procurement more expensive and reduces efficiency. In contrast, existing studies on set-aside programs provide mixed results. Using U.S. Army Corps data from

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

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

1990-1991, Denes (1997) finds no evidence that set-asides for small businesses correlate with increases in the cost of government contracts. Athey, Coey, and Levin (2013) find that small business set-asides produce a decline in revenue and efficiency for the Forest Service. More recently, Shagbazian et al. (2025) provide evidence that set-asides induce lower award prices for homogeneous goods.

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

Third, this paper contributes to the empirical literature examining the impact of demand on firm performance (Foster, Haltiwanger, and Syverson, 2016; Pozzi and Schivardi, 2016). Focusing on public procurement demand, the current literature shows that firms exposed to a demand shock experience a persistent increase in revenue and employment (Ferraz, Finan, and Szerman, 2015; Fadic, 2020; Gugler, Weichselbaumer, and Zulehner, 2020; Lee, 2021). A positive shock also increases capital investment (Hebous and Zimmermann, 2021), facilitates access to external borrowing (Goldman, 2020; Di Giovanni et al., 2022; Gabriel, 2024), promotes innovation (Czarnitzki, Hünermund, and Moshgbar, 2020), and prolongs survival (Cappelletti, Giuffrida, and Rovigatti, 2024). Our work complements this empirical literature in two ways. First, we focus on procurement-specific firm outcomes, such as the delivery of multiple product variants to the government. Second, we also exploit a persistent yet unexpected surge in demand to study targeted firms' dynamics. Most of the works use identification strategies that mainly rely on temporary shocks, namely the random awarding of a contract. One exception is provided by Coviello et al. (2022), who exploit a legislative change permanently affecting the fiscal leeway of Italian municipalities. They find that businesses facing a persistent decline in public demand respond by cutting capital.

Despite this long list of contributions, the evidence in the literature on how set-asides affect the procurement process is inconclusive. We shed light on the following four dimensions, which remain unresolved to date. First, it is unclear whether the estimated effects can be generalized to a broader range of procurement categories. Second, relevant existing work focuses exclusively on ex-ante procurement outcomes such as the award amount. Third, there is limited evidence on the effect of other types of set-asides. A recent exception is Carril and Guo (2023), which investigates the impact of a veteran-owned small business set-aside program on various firm and procurement outcomes. Other works on disadvantaged business programs focus on the effect of subcontracting requirements on targeted firms. See, for example, De Silva et al. (2012), De Silva, Kosmopoulou, and Lamarche (2017), and Rosa (2020). Fourth, the long-term effects of such programs are unexplored. Hence, it is unclear whether the policy goal of leveling the playing field between small and large firms is achieved. We contribute to all of these unanswered dimensions. First, we use a large dataset on service and construction contracts from 11 years of government contracting across the US. Second, we evaluate the impact in terms of execution-stage contract outcomes. Third, we study more restrictive set-asides (i.e., the DBSAs). Fourth, we exploit the unexpected increase in spending for a particular set-aside program combined with an event study analysis to assess the long-term implications of this policy for winning firms.

The rest of the chapter unfolds as follows. Section 2.2 provides the necessary institutional information on the U.S. government's set-aside programs. In Section 2.3, we outline a theoretical background to our empirical research. Section 2.4 presents and describes the data. We pool the contract-level analysis in Section 2.5. In particular, the related identification concerns and strategy are explained in Section 2.5.1, while findings are presented and discussed in Section 2.5.2. Section 2.6 reports our empirical exercise for assessing the firm-level implications in the long run of set-asides for target firms. Section 2.7 concludes.

2.2 The U.S. Set-aside Programs

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

2.2.1 Set-aside Programs in the U.S. Government Procurement

The set-aside program has a long history in federal acquisitions, dating back to the 1950s when the U.S. Congress passed the first Small Business Act, which established that a "fair share" of federal contracts had to be set aside for small businesses. The SBSA has by far the largest budget for its set-aside program. Currently, the federal government aims to allocate 23% of its contracting budget to small businesses.[7] In federal government contracting, revenues (or employment) below a certain threshold define a company as "small" for a specific tender category.[8] To participate in a specific SBSA tender, firms self-certify that they meet the size requirements for "small business".

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

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

[8] We provide more definition on the definition of "small business" in Appendix 2.B.

a service-disabled veteran with a service-connected disability must own a majority stake in the small business and be in a management position.[9]

2.2.2 Set-asides in Practice

According to FAR §19.502-2 (a), "Each acquisition of supplies or services that has an anticipated dollar value above the micro-purchase threshold, but not over the simplified acquisition threshold, shall be set aside for small business [...]."[10] In practice, while contracts below the simplified acquisition threshold are generally earmarked for set-aside, the agency retains the discretion not to (to) set aside a contract below (above) the threshold by justifying their decision. In particular, during the procurement planning stage, the agency decides whether to set aside a contract for small business concerns by considering two factors, namely (i) agency-specific set-aside annual spending and (ii) the Rule of Two, i.e., the number of targeted businesses in the marketplace deemed capable of performing the work at a competitive price.

The first factor implies that the more below (above) the agency is from its set-aside annual target, the higher (lower) the pressure to set-aside. Agencies must consider this aspect when drafting tenders, as the Small Business Administration evaluates their achievement of spending targets on an annual basis and issues scorecards. For agencies that fail to meet these goals, the consequences are generally not punitive, but include increased oversight by the Small Business Administration and Congress, the development and implementation of improvement plans to address shortfalls, and potential pressure from the legislative or executive branch to encourage small business participation.[11]

The second factor is commonly known as the Rule of Two, which states that the agency must have a reasonable expectation of receiving a competitive bid from at least *two* small businesses to set aside a contract. Indeed, FAR §19.502-2 (a) clearly mandates the set-aside prescription, stating that it is applicable "[...] unless the contracting officer determines there is not a reasonable expectation of obtaining offers from two or more responsible small business concerns that are competitive in terms of fair market prices, quality, and delivery." The price, quality, and effective delivery chances associated with the bid must be expected to be fair and reasonable. The following conventional methods are recommended to prove such expectations: Reviewing recent procurement history, conducting market surveys, or obtaining expert advice from an agency small business specialist.[12] Critically, agencies must document their assessment of small business participation expectations and submit them to internal

[9] For details on the other types of set-asides, see Appendix 2.B. We will discuss the service-disabled veteran-owned small business set-aside again in Section 2.6.

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

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

[12] For detailed information on practical guidance for setting aside a contract, visit https://www.sba.gov/partners/contracting-officials/small-business-procurement/set-aside-procurement. In addition, a small business self-certification database (i.e., Dynamic Small Business Search) is available to agencies, which may use this

small business offices for approval prior to acquisition. Within the Department of Defense, for example, this process is standardized through the use of the DD Form 2579. A blank version of the DD Form 2579 is provided in the Online Appendix for reference.

2.3 Theoretical Background

In this section, we set out the theoretical foundations of the impact of set-aside in the procurement process. We divide the discussion into contract- and firm-level implications according to the structure of our empirical analysis, for which we postulate trade-offs. We empirically investigate these trade-offs in Sections 2.5 and 2.6.

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

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

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

A negative effect would align with Li and Zheng (2009), who argue that more *potential* bidders result in less profitable entry due to increased competition and a lower winning chance. In contrast, a positive effect would be consistent with Cantillon (2008), who claim that reduced asymmetry in bidders' production costs leads to fiercer competition, as set-asides likely induce more symmetric value distributions. Thus, depending on the prevailing effect, set-asides could either encourage or discourage entry, influencing actual competition in procurement tenders. The latter effect would reinforce the equity impact of set-asides, confirming the entry of targeted firms, whereas the former could be attributed to the exit of larger firms only. The net outcomes depend on the mix of targeted and untargeted potential participants and the entry costs for targeted in the specific procurement market.[13]

However, how does excluding untargeted firms from bidding and reserving the participation to targeted bidders affect the efficiency of the procurement process? This question is

database for consultation but must carefully consider a small business bidder's self-certification before awarding a contract.

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

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

- Hypothesis *H.b1*: Smaller firms receiving contracts might leverage their intimate knowledge of local resources and conditions, leading to more efficient resource allocation—*advantageous selection effect*;
- Hypothesis *H.b2*: Set-aside policies might incentivize smaller or less experienced firms to exert greater effort or commitment to prove their capabilities. This heightened dedication could enhance, e.g., project management and effort—*commitment effect*;
- Hypothesis *H.b3*: Set-aside policies could inadvertently attract firms that are less capable or experienced, leading to inefficiencies in contract execution—*adverse selection effect*;
- Hypothesis *H.b4*: Firms benefiting from set-aside policies might feel less pressure to perform optimally, knowing they face repeated opportunities under the program—*moral hazard effect*.

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

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

- Hypothesis *H.c1*: the opportunities provided by set-aside contracts could empower small businesses to expand their size and become more competitive—*market-enhancement effect*;

- Hypothesis *H.c2*: persistent reliance on the set-aside program could dissuade firms from expanding to a point where they lose eligibility for participation—*deadweight-loss effect*.

2.4 Data

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

Federal Procurement Data System (FPDS). The FPDS is a publicly available database that tracks the vast majority of U.S. federal procurement spending.[14] The dataset provides a wealth of information, including over 200 variables on contract, seller, and buyer characteristics. Examples of contract-level information include the value of the contract, the start and end dates, the number of bids received, the place of performance, the type of product or service purchased, and the type of set-aside. Examples of seller-level information include the seller's ID and headquarters' location as well as the met small business standards, if any. An example of buyer-level information is the identity of the awarding agency. The FAR prescribes contracting agencies to report all awarded contracts with an estimated value above $3,500 (in the period under study) to the FPDS. Any subsequent contract modification, regardless of its amount, must also be reported and categorized in the FPDS.

Sample Selection. We start with the entire population of contracts from fiscal year (FY) 2008 to 2018. To calculate our metrics for execution-stage procurement outcomes, we need contracts that include an execution phase—i.e., with potential cost overruns and delays. Accordingly, we consider service and construction contracts. This selection implies excluding contracts where renegotiations are not meaningful to their outcomes: research and development, physical deliveries, and leasing.[15] Similarly, we exclude indefinite-delivery contracts as they are based on agreements with a supplier for an indefinite quantity of goods and services over a specified period of time, so delays and additional costs cannot be interpreted as indicators of poor quality performance. We also limit our sample to contracts that are performed within U.S. borders and with a fixed-price format.[16] Finally, for the sake of comparable processes, we exclude very small contracts from the sample, i.e., those with an expected duration of less than 2 weeks *or* expected cost of less than $25,000.[17] This ultimately leaves

[14] See www.usaspending.gov.

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

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

[17] Note that the $25,000 selection is a natural choice given that above this threshold contracts have to be posted on a centralized online platform. See Carril, Gonzalez-Lira, and Walker (2022) for more details on this threshold. In Appendix 2.D, we show that our results are robust to this sample selection.

us with a sample of 141,199 contracts (38% and 15% of which are SBSA and DBSA, respectively) with a total value of $125 billion ($11 billion yearly on average), about 570,000 bids submitted and 54,800 unique winners.

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

We stress again that these metrics are the most common execution-stage metrics in both project management (Herweg and Schwarz, 2018) and empirical economics (Decarolis, 2014). They aim to capture the *quality* of contracts, based on the idea that renegotiations of fixed-price contracts lead to adjustment costs that are suboptimal for all parties involved. Indeed, Spiller (2008) proposes this argument for the first time for the public procurement context and claims that contracts are less flexible than private contracts. As a result, they require more frequent renegotiations and provide weaker incentives to comply with contract terms. Such rigidity leads to higher adjustment costs than private contracts. Bajari, Houghton, and Tadelis (2014) quantifies this claim empirically and finds that adjustment costs in public procurement are high and can account for 7.5% to 14% of the winning bid. Moreover, cost overruns and delay in public services and construction may create important disruption for the affected citizens.

We also build two secondary outcomes measuring the extensive and (a further dimension of) intensive margins of contract amendment. The details and the results on these two variables are presented in Appendix 2.D.

Stylized Facts. Table 2.1 provides summary statistics on our outcomes of interest across control and treated groups. On average, treated tenders (i.e., SBSA or DBSA) have worse execution-stage outcomes (i.e., more delays and higher additional costs) than the control group (i.e., NSA). The difference is more pronounced for DBSAs. We find a similar pattern in terms of *Delay*. Note that both types of treated contracts receive more bids on average than non-treated contracts. Table 2.1 also displays that the average award amount of DBSAs is twice as large as that of SBSA and the former last longer, highlighting asymmetric distributions. From a descriptive viewpoint, participation restriction correlates with worse execution performance and more bids. Our first goal is to identify these effects. We do that in the next section.

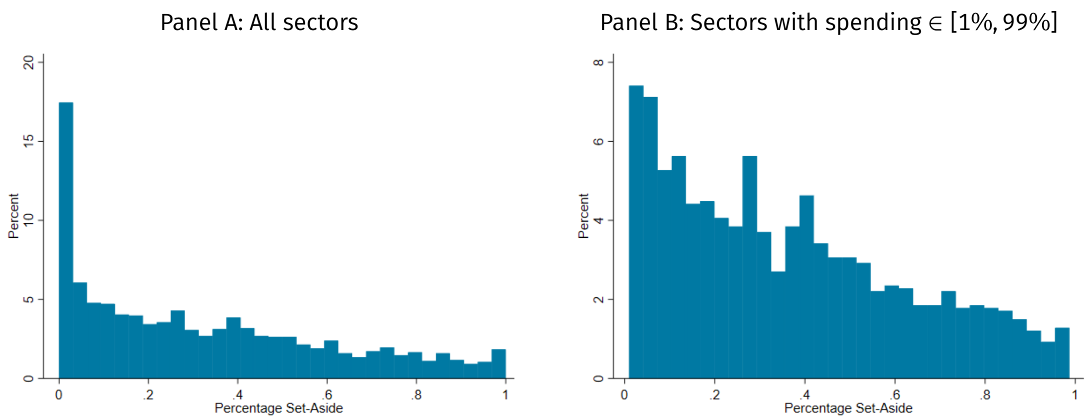
Table 2.1. Summary Statistics for Treatment and Control Groups

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

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

Is Set-aside Used Across Procurement Categories? Figure 2.1 shows the within-sector share of set-aside spending across all sectors in our data. We define the sector by considering the procurement category defined by the four-digit product or service code reported by FPDS.[18] This share is fairly uniform across sectors. In addition, the percentage of sectors in which set-asides are never or always used is 13% and 1%, respectively, which together represent $<1\%$ of our sample of contracts. These statistics are reassuring given that the differential use of set-asides across sectors could be a potential confounding factor in our pooled comparison of procurements.

Figure 2.1. Set-aside Spending Shares per Sector



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

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

2.5 Contract-level Analysis

2.5.1 Methodology

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

2.5.1.1 Identification Concerns and Strategy

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

$$Y_i \;=\; \alpha + \beta^{SA}\, SetAside_i^{SA} + \gamma\, \boldsymbol{C}_i + \epsilon_i, \tag{2.1}$$

where Y_i and C_i denote our outcomes—e.g., number of bidders, extra costs and delays—and a vector of controls of the contract i, respectively. $SetAside_i^{SA}$ takes the value 1 if the contract is awarded through a set-aside procedure—i.e., $SA = \{SBSA; DBSA\}$—and 0 if no set-aside program is implemented. The parameter of interest in Equation (2.1) is β^{SA} as we are interested in the effect of SBSA and DBSA programs on the selected outcomes, separately. We proceed with the rest of the analysis using two different subsamples of the dataset, designating one of the two *SA* groups as the treatment group and maintaining the NSA as the control group.

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

The decision to set aside a contract relies on two main aspects. In Appendix 2.C, we provide evidence that these two aspects are relevant for predicting the agency's decision. First, the agency needs to implement the Rule of Two (as discussed in Section 2.2.2). The higher the number of "targeted competitive firms" available for a particular procurement—and thus in a specific market-time—the greater the likelihood that the agency will set aside the tender. We exploit various nuances related to the availability of targeted competitive firms across different market definitions, as neither dimension is specified in detail in the regulation. Second, the agency has a specific spending target for set-aside contracts during and over the years. The further below (above) its annual target the agency is at the time of award, the more (less) likely it is that the tender i will be set aside. Our data allows us to reconstruct this record as well.

We process the constructed information with a propensity score approach, which allows us to compare set-aside contracts similar in observable characteristics to open procedure contracts. We proceed with a two-stage approach. In the first stage, using a random forest, we predict the probability (i.e., propensity score) of set-aside treatment assignment (i.e., SBSA or DBSA) conditional on observed covariates. The estimated propensity scores are used in the second stage to run a weighted version of Equation 2.1 to identify our coefficients of interest. We discuss the details of the second stage in Section 2.5.1.3. Under specific assumptions (we report them in Section 2.5.1.2 and formally define them in Appendix 2.C), this method allows us to perform an unbiased counterfactual analysis by eliminating key observable differences between the treatment and the control groups. We find no violation of such assumptions in our setting.

Finally, we argue that a propensity score-based approach is appropriate, as our setting exhibits desirable features for this method. In particular, the fact that the agency retains some margin of discretion in implementing set-aside rules is not an issue but rather an advantage for two reasons. First, discretion prevents the treatment assignment from being deterministic. A deterministic assignment would violate the common support assumption on which the validity of the propensity score relies. Second, discretion is why similar contracts have different treatment statuses (Gibson and McKenzie, 2014). Indeed, agencies must justify why they do or do not set aside a contract (see Section 2.2), making the role of private information in the decision negligible.

2.5.1.2 First Stage: Propensity Score of Set-asides

In this subsection, we explain the approach used to predict the probability that a contract will be tendered with one of the two set-asides. We do so by using a random forest, following in particular Lee, Lessler, and Stuart (2010).[19] We also provide information about the treatment predictors and the identification assumption needed for this approach.

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

Propensity Score Estimation. In this work, we are interested in the probability of each contract being set aside. We define $p_i(X_i)$ for each tender i as the estimated propensity score receiving a binary treatment, where X_i is the vector of covariates. Given the binary nature

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

of the outcome, our random forest aggregates multiple classification trees.[20] To predict the propensity score, we use a classification method with multiple binary, categorical, and continuous covariates as inputs, and we perform two separate classifications for the two subsamples, one classification for the treatment being SBSA and one for the treatment being DBSA. For each subsample, we grow a random forest of 1,000 trees by closely following the practical guidance in Breiman and Cutler (2011) and Liaw, Wiener, et al. (2002) in tuning the parameters.[21] Following Lee, Lessler, and Stuart (2010), we compute a continuous propensity score for each contract by taking the average of the predicted outcomes of each tree.[22]

Why Using the Random Forest for Propensity Score Estimation? The recent literature has shown the merits of machine learning techniques such as classification and regression trees when predicting propensity scores (e.g., Lee, Lessler, and Stuart, 2010). Among these techniques, we choose the random forest for three reasons. First, not only this method requires only two tuning parameters but it is also not sensitive to their specification (Liaw, Wiener, et al., 2002). Second, we find that it outperforms the logit—the traditional method for estimating the propensity score—in predicting the out-of-sample treatment. Third, it enables us to include a wealth of predictors, thereby fully exploiting the richness of our available dataset. Hence, the random forest also allow us to achieve covariate balance, while the logit does not. In Section 2.5.2.3, we present the empirical exercises that corroborate these arguments.

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

Which Variables Explain the Set-aside Decision the Most? We select and build variables that (i) replicate the information set to implement the Rule of Two, (ii) reconstruct set-aside spending in the agency from year-start, and (iii) control for contract and agency features.

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

[21] See Appendix 2.C for the exact steps taken in the random forest.

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

All of these covariates must satisfy the following conditions to be salient predictors: intuitively correlate with both treatment and outcomes simultaneously, are measured prior to treatment, and are orthogonal to the anticipation of treatment. We identify 71 variables the variables that meet these criteria in our data.[23] Then, we allow the random forest to select on the most relevant variables for the propensity score prediction. The agencies seem to adhere closely to the Rule of Two: Our random forest selects the (i) *percentage of small firms within a sector and division* (as defined by the Census Bureau) and (ii) *percentage of contracts previously awarded with set-aside by the same agency* as the two most significant factors for a contract to be set aside.[24]

Identification Assumptions. The propensity score approach relies on two main assumptions, which we formally define in Appendix 2.C. First, we need *common support* (or "overlap assumption"), which implies for us that each contract (i) *could* receive both treatment levels and (ii) that we cannot *perfectly* predict the probability of receiving treatment. Both conditions apply in our setting since, on the one hand, contracts of any amount could be set aside; on the other hand, the set-aside choice is not entirely deterministic, as the rules are not binding. As discussed above, the role of private information in this choice is minimal, implying that the agency makes decisions based primarily on hard information about the contact and buyer characteristics. In Section 2.5.2.3, we assess the plausibility of this assumption and find that it is verified. Second, we assume *unconfoundedness* ("or conditional independence"), i.e., no unobservable or other observables might influence selection into treatment. In Section 2.5.1.1, we documented the agency's decision-making process and argued that, through our data, we are able to reconstruct all relevant dimensions driving set-aside choices. Moreover, we do not find a violation of this assumption in our data and report the results in Section 2.5.2.3.

2.5.1.3 Second Stage: IPW Regression Using Propensity Scores as Weights

As a second stage, we employ the inverse of the propensity scores estimated to perform as weights for a weighted OLS regression of Equation 2.1. The model specification turns into

$$Y_{i,b,t} = \alpha + \delta^{SA}\, SetAside_i^{SA} + \eta \boldsymbol{X_i} + \zeta_{b,t} + \varepsilon_i, \tag{2.2}$$

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

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

[24] Among other variables, the other most important factors are the (iii) *percentage of competitive contracts awarded to small firms in the division/year*, (iv) *percentage of competitive small firms in the division/sector/year*, (v) *percentage of small firms that also win without set-aside in the sector/year*, (vi) *percentage of small firms that always win without set-aside in the sector/year*, and (vii) *percentage of small firms in the sector/year*. We refer the reader to Appendix 2.C for an in-depth discussion of the logic, grouping, and relevance of first-stage predictors X_i.

used in the first stage in the model. Nonetheless, they are captured due to the dynamic nature of the interaction fixed effects.

We are interested in the average treatment effect on the treated (ATT) contracts. Formally, in the potential outcome framework, the ATT is defined as $\theta_{att} = E(Y(1) - Y(0)|SA = 1)$, where $Y(1)$ is the outcome for the restricted solicitation, and $Y(0)$ is the counterfactual outcome had the contract had not received the treatment (Heckman and Robb Jr, 1985; Angrist, 1998).[25] To estimate an ATT, the IPW cancels out the differences in observable characteristics and gives more weight to those observations that have a low probability of receiving the treatment and vice versa—i.e., by weighting the treated by 1 and the control by $\frac{p_i(X_i)}{1-p_i(X_i)}$. For the standard errors of the estimates, we compute the variance-covariance matrix with a sandwich estimator when weights are used (Dupraz, 2013). Following the clustering recommendations of Abadie et al. (2023), we do not cluster the standard errors as the treatment decision is at the individual (i.e., contract) level. Finally, in Section 2.5.2, we check that the results are not qualitatively sensitive to the chosen approach.

2.5.2 Results

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

2.5.2.1 Baseline Results

In Table 2.2, we report the baseline results of the IPW regressions. The top panel reports the estimated coefficient $\widehat{\delta^{SA}}$ from the Equation (2.2) for $SA = SBSA$, and the bottom panel reports the results for $SA = DBSA$. First, columns 1 and 2 report the effect on the log number of bids received. We then report the estimates on the two *execution-stage* contract outcomes, i.e. *Extra Cost* (columns 3 and 4) and *Delay* (columns 5 and 6). The odd columns show point estimates with a plain model specification, while the even columns include contract-level first-stage covariates and agency×FY fixed effects. We set the rich model as our preferred specification.

We find that SBSA increases the number of bids by 23.7%—i.e., $\left(\exp(0.213) - 1\right) \times 100$. DBSAs have a slightly weaker effect on competition, i.e., +20.2%. The magnitude of these effects amounts to approximately one additional bidder. While the Rule of Two is designed to ensure a minimum level of small business participation, our results show that set-asides attract more bids than the regulatory minimum. The average receipt of 3.6 bids for NSA contracts, our control group (see Section 2.4), confirms that the increase in bids for set-aside contracts goes beyond mere regulatory compliance, indicating a meaningful

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

increase in competition. Notably, about 20% of SBSAs in our dataset received only a single bid, underscoring the unpredictability of bidder responses to set-aside auctions.

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

Table 2.2. Baseline Outcomes, IPW

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

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

All in all, these results show that restricting participation to small or disadvantaged firms increases competition in government tenders, confirming the intended equity effect of the policy. This points to a positive competition effect as put forward in Section 2.3. Second, we find an increase in spending through additional budgetary costs and delays in contract execution. Over both cost and time dimensions, set-asides fare worse in the execution phase, thus pointing toward evidence in favor of H.b3 and H.b4. In the next subsection, we discuss how to empirically disentangle these mutually non-excluding channels. Finally, in 2.5.2.3, we discuss the validity and robustness of the results.

2.5.2.2 Channels for the Inefficiency Effect

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

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

We present the results in Table 2.3 for both the execution-stage outcomes—*Extra Cost* in columns 1-4 and *Delay* in columns 5-8. We assess the impact on the same firm awarding contracts with or without SBSA or DBSA procedures (columns 2 and 6). We also interact the winner fixed effects with the award year (columns 3 and 7) or with the purchasing sector code (columns 4 and 8) to additionally capture unobserved firm-specific time trends (e.g., management practices) or firm-product-specific features (e.g., specialization), respectively. Columns 1 and 5 report the baseline estimates from Table 2.2 to facilitate comparison. For SBSA, estimates turn insignificant across the different specifications of firm fixed effects, suggesting an adverse selection effect. For DBSA, instead, we find statistically indistinguishable results from the baseline. Although we cannot exclude adverse selection, a moral hazard effect seems at play for DBSA.

Table 2.3. Execution-stage Outcomes with Seller Fixed Effects

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

Notes: In columns 1 and 5, we replicate Table 2.2 columns 4 and 6, to which we add seller fixed effects (columns 2 and 6), firm-year fixed effects (columns 3 and 7) as well as seller-one-digit-service-code fixed effects (columns 4 and 8).

Are Set-aside Winners Different? To further rationalize these findings—namely to corroborate adverse selection and exclude moral hazard for SBSAs and vice versa for DBSAs—we investigate whether SBSA and DBSA winners have different characteristics than open procedure winners. To account for winner's characteristics, we construct three different outcome variables. First, *Entry* is an indicator equal to one if the firm has not previously been awarded a contract in our data. Second, *Incumbent* is a binary variable equal to one if the winner is awarded at least one contract in the previous year. Third, we construct *Backlog* by counting the number of contracts won in the previous quarter. All these variables capture the concept of incumbency with different specifications. Incumbency is relevant as it informs about firms' experience in the procurement market. [26]

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

Table 2.4. Entry, Incumbency, and Backlog

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

Notes: The baseline model specification from Table 2.2 (columns 4 or 6) is replicate on firm-level outcomes: *Entry [N,Y]* is an indicator equal to one if the firm is winning a contract for the first time. *Incumbent [N,Y]* is a binary variable equal to one if the firm is an incumbent, i.e., won a contract in the previous year. *Backlog* counts the number of contracts won in the previous quarter.

Does Monitoring Set-asides Mitigate their Inefficiency? To draw further intuitions on the channels at play, we leverage the role of performance bonds. Performance bonds, issued by surety companies, unequivocally distinguish construction from service contracts in the U.S. federal procurement. The bonds are shown to enhance contract execution outcomes by enforcing effective screening and monitoring on suppliers, both theoretically (Calveras, Ganuza, and Hauk, 2004) and empirically (Giuffrida and Rovigatti, 2022), mitigating supply risks. Surety companies consider firms' past performance when issuing bonds, motivating suppliers to maintain high standards in their operations, even when bonds are not contractually required.[27]

This specific attribute of performance bonds provides a context to discern the underlying mechanisms of inefficiencies from set-aside contracts. Table 2.5 presents split-sample

[26] Note that these three variables refer to previous procurement activity of the firm using the entire FPDS population of service and construction contracts. Hence, we avoid missing values by executing such regressions on all years except for FY2008, the first year in the data. We replicate baseline analysis without FY2008 and results (unreported) are qualitatively and quantitatively unaffected to such exclusion.

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

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

Table 2.5. ATET – Heterogeneity by Works and Services

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

Notes: Results for the ATT of the IPW on the three main outcomes, *Log(# Offers)*, *Extra Cost* and *Delay*. We report, for each outcome, the baseline results as in Table 2.2. We then report results separately by procurement category, either works or services.

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

2.5.2.3 Validity of the Identification Assumptions

In this subsection, we first report the tests of our identification assumptions that corroborate the internal validity of our causal claims. Then, we justify the chose random forest approach and provide evidence on the better performance of the random forest compared to the logit. We refer the reader to Appendix 2.D for additional robustness checks on methodology, alternative outcomes, and sample selection. Altogether, such empirical exercises highlight the robustness of our findings.

Testing the Identification Assumptions. As discussed in Section 2.5.1.2, our chosen method relies on two assumptions on the propensity score, i.e., the overlap and unconfoundedness assumptions. We test validity of the first assumption in two different ways. First, to check the satisfaction of the overlap assumption, we verify that the propensity score distributions for both the treated and control overlap. The results of this check are shown in Figure 2.A.1 in Appendix 2.A. We find a strong overlap, and we exclude those units from the relevant population whose probability of receiving the treatment can be perfectly (or almost perfectly) predicted (Wooldridge, 2010). Therefore, we restrict our sample to units for which the propensity score is strictly between 0.01 and 0.99, implying removing 11.6% of the observations for the SBSA sample and 23.7% for the DBSA sample without loss of internal validity.

Second, in Figure 2.2, we look at the covariate balance between treated and control contracts. These results allow us to check the validity of both assumptions. Indeed, when covariate balance is achieved, the propensity score has been adequately specified, implying that treatment effect estimates is valid and unbiased (Ho et al., 2007; Zhao et al., 2016). For this purpose, we compute the standardized differences after applying the IPW (Imbens and Rubin, 2015). This procedure allows us to assess the comparability of treated and control units in the weighted sample (Austin, 2009). Moreover, the standardized differences are more robust than simply calculating the t-statistic or testing the difference in means because they do not depend directly on the sample size (Wooldridge, 2010). As long as the standardized differences do not exceed 0.25 (in absolute value), covariate balancing between groups should be satisfied (McCaffrey, Ridgeway, and Morral, 2004; Stuart, 2010; Imbens and Rubin, 2015). We obtain excellent results: for SBSA all of the variables used for propensity score prediction are below 0.1. For DBSA, 5.6% are above 0.1, and none are above 0.15.

We assess the validity of the unconfoundedness assumption following Imbens and Rubin (2015). To do so, we assess the sensitivity of our estimates to different choices of predictor variables. In these robustness checks, we remove some variables from the propensity score prediction covariates vector. In other words, we rely on the concept of subset unconfoundedness, which is a more restrictive condition than the original assumption on which the propensity score relies, the unconfoundedness assumption (Caliendo and Kopeinig, 2008). As mentioned before, this assumption is not directly testable, and consequently neither is subset unconfoundedness. However, the two assumptions combined have some testable implications: We can test whether adjusting for differences in a subset of covariates give similar point estimates as with the full set of covariates. When we do so, we find that the coefficients are very stable across different specifications and are not statistically different. We report the results in Figure 2.3. The first line in each figure reports the baseline estimates from Table 2.2, in the specification with fixed effects and additional controls. We report the coefficients for SBSA with red dots, while we use blue triangles for DBSA. In lines two and three, we report the coefficients excluding variables that we created to account for, respectively, (i) the performance in terms of extra cost and delay in the previous period, at the agency level and (ii) the level of competition in the geographical region and sector. Therefore, we can conclude that we find no evidence of (i) a lack of the overlap assumption and (ii) the violation of the unconfoundedness assumption.

Testing the Random Forest's Performance. As discussed in Section 2.5.1.1, we choose the random forest for three reasons. For each dimension, we now provide empirical evidence to strengthen our arguments. First, we show that the random forest is robust to the specification of its parameter, by twisting the main parameter of the random forest, i.e. the number of variables selected at each node.[28] In Figure 2.3 line four and five, we impose the number of variables instead of using the number of variables based on prediction accuracy.[29] We report the results for alternative numbers of variables selected at each node.[30]

Second, we test the prediction performance of random forest vis-à-vis logit. We train both models on the same sub-sample of the data and predict the treatment on the unused part of the sample in the second step.[31] The random forest has a much higher accuracy of 77%, while the logit correctly classifies the treatment in only 48% of the cases. This difference in prediction performance suggests that the random forest does a better job of using the observable information provided in the data. We note that, however, the logit performs better in terms of the overlap assumption. This can be observed in 2.A.2 in Appendix 2.A, which replicates Figure 2.A.1 using the logit.

Third, we have shown in this subsection that the random forest achieves great covariate balance. We stress here that the random forest, by allowing the inclusion a rich set of covariates and controlling for nonlinearities and interaction, achieves better covariate balance than a logit. We provide evidence of this in Figure 2.A.3 in Appendix 2.A, replicating Figure 2.2. We observe that, for some of the buyer characteristics, the covariate balance is not achieved with the logit, as some standardized differences are above the absolute value of 0.25. Therefore, the validity of the coefficient is not met for the logit approach, while it is for the random forest.

For completeness, in a robustness check, we also replicate the main analyses using the logit instead of the random forest. In Appendix 2.D, we report the results with the logit and show that they are quantitatively and qualitatively similar to our baseline. Moreover, the coefficients with the logit are always larger than with the random forest. If anything, this could suggest that our chosen methodology underestimates the effect. We stress again that the logit estimates cannot be causally interpreted, given that with this method covariate balance is not achieved.

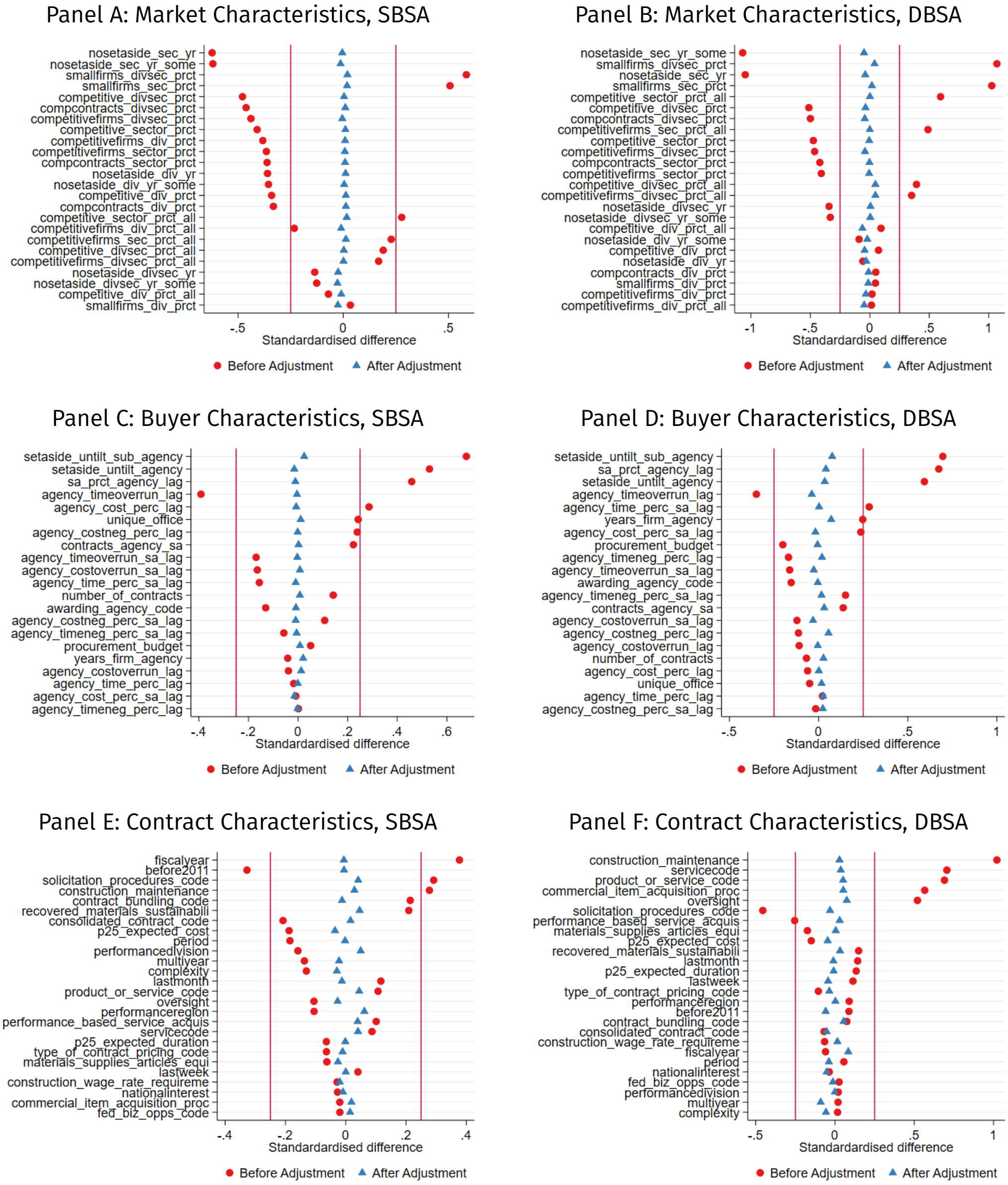
[28] Note that the second important parameter for the random forest is the number of trees (Liaw, Wiener, et al., 2002). However, this is not relevant in our case as we are already using 1,000 trees.

[29] We refer to Appendix 2.C for more details about accuracy and its definition in this context.

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

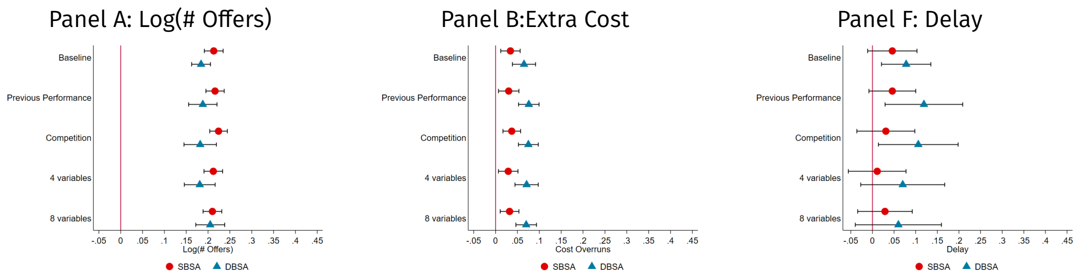
[31] We split the dataset and use 80% of it as training sample and the remaining 20% as testing sample.

Figure 2.2. Standardized Differences



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

Figure 2.3. Robustness Checks, Functional



Notes: This figure reports the different robustness checks performed on our analysis. The panel title report the outcome variable of interest. The red dot reports the coefficient for small business set-asides (SBSA) as treatment group, the blue triangle for disadvantaged business set-asides (DBSA). The black confidence interval are at the 95% confidence level. We report the different functional robustness checks performed on the propensity score prediction with the random forest. The first line reports the baseline regression, as in Table 2.2, with 16 (8) variables selected at each node of the random forest, seed equal to 1994 and all variables. In the second and third line, we remove some variables related to previous performance and competition. In lines four and five, we change the number of variables selected by the random forest.

2.6 Firm-level Analysis

In the previous section, we show that set-asides boost the participation of targeted firms, promoting equity in the process, and induce poorer contract outcomes, both in terms of cost and time dimensions. In this way, excluding untargeted competitors creates inefficiencies in the government procurement market. Can this additional cost to the taxpayer be viewed as an "investment" from the policymaker's perspective? In other words, do firms that benefit from set-aside contracts perform better in the future, thus potentially leading to increased efficiency in the long run? To provide evidence for a preliminary answer to these questions, in this section we empirically investigate the underlying trade-off for the government between the market-enhancement effect (i.e., *H.c1*) and the deadweight-loss effect (i.e., *H.c2*) as posited in Section 2.3.

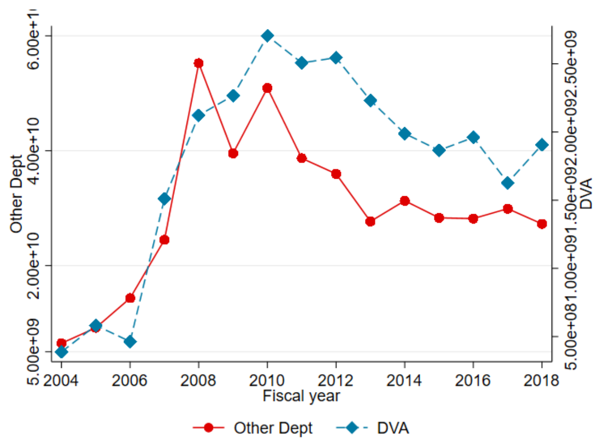
Identifying a firm-level effect of set-asides poses empirical challenges. A primary concern is that winning *specific* contracts can be anticipated by a firm, making winning a set-aside contract endogenous to a firm's unobserved characteristics. For instance, we are agnostic about the political connectedness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

To address such challenges, given that demand variation is a crucial driver of firm dynamics (see Section 2.1), we propose an event-study exercise that builds on a large, permanent, and unexpected construction spending increase by the Department of Veteran Affairs (DVA) on a given category of DBSA (i.e., the service-disabled veteran-owned small business set-aside, henceforth VBSA for convenience).[32] In addition to the policy-induced variation, we leverage the variation brought about by the differential pre-period exposure of recipients eligible to the spending spike.

2.6.1 The VBSA Spending Surge

The Recovery Act and Veteran Policies in 2009. We leverage the two government policies that were independently introduced during the same period. First, newly elected President Obama signed the Recovery Act into law in February 2009. This stimulus package contained provisions for the construction industry beginning in FY2009 to increase investment in the nation's physical infrastructure and cope with the ongoing economic downturn. Notably, such resources were mostly channeled through government procurement contracts.[33] Second, the years of the fiscal stimulus also saw policies in line with the stated mission of President Obama to increase subsidies for veterans. As a result, the President's budget request in FY2009 included $140 billion for the DVA—a 40% increase in funding from FY2008.[34] Figure 2.4 shows how the combination of these policies increased DVA's total construction spending by approximately 50% from 2008 to 2010.[35]

Figure 2.4. Total Spending by the DVA vs all Other Departments, Construction



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

[32] As this section focuses on VBSA, we replicate our contract-level analysis of DBSA versus NSA using VBSA as the treatment group. This auxiliary exercise documents that VBSA is also disruptive for execution-stage outcomes. In particular, extra costs increase, but we find no meaningful effect in delays. We report the results in Table 2.A.1 in Appendix 2.A.

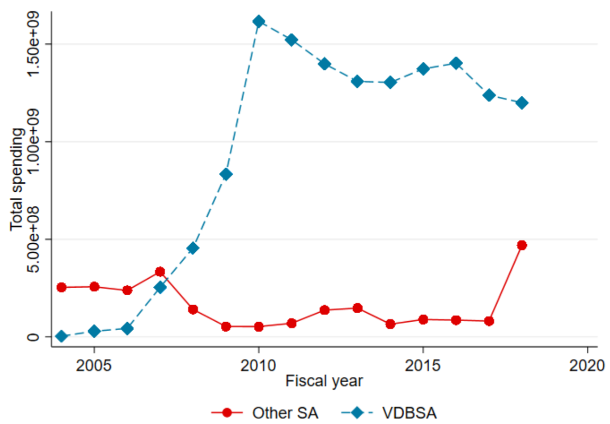
[33] For more information, see here.

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

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

Service-disabled veterans were a core target of the new veteran policy. As a result, the combined implicit effect of the Recovery Act and the new veteran policy was to increase VBSA spending in constructions. As Figure 2.5 shows, the absolute increase in construction spending in FY2009-2010 (compared to FY2008) for VBSAs was threefold, rising from approximately \$0.5B in FY2008 to \$1.5B in FY2010. Such spending amount holds in the years after, thus highlighting a persistent spending surge. Moreover, the same figure emphasizes that a similar increase in construction spending from the DVA for other types of set-asides is not observed in our data.

Figure 2.5. Total Spending in VBSA vs Other Set-Asides by the DVA, Construction



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

Contamination Concerns and Firm Sample Selection. This discussion raises the following question: Which firms benefited from the VBSA spending surge? In theory, any *small* construction firms classified as *service-disabled veteran-owned* were eligible to compete for the increased VBSA contract opportunities. As shown in Figure 2.6, the cumulative yearly number of bidders in VBSA auctions increases strongly from 2009 onward. The increasing opportunities for service-disabled veteran-owned small businesses likely drive this statistic as incentivizing incumbent businesses to participate in VBSA tenders. Also, small firms (with no further designation) could have been incentivized to switch to the small service-disabled veteran-owned category to take advantage of the increased government demand. For example, an existing small firm could have invited a disabled veteran to become a major shareholder in the company. Moreover, at the beginning of FY2009—i.e., after the election victory—President Obama also announced the withdrawal of (most) U.S. troops from Iraq. Accordingly, many veterans returned home between 2009 to the end of 2011, suggesting a growing number of service-disabled veterans establishing startups and entering the construction industry.[36] The latter two aspects jointly highlight a possible concurrent supply shock

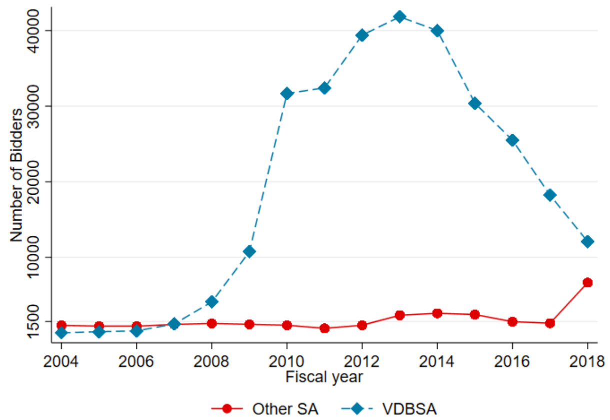
[36] This trend has been observed and reported by the media. Most veterans who start businesses do so in the construction industry because construction jobs best match veterans' skills. For example, see https://www.nvti.org/2023/08/31/employment-in-the-construction-industry-for-veterans/. Recently, Coile, Duggan, and Guo

of service-disabled veteran-owned small bidders that might contaminate the exogeneity of contract recipience and introduce self-selection in our sample.

The post-FY2009 service-disabled veteran-owned small firm pool is thus most likely composed of incumbents and entrants. Specifically, we define incumbents as firms that were awarded at least one VBSA construction contract before FY2009 *or* labeled service-disabled veteran-owned despite being awarded construction contracts outside the VBSA program. We define entrants as newly re-labeled or newly established service-disabled veteran-owned firms in the data after 2009. Our analysis restricts the attention to incumbents to reduce self-selection issues and increase firm comparability. In doing so, we spotlight 999 firms potentially exposed to the persistent demand surge equally unexpected. Indeed, they could not perfectly anticipate in FY2008 the recession—the first GDP decline was registered at the end of summer 2008, that is, shortly before the start of FY2009—and the resulting introduction of the Recovery Act. The same applies to the launch of Obama's veteran policy, as presidential elections came in November 2008 (second month of FY 2009).

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

Figure 2.6. Number of Bids in VBSA vs Other Set-Asides for Contracts Awarded by the DVA, Construction



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

(2021) has shown that 30 percent of service-disabled veterans of the Vietnam War have transitioned into self-employment.

2.6.2 Long-run Implications of VBSAs

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

(1) *Log(VBSA Total Sales)*: the log total sales to the government through VBSA contracts;
(2) *Set-Aside Share*: the share of set-aside sales over total procurement sales;
(3) *# of States*: the number of U.S. states in which the firm performs its contracting activity, weighted by the yearly number of contracts won by a firm;
(4) *# of Categories*: the number of different procurement categories associated with the firm sales, weighted by the yearly number of contracts won by a firm.

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

The Event Study. To identify the effect of permanent variation in VBSA demand on incumbents' outcomes, we employ an event study framework comparing incumbents *differentially exposed* to the demand shock. The idea is that the ex-post dynamics of the less exposed incumbent firms trace out the dynamics that we would have observed in the more exposed treated observations absent the treatment.[37] We define the exposure as $Share_i^{pre}$, measured as the annual average non-VBSA sales over total procurement sales for the five years before the treatment (i.e., between FY2004 and FY2008). Formally, this is equivalent to

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

Our hypothesis is that incumbents that rely less heavily on VBSAs before the demand surge have more scope to be affected than incumbents that rely less on it. The logic is that, even

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

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

Figure 2.A.4 shows the frequency of $Share_i^{pre}$. About 26.7% of the firms exhibit a share equal to zero, meaning that the entirety of their public revenues stems from VBSA awards. Instead, 40.9% obtain all their revenues without winning any VBSA contracts. The median firm's share is 75%. This descriptive evidence further highlights variation in the exposure useful for identification.

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

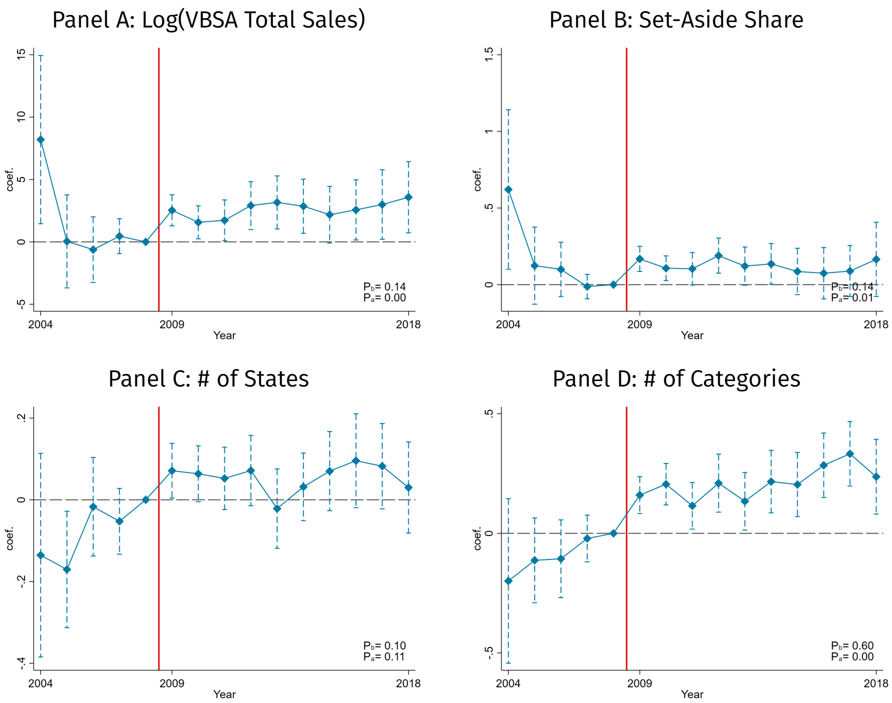
$$W_{i,t} = \theta_i + \iota_t + \sum_{T \neq 2008} b_t \cdot 1(t = T) \times Share_i^{pre} + \omega_{i,t}, \tag{2.3}$$

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

The Long-run Effects. Panels A through D of Figure 2.7 report a visual representation of the $\hat{b_t}$'s for each outcome, where FY2008 is chosen as the base year $t-1$. By observing the 95-percent confidence interval in the figures, we see that all but two coefficients before the treatment across the four outcomes are statistically insignificant. Furthermore, we test the absence of pre-trends non-parametrically by checking the joint statistical significance of the point estimates. On the bottom-right of each figure, we report P_b, the p-value of the joint test on the pre-treatment coefficients, which does not reject the parallel trend assumption for any outcome.

Panel A of Figure 2.7 confirms our exposure assumption and provides "first stage" evidence. Indeed, we find that the higher the exposure, the more the incumbents increase their VBSA procurement sales after the demand surge. They do so permanently as the shock. We report the results for *Set-Aside Share* in Panel B. We find evidence that exposed firms increase their share of set-aside sales. This suggests that the intensity of unexpected VBSA demand exposure does not lead to a size or efficiency effects for recipient firms. We consider the impact on our proxies for scale, i.e., *# of States* and *# of Categories*, in Panels C and D, respectively. We find (weak) evidence that exposed firms become less localized. Recipients sell additional product categories, as the coefficients are jointly statistically significant in the post-period. Jointly considered, the latter results suggest a positive scale effect.

Figure 2.7. Firm-Level Event Study



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

Robustness Checks. We propose two exercises for the robustness of these results. First, the results remain virtually unchanged when we focus on post-period contracts with more than one bid, since they are more likely to have been won unexpectedly. The graphical evidence can be found in Figure 2.A.6 in Appendix 2.A. Second, we replicate Panel C separately for each U.S. region as defined by the U.S. Census. Indeed, it might be easier for a firm to enter new states if located in regions consisting of several small states (e.g., the Northeast) than in regions with a few large states (e.g., the West). We report our results in Figure 2.A.5 in Appendix 2.A. We find similar effects in all regions, i.e., a (weak) increase in the number of states.

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

2.7 Conclusions

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

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

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

References

Abadie, Alberto, Susan Athey, Guido W Imbens, and Jeffrey M Wooldridge. 2023. “When should you adjust standard errors for clustering?” *Quarterly Journal of Economics* 138 (1): 1–35. [74]

Ağca, Şenay, and Deniz Igan. 2023. “The Lion’s Share: Evidence from Federal Contracts on the Value of Political Connections.” *Journal of Law and Economics* 66 (3): 609–38. [82]

Akcigit, Ufuk, Salomé Baslandze, and Francesca Lotti. 2023. “Connecting to Power: Political Connections, Innovation, and Firm Dynamics.” *Econometrica* 91 (2): 529–64. https://doi.org/https://doi.org/10.3982/ECTA18338. eprint: https://onlinelibrary.wiley.com/doi/pdf/10.3982/ECTA18338. [82]

Alcalde, José, and Matthias Dahm. 2024. “On the trade-off between supplier diversity and cost-effective procurement.” *Journal of Economic Behavior and Organization* 217: 63–90. https://doi.org/https://doi.org/10.1016/j.jebo.2023.10.038. [60]

Andini, Monica, Michela Boldrini, Emanuele Ciani, Guido De Blasio, Alessio D’Ignazio, and Andrea Paladini. 2022. “Machine learning in the service of policy targeting: the case of public credit guarantees.” *Journal of Economic Behavior & Organization* 198: 434–75. [72]

Andini, Monica, Emanuele Ciani, Guido de Blasio, Alessio D’Ignazio, and Viola Salvestrini. 2018. “Targeting with machine learning: An application to a tax rebate program in Italy.” *Journal of Economic Behavior & Organization* 156: 86–102. [72]

Angrist, D. 1998. “Estimating the Labor Market Impact of Voluntary Military Service Using Social Security Data on Military Applicants.” *Econometrica 66: 2.* [74]

Antulov-Fantulin, Nino, Raffaele Lagravinese, and Giuliano Resce. 2021. “Predicting bankruptcy of local government: A machine learning approach.” *Journal of Economic Behavior & Organization* 183: 681–99. [72]

Athey, Susan, Dominic Coey, and Jonathan Levin. 2013. “Set-asides and subsidies in auctions.” *American Economic Journal: Microeconomics* 5 (1): 1–27. [58, 62, 89]

Austin, Peter C. 2009. “Balance diagnostics for comparing the distribution of baseline covariates between treatment groups in propensity-score matched samples.” *Statistics in medicine* 28 (25): 3083–107. [79]

Austin, Peter C. 2011. “An introduction to propensity score methods for reducing the effects of confounding in observational studies.” *Multivariate Behavioral Research* 46 (3): 399–424. [106]

Austin, Peter C, and Elizabeth A Stuart. 2015. “Moving towards best practice when using inverse probability of treatment weighting (IPTW) using the propensity score to estimate causal treatment effects in observational studies.” *Statistics in medicine* 34 (28): 3661–79. [106]

Bajari, Patrick, Stephanie Houghton, and Steven Tadelis. 2014. “Bidding for incomplete contracts: An empirical analysis of adaptation costs.” *American Economic Review* 104 (4): 1288–319. [58, 61, 68, 109]

Bandiera, Oriana, Andrea Prat, and Tommaso Valletti. 2009. “Active and Passive Waste in Government Spending: Evidence from a Policy Experiment.” *American Economic Review* 99 (4): 1278–308. [61]

Beheshti, David. 2022. “The impact of opioids on the labor market: Evidence from drug rescheduling.” *Journal of Human Resources.* [86]

Best, Michael Carlos, Jonas Hjort, and David Szakonyi. 2023. “Individuals and Organizations as Sources of State Effectiveness.” *American Economic Review* 113 (8): 2121–67. https://doi.org/10.1257/aer.20191598. [61, 67]

Beuve, Jean, Marian W Moszoro, and Stéphane Saussier. 2019. “Political contestability and public contract rigidity: An analysis of procurement contracts.” *Journal of Economics & Management Strategy* 28 (2): 316–35. [58]

Breiman, Leo. 2001. “Random forests.” *Machine Learning* 45 (1): 5–32. [102, 103]

Breiman, Leo, and Adele Cutler. 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.* [72, 104]

Bruce, Joshua R, John M de Figueiredo, and Brian S Silverman. 2019. "Public contracting for private innovation: Government capabilities, decision rights, and performance outcomes." *Strategic Management Journal* 40 (4): 533–55. [74]

Butler, Jeffrey V, Enrica Carbone, Pierluigi Conzo, and Giancarlo Spagnolo. 2020. "Past performance and entry in procurement: An experimental investigation." *Journal of Economic Behavior and Organization* 173: 179–95. [77]

Caliendo, Marco, and Sabine Kopeinig. 2008. "Some practical guidance for the implementation of propensity score matching." *Journal of Economic Surveys* 22 (1): 31–72. [59, 79, 107]

Calveras, Aleix, Juan-Jose Ganuza, and Esther Hauk. 2004. "Wild bids. Gambling for resurrection in procurement contracts." *Journal of Regulatory Economics* 26: 41–68. [77]

Calvo, Eduard, Ruomeng Cui, and Juan Camilo Serpa. 2019. "Oversight and Efficiency in Public Projects: A Regression Discontinuity Analysis." *Management Science* 65 (12): 5651–75. https://doi.org/10.1287/mnsc.2018.3202. [102, 109]

Calzolari, Giacomo, and Giancarlo Spagnolo. 2009. "Relational Contracts and Competitive Screening." *CEPR Discussion Papers,* https://ideas.repec.org/p/cpr/ceprdp/7434.html. [61]

Cantillon, Estelle. 2008. "The effect of bidders' asymmetries on expected revenue in auctions." *Games and Economic Behavior* 62 (1): 1–25. [65]

Cappelletti, Matilde, Leonardo M Giuffrida, and Gabriele Rovigatti. 2024. "Procuring survival." *Journal of Industrial Economics* 72 (4): 1451–506. [62]

Carril, Rodrigo, et al. 2021. "Rules versus discretion in public procurement." *Barcelon GSE Working Paper Series,* no. 1232. [102]

Carril, Rodrigo, and Mark Duggan. 2020. "The impact of industry consolidation on government procurement: Evidence from Department of Defense contracting." *Journal of Public Economics* 184 (C). https://EconPapers.repec.org/RePEc:eee:pubeco:v:184:y:2020:i:c:s0047272720300050. [61]

Carril, Rodrigo, Andres Gonzalez-Lira, and Michael S. Walker. 2022. "Competition under incomplete contracts and the design of procurement policies." *Economics Working Papers. Department of Economics and Business, Universitat Pompeu Fabra,* https://ideas.repec.org/p/upf/upfgen/1824.html. [61, 67]

Carril, Rodrigo, and Audrey Guo. 2023. "The impact of preference programs in public procurement: Evidence from veteran set-asides." *BSE Working Papers,* no. 1417. [62]

Coile, Courtney, Mark Duggan, and Audrey Guo. 2021. "To Work for Yourself, for Others, or Not At All? How disability benefits affect the employment decisions of older veterans." *Journal of Policy Analysis and Management* 40 (3): 686–714. [84]

Coviello, Decio, Andrea Guglielmo, and Giancarlo Spagnolo. 2017. "The effect of discretion on procurement performance." *Management Science* 64 (2): 715–38. [61]

Coviello, Decio, Immacolata Marino, Tommaso Nannicini, and Nicola Persico. 2022. "Demand Shocks and Firm Investment: Micro-evidence from fiscal retrenchment in Italy." *Economic Journal* 132 (642): 582–617. [62, 86]

Czarnitzki, Dirk, Paul Hünermund, and Nima Moshgbar. 2020. "Public Procurement of Innovation: Evidence from a German Legislative Reform." *International Journal of Industrial Organization* 71: 102620. https://doi.org/https://doi.org/10.1016/j.ijindorg.2020.102620. [62]

De Silva, Dakshina G, Timothy Dunne, Georgia Kosmopoulou, and Carlos Lamarche. 2012. "Disadvantaged business enterprise goals in government procurement contracting: An analysis of bidding behavior and costs." *International Journal of Industrial Organization* 30 (4): 377–88. [62]

De Silva, Dakshina G, Georgia Kosmopoulou, and Carlos Lamarche. 2017. "Subcontracting and the survival of plants in the road construction industry: A panel quantile regression analysis." *Journal of Economic Behavior & Organization* 137: 113–31. [62]

Decarolis, Francesco. 2014. "Awarding Price, Contract Performance and Bids Screening: Evidence from Procurement Auctions." *American Economic Journal: Applied Economics* 6 (1): 108–32. [58, 61, 68]

Decarolis, Francesco. 2018. “Comparing Public Procurement Auctions.” *International Economic Review* 59 (2): 391–419. [61]

Decarolis, Francesco, Leonardo M Giuffrida, Elisabetta Iossa, Vincenzo Mollisi, and Giancarlo Spagnolo. 2020. “Bureaucratic Competence and Procurement Outcomes.” *Journal of Law, Economics, and Organization* 36 (3): 537–97. [61, 109]

Decarolis, Francesco, Gaétan de Rassenfosse, Leonardo M. Giuffrida, Elisabetta Iossa, Vincenzo Mollisi, Emilio Raiteri, and Giancarlo Spagnolo. 2021. “Buyers’ role in innovation procurement: Evidence from US military R&D contracts.” *Journal of Economics & Management Strategy* 30 (4): 697–720. [61, 67, 74, 109]

Denes, Thomas A. 1997. “Do small business set-asides increase the cost of government contracting?” *Public Administration Review,* 441–44. [58, 62]

Di Giovanni, Julian, Manuel García-Santana, Priit Jeenas, Enrique Moral-Benito, and Josep Pijoan-Mas. 2022. “Government procurement and access to credit: firm dynamics and aggregate implications.” [62]

Dilger, Robert J. 2024. “An Overview of Small Business Contracting.” Report R45576. Congressional Research Service. https://crsreports.congress.gov/product/pdf/R/R45576. [59, 64]

Dupraz, Yannick. 2013. “Using weights in Stata.” Working paper. Paris School of Economics. [74]

EC. 2017. *European Semester: Thematic Factsheet–Public Procurement.* [57]

Fadic, Milenko. 2020. “Letting luck decide: Government procurement and the growth of small firms.” *Journal of Development Studies* 56 (7): 1263–76. [62]

Ferraz, Claudio, Frederico Finan, and Dimitri Szerman. 2015. “Procuring firm growth: the effects of government purchases on firm dynamics.” Working paper. National Bureau of Economic Research. [62]

Fischbacher, Urs, Christina M Fong, and Ernst Fehr. 2009. “Fairness, errors and the power of competition.” *Journal of Economic Behavior and Organization* 72 (1): 527–45. [61]

Foster, Lucia, John Haltiwanger, and Chad Syverson. 2016. “The Slow Growth of New Plants: Learning about Demand?” *Economica* 83 (329): 91–129. https://doi.org/https://doi.org/10.1111/ecca.12172. eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/ecca.12172. [62]

Gabriel, Ricardo Duque. 2024. “The credit channel of public procurement.” *Journal of Monetary Economics* 147: 103601. [62]

Gerardino, Maria Paula, Stephan Litschig, and Dina Pomeranz. 2017. “Can audits backfire? Evidence from public procurement in Chile.” *NBER Working Paper Series,* no. 23978. [61]

Gibson, John, and David McKenzie. 2014. “The development impact of a best practice seasonal worker policy.” *Review of Economics and Statistics* 96 (2): 229–43. [59, 71]

Giuffrida, Leonardo M, and Gabriele Rovigatti. 2022. “Supplier selection and contract enforcement: Evidence from performance bonding.” *Journal of Economics & Management Strategy* 31 (4): 980–1019. [77, 102, 109]

Goldman, Jim. 2020. “Government as customer of last resort: The stabilizing effects of government purchases on firms.” *Review of Financial Studies* 33 (2): 610–43. [62]

Goller, Daniel, Michael Lechner, Andreas Moczall, and Joachim Wolff. 2020. “Does the estimation of the propensity score by machine learning improve matching estimation? The case of Germany’s programmes for long term unemployed.” *Labour Economics* 65: 101855. [59, 72]

Gugler, Klaus, Michael Weichselbaumer, and Christine Zulehner. 2020. “Employment behavior and the economic crisis: Evidence from winners and runners-up in procurement auctions.” *Journal of Public Economics* 182: 104112. [62]

Hart, Oliver, and John Moore. 1988. “Incomplete Contracts and Renegotiation.” *Econometrica* 56 (4): 755–85. Accessed October 17, 2023. http://www.jstor.org/stable/1912698. [58, 100]

Hart, Oliver D, and Bengt Holmstrm. 1986. “The theory of contracts.” [57]

Hebous, Shafik, and Tom Zimmermann. 2021. "Can government demand stimulate private investment? Evidence from U.S. federal procurement." *Journal of Monetary Economics* 118: 178–94. https://doi.org/https://doi.org/10.1016/j.jmoneco.2020.09.005. [62]

Heckman, James J, and Richard Robb Jr. 1985. "Alternative methods for evaluating the impact of interventions: An overview." *Journal of Econometrics* 30 (1-2): 239–67. [74]

Hersh, Jonathan, Bree J Lang, and Matthew Lang. 2022. "Car accidents, smartphone adoption and 3G coverage." *Journal of Economic Behavior & Organization* 196: 278–93. [72]

Herweg, Fabian, and Marco A Schwarz. 2018. "Optimal Cost Overruns: Procurement Auctions with Renegotiation." *International Economic Review* 59 (4): 1995–2021. [58, 68]

Ho, Daniel E, Kosuke Imai, Gary King, and Elizabeth A Stuart. 2007. "Matching as nonparametric preprocessing for reducing model dependence in parametric causal inference." *Political analysis* 15 (3): 199–236. [79]

Imbens, Guido W, and Donald B Rubin. 2015. *Causal inference in statistics, social, and biomedical sciences.* Cambridge university press. [79, 81, 96, 106]

Jehiel, Philippe, and Laurent Lamy. 2015. "On Discrimination in Auctions with Endogenous Entry." *American Economic Review* 105 (8): 2595–643. https://doi.org/10.1257/aer.20131580. [65]

Jehiel, Philippe, and Laurent Lamy. 2020. "On the Benefits of Set-Asides." *Journal of the European Economic Association* 18 (4): 1655–96. [89]

Kang, Karam, and Robert A Miller. 2022. "Winning by Default: Why is There So Little Competition in Government Procurement?" *Review of Economic Studies* 89 (3): 1495–556. [61, 109]

Krasnokutskaya, Elena, and Katja Seim. 2011. "Bid preference programs and participation in highway procurement auctions." *American Economic Review* 101 (6): 2653–86. [61]

Laffont, Jean-Jacques, and Jean Tirole. 1993. *A theory of incentives in procurement and regulation.* MIT press. [57]

Lee, Brian K, Justin Lessler, and Elizabeth A Stuart. 2010. "Improving propensity score weighting using machine learning." *Statistics in Medicine* 29 (3): 337–46. [59, 71, 72]

Lee, Munseob. 2021. "Government purchases and firm growth." *Available at SSRN 3823255.* [62]

Li, Tong, and Xiaoyong Zheng. 2009. "Entry and Competition Effects in First-Price Auctions: Theory and Evidence from Procurement Auctions." *Review of Economic Studies* 76 (4): 1397–429. [65]

Liaw, Andy, Matthew Wiener, et al. 2002. "Classification and regression by randomForest." *R news* 2 (3): 18–22. [72, 80]

Liebman, Jeffrey B, and Neale Mahoney. 2017. "Do expiring budgets lead to wasteful year-end spending? Evidence from federal procurement." *American Economic Review* 107 (11): 3510–49. [61, 101]

Lippmann, Quentin. 2021. "Are gender quotas on candidates bound to be ineffective?" *Journal of Economic Behavior & Organization* 191: 661–78. [59, 72]

Marion, Justin. 2007. "Are bid preferences benign? The effect of small business subsidies in highway procurement auctions." *Journal of Public Economics* 91 (7-8): 1591–624. [58, 61]

Marion, Justin. 2009. "How costly is affirmative action? Government contracting and California's Proposition 209." *Review of Economics and Statistics* 91 (3): 503–22. [58, 61]

McCaffrey, Daniel F, Greg Ridgeway, and Andrew R Morral. 2004. "Propensity score estimation with boosted regression for evaluating causal effects in observational studies." *Psychological methods* 9 (4): 403. [79]

Nakabayashi, Jun. 2013. "Small business set-asides in procurement auctions: An empirical analysis." *Journal of Public Economics* 100: 28–44. [58, 62]

OECD. 2019. *Health at a Glance 2019: OECD Indicators.* Paris: OECD Publishing. https://doi.org/10.1787/4dd50c09-en. [57]

Pozzi, Andrea, and Fabiano Schivardi. 2016. "Demand or productivity: What determines firm growth?" *RAND Journal of Economics* 47 (3): 608–30. https://doi.org/https://doi.org/10.1111/1756-2171.12142. eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/1756-2171.12142. [62]

Rosa, Benjamin V. 2019. "Resident bid preference, affiliation, and procurement competition: Evidence from new mexico." *Journal of Industrial Economics* 67 (2): 161–208. [61]

Rosa, Benjamin V. 2020. "Subcontracting requirements and the cost of government procurement." *RAND Journal of Economics.* [62]

Shagbazian, Gegam, Paola Valbonesi, Andrey Tkachenko, and Elena Shadrina. 2025. "Set-aside auctions and small businesses' participation in public procurement: an empirical analysis." *Applied Economics,* 1–15. [58, 62]

Spiller, Pablo T. 2008. "An Institutional Theory of Public Contracts: Regulatory Implications." *National Bureau of Economic Research,* no. 14152, https://ideas.repec.org/p/nbr/nberwo/14152.html. [58, 68]

Spulber, Daniel F. 1990. "Auctions and Contract Enforcement." *Journal of Law, Economics, and Organization* 6 (2): 325–44. [61]

Strobl, Carolin, Anne-Laure Boulesteix, Achim Zeileis, and Torsten Hothorn. 2007. "Bias in random forest variable importance measures: Illustrations, sources and a solution." *BMC bioinformatics* 8 (1): 25. [105]

Stuart, Elizabeth 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. [79]

Williamson, Oliver E. 1971. "The vertical integration of production: market failure considerations." *American Economic Review* 61 (2): 112–23. [58]

Wooldridge, Jeffrey M. 2010. *Econometric analysis of cross section and panel data.* MIT press. [78, 79]

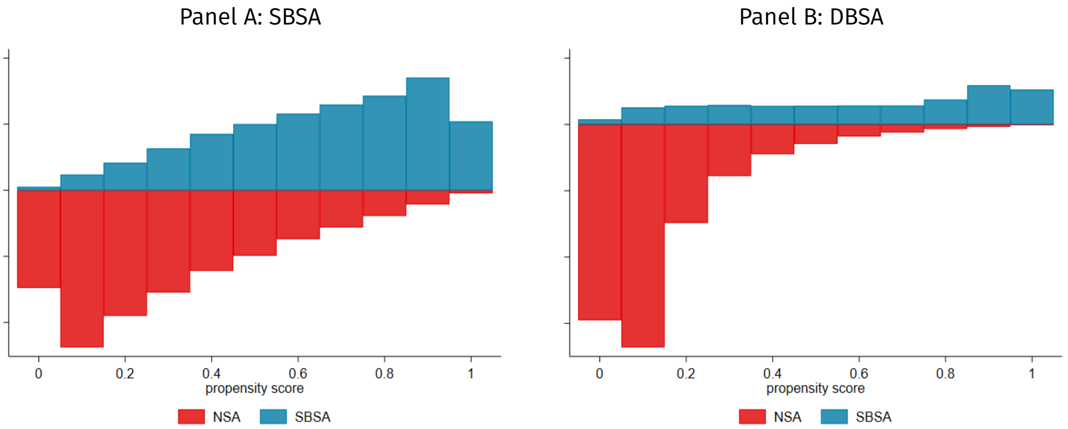
Zhao, Peng, Xiaogang Su, Tingting Ge, and Juanjuan Fan. 2016. "Propensity score and proximity matching using random forest." *Contemporary Clinical Trials* 47: 85–92. [71, 79]

Zhao, Zhong. 2004. "Using matching to estimate treatment effects: Data requirements, matching metrics, and Monte Carlo evidence." *Review of Economics and Statistics* 86 (1): 91–107. [104]

Appendices to Chapter 2

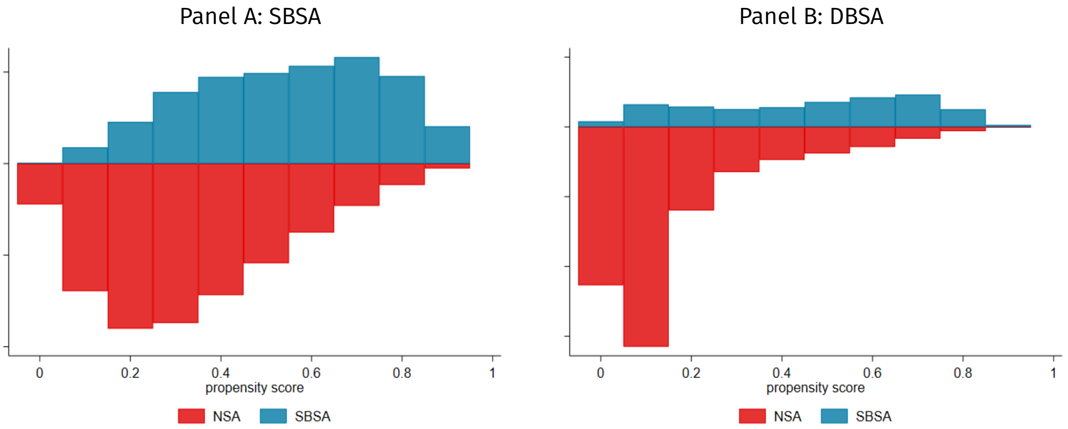
Appendix 2.A Additional Figures and Tables

Figure 2.A.1. Propensity Score Distribution for the Treated and the Untreated



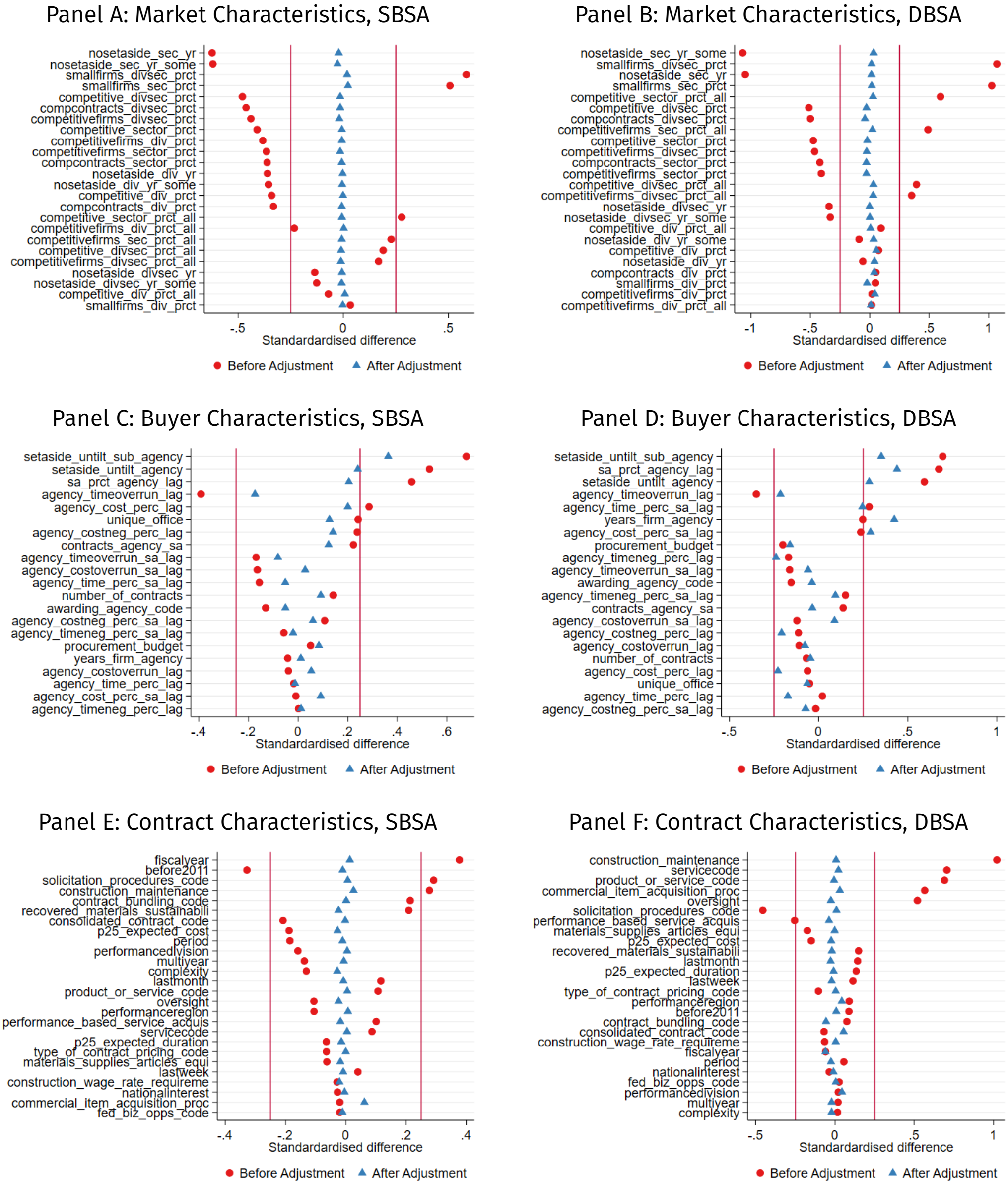
Notes: This figure reports the distribution of treated and untreated observations according to the value of the propensity score. This figure provides some evidence that the overlap assumption is satisfied, since we do not observe values of the propensity score that are taken by one of the two groups.

Figure 2.A.2. Propensity Score Distribution for the Treated and the Untreated, with Logit



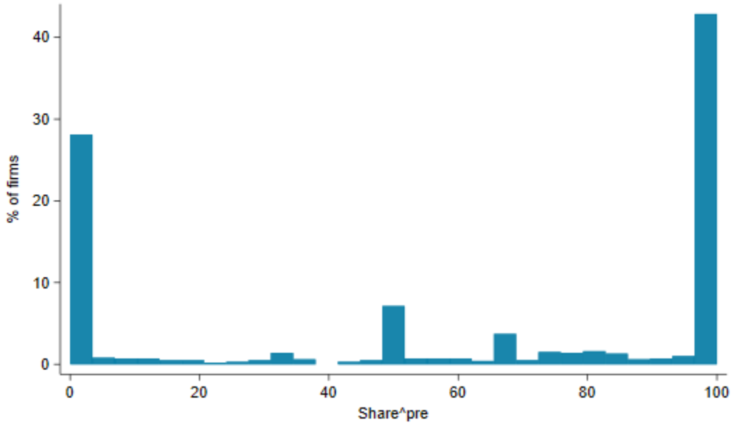
Notes: This figure reports the distribution of treated and untreated observations according to the value of the propensity score. The latter is estimated using logit. This figure provides some evidence that the overlap assumption is satisfied, since we do not observe values of the propensity score that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

Figure 2.A.3. Standardized Differences, with Logit



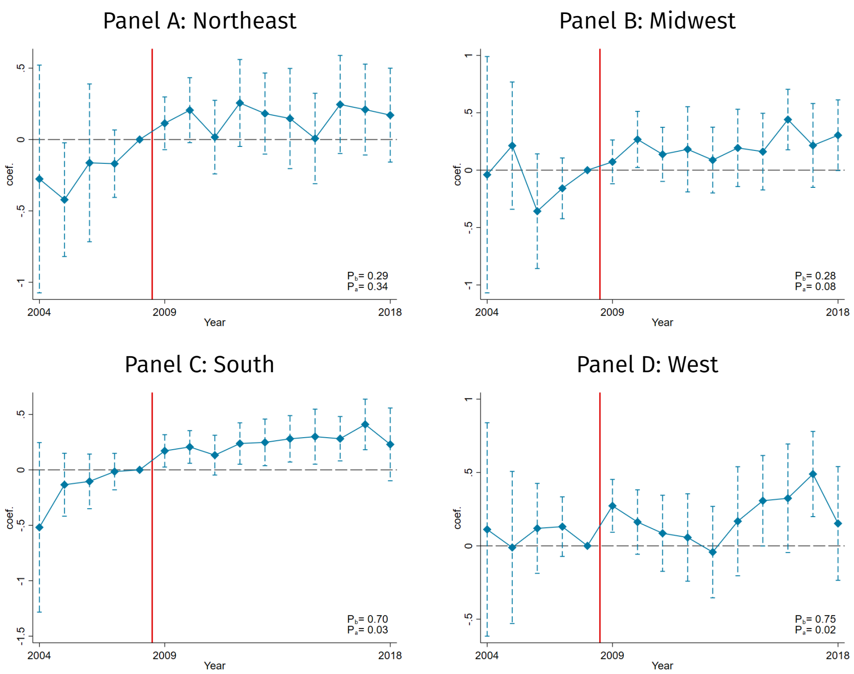
Notes: This figure reports the standardized differences before and after the IPW. The propensity score is estimated using logit. The treatment for this sample is specified in the panel title; the control is non set-aside contract. We predict the propensity score using 71 variables. We report variables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

Figure 2.A.4. Firm Exposure to VBSA Spending Increase



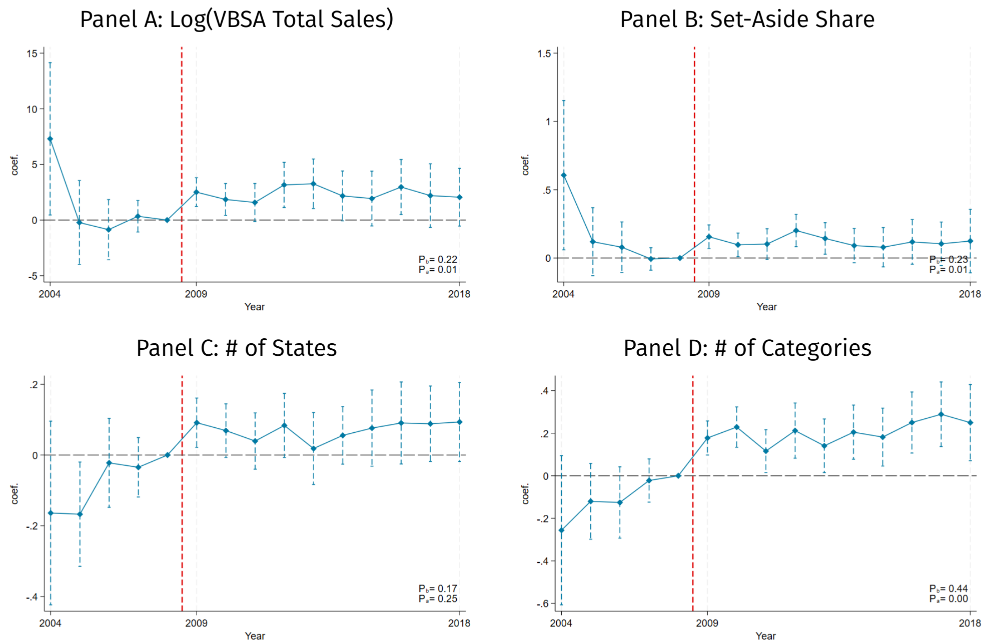
Notes: This figure reports the distribution of $Share_i^{pre}$, i.e., the annual average non-VBSA revenues out of total procurement revenues for VBSA incumbents between FY2004 and FY2008.

Figure 2.A.5. Event Study: # of States by U.S. Region



Notes: Results for the event study analysis, reporting the coefficients of the interaction of $Share_{it}^{pre}$ with FY FEs. The policy was enacted in FY2009 (at time *t*), shown by the red vertical line. FY2008, time *t-1*, is chosen as the base year and all reported coefficients are relative to it. The dependent variables are constructed at the firm-year level. Panels A–D present the results for the number of states in which each firm operates, split by the U.S. region of the firm's headquarters. We report the 95 percent confidence intervals using standard errors clustered at the firm level. P_b reports the joint-significance of the coefficients in the before period, i.e. before FY2009, while P_a in the after period.

Figure 2.A.6. Event Study, Competitive Awards Only



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

Table 2.A.1. VBSA vs NSA

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

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

Appendix 2.B Additional Information on Set-asides

Definition of a Small Business. The Small Business Authority defines a vendor as small based on the particular service category to which the contract belongs and its characteristics, whether revenues or revenues and number of employees. The list is revised every five year. The NAICS code defines the industry. For example, consider a company with 700 employees that is involved in “Wind Electric Power Generation" (NAICS 221115) or “Tire Manufacturing except retreading" (NAICS 326211). The firm would not be considered small in the former case because it has more than 250 employees (i.e., the threshold for this NAICS category). However, in the latter category, it would be considered small because it has fewer than 1,500 employees.[38]

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

Table 2.B.1. Types of Set-Asides

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

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

Table 2.B.1 shows that 53.1% of contracts in the working sample are awarded through some type of set-aside procurement, 38.3% through SBSA, and 14.7% through DBSA. These percentages change for the DVA, mostly due to the VBSA program—a major DBSA category: only 12.8% are awarded through SBSA, and 46.8% are awarded through DBSA.

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

Appendix 2.C Additional Details on the Contract-level Empirical Strategy

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

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

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

  We define a contract as *competitive* if (i) it is awarded without set-aside and (ii) at least two bids were submitted. A targeted firm is competitive if it secures competitive contracts. Our rational is that in a market with many (few) competitive targeted firms, the Rule of Two will be more (less) likely satisfied. Consequently, agencies might be more likely to set aside contract in such markets. Therefore, we first construct proxies for the competitiveness of a market, such as the share of contracts that are competitive in a given market. In addition, we calculate different proxies looking at small firms. For instance, we calculate percentage of competitive small businesses relative to all small businesses. We also control for the percentage of small firms winning without set-aside, as well as the percentage of small firms in a given market.
- *Buyer characteristics.* With buyer characteristics, we are approximating agency quality, experience, as well as set aside culture and goals. It is often impractical to specify fully complete procurement contracts (Hart and Moore, 1988). However, some buyers may

be better at drafting more complete contracts, resulting in less renegotiations and more bids. Competence in drafting contracts could be correlated with a better knowledge in setting contracts aside, with consequent effects on the contract outcomes. Therefore, agency's past experience should be considered for the treatment prediction. We construct a proxy that takes into account the agency's experience with set-asides to control for the contracting agency's performance on past delays and additional costs on awarded contracts. We include the proportion of contracts where cost or schedule overruns were reported prior to each award in *all* set-aside contracts.

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

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

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

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

Finally, we control for a large number of contract characteristics that are reported in FDPS, for instance, *Nationalinterest* indicates whether the contract was created for the national interest. This might be relevant, as such contracts might be less likely to be set-aside to, e.g., quickly limit damages provoked by hurricanes. We also observe whether for the contract some kind of environmental-related clause is applied. *Oversight* is a dummy variable that has a value of 1 if the contract is subject to oversight by the agency (in addition to the surety company on performance bonds if the contract is construction-related).[39]

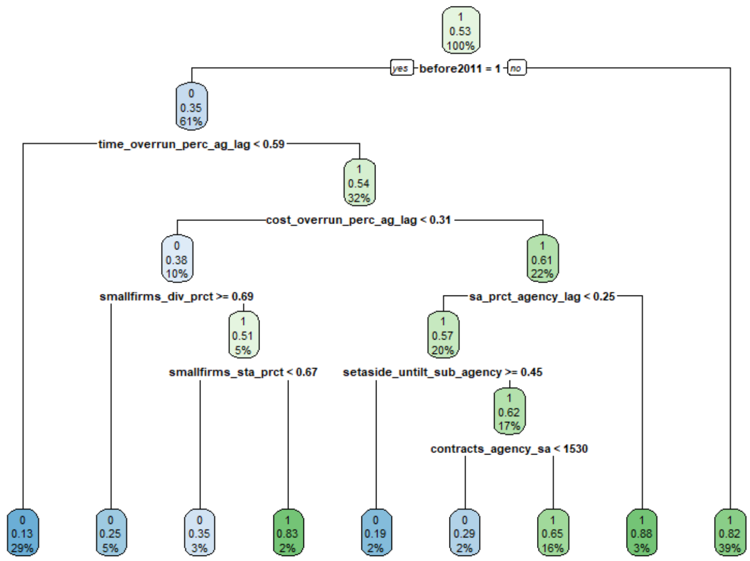
The Random Forest. The random forest is a machine learning method developed by Breiman (2001). Its goal is to predict outcomes based on the available covariates. For instance, it can be used to predict the selling price of a house given its characteristics, such as the number of rooms and the location.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

Important features of the trees are *nodes* and *branches*, whose meaning can best be illustrated with the help of Figure 2.C.1, where the treatment is SBSA. Starting from the green ellipse on top of the figure, “Before2011 = 1" is the first *node* or *split*. At this point, 100% of the sample is at this node, and 53% of the sample has received the treatment. If the statement is true, namely, if the contract is awarded before 2011, i.e. *Before2011* equals 1, then we move to the left *branch*. We reach the second node, where we find 61% of the sample, and 35% of it has received the treatment. At this node, we can split the sample by

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

asking another question, namely, whether the given variable takes a value below 0.59. We continue until we have reached a final node at the bottom of the figure. In case the variable takes a value below 0.59 for a given contract, we predict that the contract will not be set aside.

Figure 2.C.1. Example of a Classification Tree



Notes: This decision tree is obtained with the R package "rpart". The tree predicts the probability of obtaining the treatment for each observation (i.e. whether the contract is set aside). The Sample and variables are the same as for the random forest. However, this tree is only reported as an illustrative example to show how decision trees *might* be built. Random forests build on classification and regression trees but are more complex and less prone to overfitting. Hence, this tree is not necessarily representative of the trees obtained when running a random forest. For more precise information regarding the variables that are important for random forest, refer to Figure 2.C.2.

Although decision trees can be a helpful tool for prediction, they can often lead to overfitting. Random forests address precisely this concern since they work in the following way: several trees are grown using a different bootstrap sample of the data every time, and then, a majority vote is taken to predict the outcome of each observation. Random forests add an extra level of randomness, as they select a random subset of features at each split (or node). In a decision tree, the variable used at each node is the best *among all variables* in the data. Instead, in a random forest, only a subset of variables is randomly selected at each node, and the ones chosen are the best *among a subset of variables*.[40]

When growing each tree, we obtain an estimate of the error term, called *out-of-bag error*, by predicting the outcome for the observations not in the bootstrap sample (Breiman,

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

2001). This estimate is essential for the random forest, as it is used to determine the importance of each variable. Different ways exist to determine variable importance. In this paper, we use the minimal depth. The lower the number of minimal depth, the more important the variable. The number indicates the depth of the node—the closer the node is to the root of the tree, the lower the minimal depth. Hence, a low minimal depth means that the variable splits many observations into two groups. In the example reported in Figure 2.C.1, *Before2011* would have depth 1.

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

For tuning the optimal number of randomly selected variables at each node, we follow again Breiman and Cutler (2011). The authors suggest trying different numbers of randomly selected variables combined with a relatively small number of trees. An exact number of trees is not provided. Thus, we use 200 trees. We train the data with a sample, which constitutes 20% of the original dataset, and we keep the remaining 80% aside for testing the error prediction. We choose three different numbers of variables to conduct three trials. First, we compute our number of variables as the square root of the total number of variables (i.e., 8 since we have 71 variables). Then, we use half as many variables, and finally, twice as many. For the baseline regressions, we obtain the lowest prediction error for 16 (8) variables for the SBSA (DBSA) treatment, so we use it throughout the analysis. Indeed, it has the highest accuracy, which is the percentage of correct out-of-sample prediction when using the fitted random forest. Note, however, that we estimate the propensity score for the three different numbers of randomly selected variables and obtain stable estimates of the ATT.[41] This shows that the results are not sensitive to the different specifications of the propensity score.[42]

First-stage Results. In this subsection, we identify key variables for the decision to restrict participation to targeted firms. Predicting the propensity score using random forests enables us to do so. In Figure 2.C.2, we report the distribution of the minimal depth for the ten variables that are selected most often in one of the first nodes of the tree. In other words, those are the most critical factors in predicting the treatment as a low number of minimal depth means that a lot of observations are divided into groups based on this variable.

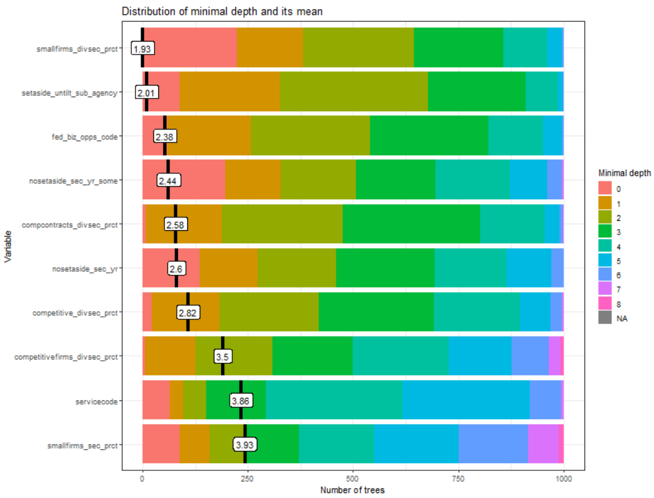
We had previously argued that, following the FAR regulations, the the agency's decision to set-aside a contract is driven by two factors: (i) specific rules based on an assessment of potential competition between targeted firms and (ii) agency–year spending targets for

[41] The results of this test are reported in Figure 2.3.

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

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

Figure 2.C.2. Variable Importance According to Minimal Depth



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

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

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

$$Y(0), Y(1) \perp\!\!\!\perp D\,|X, \forall X \tag{2.C.1}$$

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

$$0 < P(D = 1|X) < 1 \tag{2.C.2}$$

In Section 2.5.1, we have shown that the common support assumption is satisfied given that covariates are balanced. Therefore, we are able to obtain unbiased estimators of the ATT. Note that for checking whether covariates are balanced, we use the standardized difference instead of computing t-statistics or testing the difference in means. The standardized difference “compares the difference in means in units of the pooled standard deviation" (Austin (2011), p. 412). According to Imbens and Rubin (2015), it is formally defined as follows:

$$\hat{\Delta}_{ct} = \frac{\bar{X}_t - \bar{X}_c}{\sqrt{\frac{\sigma_c+\sigma_t}{2}}} \tag{2.C.3}$$

For the weighted mean, we can substitute in Equation (2.C.3): $\bar{W}_{weight} = \frac{\sum \omega_i X_i}{\sum \omega_i}$, where ω_i is the weight for each unit (Austin and Stuart, 2015). The weighted sample variance is defined as:

$$s^2_{weight} = \frac{\sum \omega_i}{(\sum \omega_i)^2 - \sum \omega_i^2} \sum \omega_i (X_i - \bar{X}_{weight})^2 \tag{2.C.4}$$

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

$$\hat{\tau}_{ATT} = N^{-1} \sum_{i=1}^{N} SetAside_i Y_i - \frac{(1 - SetAside_i)\ Y_i\ p_i(X_i)}{1 - p_i(X_i)}, \tag{2.C.5}$$

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

Appendix 2.D Additional Robustness Checks

2.D.1 Robustness to Methodology

In Table 2.D.1, we check whether our takeaways from the baseline analysis are sensitive to changes in methodology. Columns 1-3 report estimates on *Log(# Offers)*, columns 4-6 on *Extra Cost* and columns 7-9 on *Delay*. The top panel reports again the estimated coefficient for the SBSA, while the bottom one reports the estimated coefficients for the DBSA. Columns 1, 4 and 7 report baseline estimates from, respectively, columns 2, 4, and 6 in Table 2.2.

Table 2.D.1. Alternative Second-stage Methods

	Log(# Offers)			Extra Cost			Delay		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	IPW (RF)	PSM (RF)	IPW (logit)	IPW (RF)	PSM (RF)	IPW (logit)	IPW (RF)	PSM (RF)	IPW (logit)
SBSA	**0.213**	**0.298**	**0.261**	**0.035**	**0.021**	**0.078**	**0.047**	**0.058**	**0.142**
	(0.0112)	(0.0084)	(0.0068)	(0.0113)	(0.0081)	(0.0061)	(0.0292)	(0.0219)	(0.0171)
	0.837	0.754	0.875	0.253	0.234	0.231	0.952	0.921	0.927
	74,934	*106,432*	*81,716*	*74,425*	*105,730*	*81,194*	*74,928*	*106,418*	*81,708*
DBSA	**0.184**	**0.252**	**0.257**	**0.075**	**0.054**	**0.098**	**0.099**	**0.083**	**0.221**
	(0.0189)	(0.0192)	(0.0104)	(0.0138)	(0.0191)	(0.0092)	(0.0467)	(0.0515)	(0.0255)
	0.856	0.761	0.837	0.268	0.247	0.245	0.995	0.960	0.947
	46,533	*66,337*	*56,804*	*46,242*	*65,956*	*56,483*	*46,529*	*66,331*	*56,799*
Buyer*FY FE	✓	✓	✓	✓	✓	✓	✓	✓	✓
Market Controls	✓	✓	✓	✓	✓	✓	✓	✓	✓
Contract Controls	✓	✓	✓	✓	✓	✓	✓	✓	✓

Notes: Results for the ATT of the IPW on the three main outcomes, *Log(# Offers)*, *Extra Cost* and *Delay*. We report, for each outcome, the baseline results as in Table 2.2 (i.e., IPW (RF)). Column 2, 5, and 8 reports the traditional propensity score matching (PSM) approach with the previously estimated propensity score. Note that Stata 17 user-written program `psmatch2` does not allow to specify both a variable list and a propensity score, so only the propensity score is specified. Column 3, 6, and 9 reports results from IPW estimated with the propensity score predicted with a logistic regression.

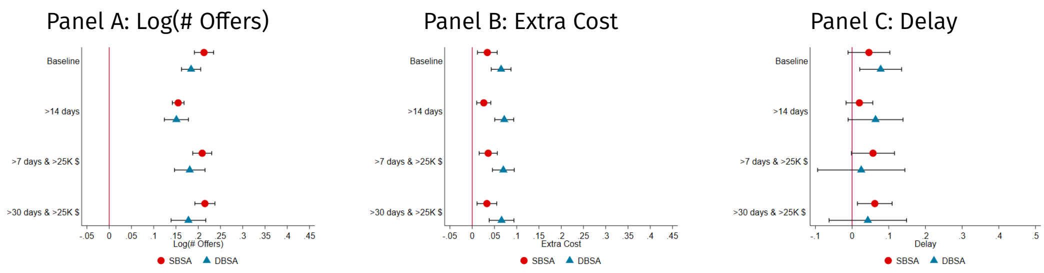
Using the same propensity score obtained estimated with random forest, we first test the robustness of the chosen second-stage estimator—i.e., the IPW. Columns 2, 5, and 8 report the same coefficients using a propensity score matching approach for which we employ a more traditional matching approach: A kernel matching with a biweight distance, which defines a neighborhood for each treated observation and constructs the counterfactual using all control observations within the neighborhood. The method assigns a positive weight to all those observations within the previously defined neighborhood and a zero weight to the remaining observations (Caliendo and Kopeinig, 2008).[43] Finally, in columns 3, 6, and 9, we run the IPW by calculating the propensity score with a classic logistic regression using the same set of variables employed for the random forest prediction. To sum up, on the one hand, the estimates confirm the sensibility of the choice of IPW since it does a better job of balancing the covariates without losing many observations. On the other hand, the results are qualitatively in line with different methods and specifications.

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

2.D.2 Robustness to Sample Selection

We then test the robustness of our results by using a different sample selection in terms of contract amount and duration, as reported in the right-side panels of Figure 2.D.1. We replicate the results for our preferred specification, with fixed effects and additional controls. We recall that our baseline sample described in Section 2.4 includes contracts expected to be worth more than $25,000 *and* last more than 14 days.

Figure 2.D.1. Robustness Checks, Sample



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

The first line reports the baseline estimates from Table 2.2; the second line includes all contracts above 14 days, irrespective of their size; the third adds contracts between 7 and 14 days to the baseline sample; the fourth removes contracts between 14 and 30 days from the baseline sample. Note that we make the sample selection before predicting propensity scores, that is before running the random forest, which is therefore estimated again for each subsample. The results for *Log(# Offers)* are very stable and differ statistically only slightly when including all contracts above 14 days. For the outcome *Extra Cost* we find that the results both qualitatively and quantitatively almost identical. The same is also true for *Delay*, though we find a higher variability in the estimates, which is probably due to the presence of huge outliers in our data. In general, we can conclude that for all the three variables the results for the different samples are not statistically different and hence do not depend on the chosen specification. This is true for both treatments that we consider.

2.D.3 Robustness to Alternative Outcome Specifications

As mentioned in Section 2.4, we also build two secondary outcomes, *Renegotiation* and *# Renegotiations*, which measure the extensive and intensive margins of contract amendment, respectively. *Renegotiation* is an indicator that takes the value 1 if we observe at least one *in-scope* modification that increases the final cost of a given contract.[44] Note that we consider

[44] Given the available data, we cannot proceed in the same way for an increase in final duration due to the construction of the date variable. Indeed, the dataset does not report the reasons of modification for an increase in final duration, it only reports expected end date and actual end date. The construction of these variables is

a renegotiation as *in-scope* if the reason for the modification is consistent with the original contract terms.[45] *Out-of-scope* would be adding a new task to the contract in disregard of the original plan as to type, scope of work, period of performance, and method of performance.

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

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

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

As we report in Table 2.D.2, we observe that set-asides cause the probability for a contract to undergo an amendment to increase by 5.3 and 4.8 p.p.. This is the most reliable point estimate as IPW best performs with dichotomous treatment and outcome variables. The coefficient on *# Renegotiations* is also positive, though not statistically significant for SBSA, while it is negative and significant for DBSA. This finding implies that contracts awarded under DBSA are more likely to be renegotiated, but such renegotiations tend to

similar to Calvo, Cui, and Serpa (2019), Decarolis et al. (2020), Giuffrida and Rovigatti (2022), and Kang and Miller (2022).

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

be larger in size, as we observe an increase in *Extra Cost* and *Delay* and a concomitant decrease in *# Renegotiations*. The results hold as compared to the baseline and tend to speak in favor of a deterioration in contract performance due to the set-aside's restriction to entry—in particular for SBSA—consistent with our baseline execution-stage findings.

Table 2.D.2. Additional Contract Outcomes

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

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

Appendix 2.E Institutional Background

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

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

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

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

7a. ITEM AND/OR SERVICE DESCRIPTION

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

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

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

10. RECOMMENDATION *(X all that apply)*

☐ a. SMALL BUSINESS SET-ASIDE *(X one)* ☐ 100% ☐ Partial ________%	☐ b. SECTION 8(a) *(X one)* ☐ Competitive ☐ Sole Source
☐ c. HISTORICALLY UNDERUTILIZED BUSINESS ZONE (HUBZone) SMALL BUSINESS *(X one)* ☐ Competitive ☐ Sole Source	☐ d. SERVICE-DISABLED VETERAN-OWNED SMALL BUSINESS (SDVOSB) *(X one)* ☐ Competitive ☐ Sole Source
☐ e. ECONOMICALLY DISADVANTAGED WOMEN-OWNED SMALL BUSINESS (EDWOSB) SET-ASIDE	☐ f. WOMEN-OWNED SMALL BUSINESS (WOSB) ELIGIBLE UNDER WOSB PROGRAM SET-ASIDE
☐ g. OTHER SET-ASIDE *(Cite authority, e.g., FAR 26.202-1 or 6.208; or DFARS 226.71)*	☐ h. OTHER THAN FULL AND OPEN COMPETITION NOT PREVIOUSLY ADDRESSED
☐ i. FULL AND OPEN COMPETITION *(Complete block 13)* ☐ HUBZONE PRICE EVALUATION PREFERENCE *(Ref. FAR 19.1307)*	☐ j. MULTIPLE AWARD ☐ Contract ☐ Delivery/Task Order ☐ Reserves *(FAR 19.5) (List type(s) of small business, e.g., WOSB, SDVOSB)*

11a. MARKET RESEARCH/ACQUISITION PLAN

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

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

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

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

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

DD FORM 2579, AUG 2015 PREVIOUS EDITION IS OBSOLETE. **Page 1 of 2**

14. ACQUISITION HISTORY

a. IS THIS A NEW REQUIREMENT? *(X one)*

☐ Yes *(Proceed to Block 15)* ☐ No *(Continue to Blocks a(1) through (10), marking all that apply for the immediately preceding acquisition.)*

☐ (1) SMALL BUSINESS SET-ASIDE *(X one)* ☐ 100% ☐ Partial ________%	☐ (2) SECTION 8(a) *(X one)* ☐ Competitive ☐ Sole Source
☐ (3) HISTORICALLY UNDERUTILIZED BUSINESS ZONE (HUBZone) SMALL BUSINESS *(X one)* ☐ Competitive ☐ Sole Source	☐ (4) SERVICE-DISABLED VETERAN-OWNED SMALL BUSINESS (SDVOSB) *(X one)* ☐ Competitive ☐ Sole Source
☐ (5) ECONOMICALLY DISADVANTAGED WOMEN-OWNED SMALL BUSINESS (EDWOSB) SET-ASIDE	☐ (6) WOMEN-OWNED SMALL BUSINESS (WOSB) ELIGIBLE UNDER WOSB PROGRAM SET-ASIDE
☐ (7) OTHER SET-ASIDE *(Cite authority, e.g., FAR 26.202-1 or 6.208; or DFARS 226.71)*	☐ (8) OTHER THAN FULL AND OPEN COMPETITION NOT PREVIOUSLY ADDRESSED
☐ (9) FULL AND OPEN COMPETITION *(Complete block 13)* ☐ HUBZONE PRICE EVALUATION PREFERENCE *(Ref. FAR 19.1307)*	☐ (10) MULTIPLE AWARD ☐ Contract ☐ Delivery/Task Order ☐ Reserves *(FAR 19.5) (List type(s) of small business, e.g., WOSB, SDVOSB)*

b. PREVIOUSLY CONSOLIDATED OR BUNDLED? *(X one)*

(1) CONSOLIDATED ☐ YES ☐ NO (2) BUNDLED ☐ YES ☐ NO

c. DETAILS OF PREVIOUS AWARD(S) *(List details requested in instructions. Attach additional page(s) if necessary.)*

15. CONTRACTING OFFICER

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

16. SMALL BUSINESS PROFESSIONAL/SMALL BUSINESS DIRECTOR REVIEW

☐ Concur ☐ Non-concur	a. NAME *(Last, First, Middle Initial)*	b. E-MAIL ADDRESS

c. SMALL BUSINESS PROFESSIONAL/SMALL BUSINESS DIRECTOR REMARKS

d. SIGNATURE	e. DATE SIGNED *(YYYYMMDD)*	f. DATE ACQUISITION PACKAGE PROVIDED TO SBA *(FAR 19.202-1(e)) (YYYYMMDD)*

17. SBA PROCUREMENT CENTER REPRESENTATIVE REVIEW

☐ Concur ☐ Non-concur	a. NAME *(Last, First, Middle Initial)*	b. E-MAIL ADDRESS

c. SBA PROCUREMENT CENTER REPRESENTATIVE REMARKS

d. SIGNATURE	e. DATE SIGNED *(YYYYMMDD)*

18. CONTRACTING OFFICER REVIEW

a. CONTRACTING OFFICER REMARKS

☐ Concur with PCR recommendation ☐ Reject PCR recommendation	d. SIGNATURE	e. DATE SIGNED *(YYYYMMDD)*

DD FORM 2579, AUG 2015 PREVIOUS EDITION IS OBSOLETE. **Page 2 of 2**

Full List of Treatment Predictors

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

Market Characteristics.

- What share of contracts is competitive?
  - *compcontracts_div_prct*: Share of competitive contracts out of all contracts, by division, per fiscal year.
  - *compcontracts_sector_prct*: Share of competitive contracts out of all contracts, by sector, per fiscal year.
  - *compcontracts_divsec_prct*: Share of competitive contracts out of all contracts, by division and sector, per fiscal year.
- What share of contracts won by small firms is competitive?
  - *competitive_div_prct*: Share of competitive contracts out of all contracts awarded to small firms, per division and year.
  - *competitive_sector_prct*: Share of competitive contracts out of all contracts awarded to small firms, per sector and year.
  - *competitive_divsec_prct*: Share of competitive contracts out of all contracts awarded to small firms, per division, sector and year.
- What share of competitive contract is won by small firms?
  - *competitive_div_prct_all*: Share of contracts won by small firms out of competitive contracts, per division and year.
  - *competitive_sector_prct_all*: Share of contracts won by small firms out of competitive contracts, per sector and year.
  - *competitive_divsec_prct_all*: Share of contracts won by small firms out of competitive contracts, per division, sector and year.
- What share of small firms is competitive?
  - *competitivefirms_div_prct*: Share of competitive small firms out of all small firms, per division and year.
  - *competitivefirms_sector_prct*: Share of small competitive firms out of all small firms, per sector and year.
  - *competitivefirms_div_prct*: Share of competitive small firms out of all small firms, per division, sector and year.
- What share of competitive firms is small?
  - *competitivefirms_div_prct_all*: Share of competitive small firms out of all competitive firms, per division and year.
  - *competitivefirms_sec_prct_all*: Share of small competitive firms out of all competitive firms, per sector and year.
  - *competitivefirms_divsec_prct_all*: Share of competitive small firms out of all competitive firms, per division, sector and year.
- What share of firms is small and win without set-aside? (always or sometimes)

- *nosetaside_div_yr*: Share of small firms out of all firms that always win without set-aside, per division and year.
- *nosetaside_div_yr_some*: Share of small firms out of all firms that sometimes win without set-aside, per division and year.
- *nosetaside_sec_yr*: Share of small firms out of all firms that always win without set-aside, per sector and year.
- *nosetaside_sec_yr_some*: Share of small firms out of all firms that sometimes win without set-aside, per sector and year.
- *nosetaside_divsec_yr*: Share of small firms out of all firms that always win without set-aside, per division, sector and year.
- *nosetaside_divsec_yr_some*: Share of small firms out of all firms that sometimes win without set-aside, per division, sector and year.

- What is the share of small firms?
  - *smallfirms_div_prct*: Share of small firms per division per year.
  - *smallfirms_sec_prct*: Share of small firms per sector per year.
  - *smallfirms_divsec_prct*: Share of small firms per division and sector per year.

Buyer Characteristics.

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

  - *agency_timeneg_perc_lag*: Lag of ratio of contracts with a negative delay out of all contracts by agency per year.
  - *agency_timeneg_perc_sa_lag*: Lag of ratio of set-aside contracts with a negative delay out of all set-aside contracts by agency per year.
  - *agency_timeoverrun_lag*: Lag of the absolute delay of the agency divided by the absolute sum of the expected duration by agency per year.
  - *agency_timeoverrun_sa_lag*: Lag of the absolute delay of the agency divided by the absolute sum of the expected duration by agency per year, calculated for set-aside contracts only.
- Agency size:
  - *Contracting Offices*: Number of distinct contracting units within an agency.
  - *Contracts*: agency's total yearly number of contracts awarded for all purchase categories.
  - *Procurement Budget*: agency's total yearly procurement spending for all purchase categories.
- Percentage of set-aside in the previous year:
  - *sa_prct_agency_lag*: Ratio of the previous fiscal year, i.e. set-aside contracts out of all contracts awarded by a given agency.
- Set-aside contracts until that award, a proxy for pressure to set-aside a contract to reach the agency goal:
  - *contracts_agency_sa*: Cumulative count of set-aside contracts grouped by awarding agency and year (and sorting by action date).
  - *setaside_untilt_agency*: Cumulative sum of contracts awarded with set-aside out of the cumulative sum of all contracts per agency (contracts are sorted by awarding agency and starting date).
  - *setaside_untilt_sub_agency*: Cumulative sum of contracts awarded with set-aside out of the cumulative sum of all contracts per subagency.
- Number of years in which the same firms wins with same agency, a proxy for relational contracts:
  - *years_firm_agency*: Number of years in which the same firm wins with the same agency.

Contract Characteristics.

- Time dimension:
  - *before2011*: Indicator equal to 1 if the contract was awarded before 2011.
  - *fiscalyear*: Year of the contract award. The fiscal year starts October 1st each calendar year.
  - *lastmonth*: Indicator variable equal to one if the contract was awarded during the last month of the fiscal year, i.e. September.

  - *lastweek*: Indicator variable equal to one if the contract was awarded during the last week of September.
  - *period*: This indicator equals to one if the contract was awarded after December 2007 and before June 2009. It accounts for contracts awarded during the Great Recession.
- Geographical dimension:
  - *performancedivision*: Indicates each of the nine divisions of the place of performance of the contract, i.e., New England, Middle Atlantic, East North Central, West North Central, South Atlantic, East South Central, West South Central, Mountain, Pacific.
  - *performanceregion*: Indicates each of the four regions (West, South, Midwest, and North East) of the place of performance of the contract.
- Sector-related variables:
  - *construction_maintenance*: An indicator variable equal to one if the contract is based on the execution of a construction or maintenance project.
  - *product_or_service_code*: The code that best identifies the product or service procured.
  - *servicecode*: The code that best identifies the product or service procured, but only considering the first letter of the code, such as construction, quality control, information technology, and so on.
- Proxies for complexity:
  - *complexity*: A categorical variable taking the value of one for contracts in the first and second quartile for *both* amount and duration. It is equal to three for the fourth quartile in *both* categories and equals to two in all other cases.
  - *multiyear*: Indicator variable equal to one if the expected duration is over 365 days.
  - *oversight*: An indicator variable equal to one if the contract is subject to oversight by the buyer (in addition to the surety company when the contract is construction-based). Expected costs of above $100,000 was the threshold to apply for oversight until 2011, and expected costs of above $150,000 was the threshold after 2011.
  - *p25_expected_cost*: This variable is a proxy of the complexity of the contract. It places the contract into a quartile according to the expected cost grouped by sector.
  - *p25_expected_duration*: This variable is a proxy of the complexity of the contract. It places the contract into a quartile according to the expected duration grouped by sector.
- Other controls for contract heterogeneity:
  - *commercial_item_acquisition_pr*: Indicates whether the solicitation meets the special requirements for the acquisition of commercial items intended to more closely resemble those customarily used in the commercial marketplace as defined by FAR Part 12.

- *consolidated_contract_code*: Indicates whether the contract is a consolidated contract. This is only 'True' if the Funding Agency or the contracting agency is a Department of Defense Agency.
- *construction_wage_rate_requireme*: An indicator equal to one if the transaction is subject to the Construction Wage Rate Requirements. The latter states that "all laborers and mechanics employed or working upon the site of the work will be paid unconditionally and not less often than once a week, and without subsequent deduction or rebate on any account."
- *contract_bundling_code*: Designates that the value of the contract, including all options, is expected to exceed the threshold whose value is: (1) $5 million until 09/27/2006, (2) $5.5 million from 09/28/2006 to 09/30/2010, (3) $6 million from 10/01/2010. It indicates the reason why the agency bundled contract requirements. 'Bundling' refers to the consolidation of two or more requirements for goods or services previously provided or performed under separate smaller contracts into a solicitation for a single contract that is likely to be unsuitable for award to a small business.
- *fed_biz_opps_code*: Description tag (by way of the FPDS Atom Feed) that explains the meaning of the code provided in the FedBizOpps Field.
- *materials_supplies_articles_eq*: Description tag (by way of the FPDS Atom Feed) that explains the meaning of the code provided in the Contracts for Materials, Supplies, Articles, and Equipment Exceeding $15,000 Field.
- *nationalinterest*: Indicator variable equal to one if the contract is created for the national interest, e.g. projects to limit damages provoked by hurricanes.
- *performance_based_service_acqu*: This variable describes the requirements in terms of results required rather than the methods of performance of the work.
- *recovered_materials_sustainabili*: Designates whether Recovered Material Certification and/or Estimate of Percentage of Recovered Material Content for EPA-Designated Products clauses were included in the contract. For instance, if the contract considered energy efficiency in the award.
- *solicitation_procedures_code*: Report this code for the type of solicitation procedure used.
- *type_of_contract_pricing_code*: The type of contract as defined in FAR Part 16 that applies to this procurement. As we keep only fixed price contracts, these are different types of fixed price, i.e., award fee, incentive, etc.

Chapter 3

Procuring Survival

Matilde Cappelletti, Leonardo Giuffrida, and Gabriele Rovigatti[*]

3.1 Introduction

A firm's life expectancy at birth is low. Across countries and sectors, most startups survive the first year, but less than half remain in the market after seven years (Agarwal and Gort, 2002; Bartelsman, Haltiwanger, and Scarpetta, 2009; Calvino, Criscuolo, and Menon, 2016). These statistics are partially ascribed to the natural selection of the most efficient firms; yet, as demonstrated by the government support packages for firms put in place in response to economic fallouts, business survival has intrinsic value for socioeconomic cohesion. The former observation and the latter consideration have spawned a body of research on the determinants of business survival. In a nutshell, marginal survival probability is robustly found to increase with age and size (Evans, 1987a, 1987b; Hall, 1987; Dunne, Roberts, and Samuelson, 1989; Clementi and Hopenhayn, 2006). Other major identified determinants include idiosyncratic productivity (Ugur and Vivarelli, 2021), industry characteristics (Zingales, 1998), and geography (Choi et al., 2021).

The role of demand constraints on firm dynamics has been less widely analyzed (Syverson, 2011; Foster, Haltiwanger, and Syverson, 2016; Pozzi and Schivardi, 2016). In this paper, we spotlight the *nature* of demand and explore the role of government-based public demand—as opposed to market-based private demand—in determining firm survival. At the macroeconomic level, government spending, its optimal level, and its structural role in guiding the economy have been at the center of debate for decades (Ramey, 2019). Several contributions have shown how shifting the amount of public spending has cascading effects throughout the productive sector, making it the most effective policy tool to prop up the economy during downturns. At the microeconomic level, however, the impact of procurement spending—i.e., a specific component of government outlay explicitly targeted to

[*] Published in *The Journal of Industrial Economics*, 72(4), 1451-1506, 2025. We acknowledge the fruitful collaboration with DBInformation S.p.A. (Telemat Division) on dataset setup. The support from the Leibniz Association for its support through the project 'Market Design by Public Authorities' is gratefully acknowledged. The views expressed herein are those of the authors and do not involve the responsibility of the Bank of Italy.

firms—on business outcomes has been studied only recently (e.g., Ferraz, Finan, and Szerman, 2015; Goldman, 2020; Gugler, Weichselbaumer, and Zulehner, 2020) and its effect on business survival is underinvestigated (De Silva, Kosmopoulou, and Lamarche, 2009).[1]

A priori, a differential survival effect between public and private demand is uncertain, given that both entail revenues. The former does not necessarily entail higher profits than the latter—and profitability tends to be a better predictor of survival than revenues (Jovanovic, 1982). Indeed, higher revenues from public sales could be associated with higher costs due to the administrative burden lowering the profitability of public demand. Government-linked firms may also see fewer incentives to invest in intellectual capital and thus do not become more productive. On the other hand, firms selling to the government may see frictions reduced. For example, they may have easier access to credit since the certainty of a government-backed cash flow decreases their implied risk (Giovanni et al., 2022). Or, focusing on small businesses, public awards may help build a firm's customer base and reputation (Foster, Haltiwanger, and Syverson, 2016).

We address our research question empirically using a novel combination of extensive and highly detailed data on Italy, the laboratory for this study. We combine a decade of individual balance-sheet and income-statement records on the quasi-universe of limited companies with administrative data reporting official business registration (i.e., market entry) and deregistration (i.e., market exit). We match this panel of firms—including records on survival, age, revenues, employment, and labor productivity—with a database on government procurement contracts provided by the National Anti-Corruption Authority (hereafter ANAC), which is the public procurement regulator in the country. The database contains comprehensive information on tenders solicited by any public agency with a value of more than €40 thousand and the related contracts, totaling a cumulative average yearly value of €156 billion—representing 9% of Italian GDP and 90% of total public procurement spending.[2] The data include information on the contract value and duration, the procurement category, the award mechanism, and, importantly, the winner's identity.

Thanks to the granularity of our combined dataset, we can pinpoint firms that receive public money through public procurement contracts ('*procurement firms*') and compare them to firms that receive no such contracts ('*non-procurement firms*'). A puzzling piece of descriptive evidence emerges from Figure 3.1 and motivates our endeavor. Fixing the quantile of industry-specific characteristic distribution (i.e., age, size, or labor productivity) or geographic area (northern and central regions versus southern and island regions)—i.e., accounting for major survival predictors identified by the literature—procurement firms in the data display better survival prospects than non-procurement firms.[3] However, any naive comparison between procurement and non-procurement firms would overlook the crucial

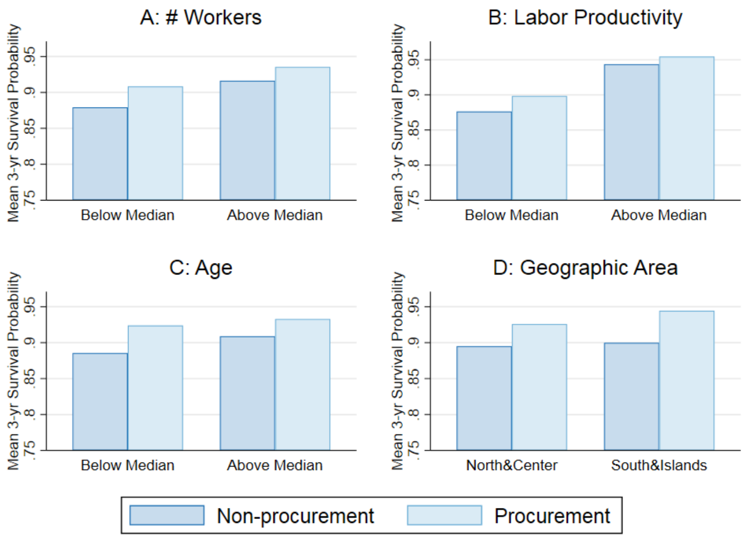
[1] In 2018, government procurement spending in the median OECD country amounted to 13% of GDP and 41% of total government outlay (OECD, 2019).

[2] The reporting threshold is lower than any category-specific EU regulatory threshold and, in particular, much lower than that for public works contracts, which is around €5.5 million in 2019.

[3] For sake of tractability, we distinguish between procurement and non-procurement firms by leveraging public contracts awarded between 2013 and 2016. We compare the share of firms still active in 2019 (i.e., three years later) between the two groups, partialling out each of the four dimensions discussed. A similar qualitative conclusion emerges from our data when we replicate such exercise on other years.

role of unobservable firm characteristics (e.g., management quality, political connections) that correlate with the probability of participating in and winning a public auction on the one hand, and with the ability to stay in the market on the other.

Figure 3.1. Procurement vs. Non-procurement Firms – Average Survival Rate



Notes: We report the average three-year survival rate for procurement and non-procurement firms in 2016. In each panel, we partial out a predictor of survival: the number of workers (A), the labor productivity (B), firm age (C), and geographic area (D). Labor productivity is defined as value-added divided by the number of workers.

To address these concerns and identify the impact of public versus private demand on firm survival, we focus on auctions in the construction sector, which accounts for 19% of procurement spending and involves 13% of firms in our dataset. For this subset of contracts, we can extract information about the bidding process (i.e., individual bids, the identity of bidders, and final ranking) directly from the official tender documents when available. We employ such additional pieces of information to gauge causality. Leveraging the gap between the winning and losing bids, we define a running variable and a cutoff (i.e., at the runner-up bid) to implement a regression discontinuity (RD) analysis that compares firms that win a public contract with firms that lose by a small margin. The key identifying assumptions are that (i) firms cannot perfectly manipulate the award assignment around the cutoff, (ii) the award is as good as random for bids in the vicinity of the cutoff, and (iii) winners and runners-up are ex ante exchangeable. In the paper, we provide evidence that validates these assumptions. Moreover, the auction-level analysis allows us to consider self-selection issues, both in the procurement market and specific auction.

Our causal estimates confirm the descriptive comparison in Figure 3.1. We find a substantial increase in survival probability as a response to the awarding of contracts. Specifically, we estimate that winning a government contract—whose median duration is about six months—causes an increase in the 24-month (36-month) survival probability by 1.9 (3.4) p.p. on top of a baseline 97.9% (96%) survival rate—i.e., an 85% decrease in exit rate. We show that the results are robust to the risk of collusive bidding behavior around the cutoff—a concern when assuming quasi-random contract allocation—as well as to the risk of contamination with other awards.

The estimated effect may arise from the combination of a potential scale effect (i.e., additional revenues from the contract award) and a composition effect, which rebalances the firms' source of income toward public money. We find no scale effect at play and that firms absorb the marginal public demand boost ($\approx$+10%) by substituting approximately 17% of their revenues from the private demand in the award year. Thus, we can interpret the boost in survival effect as depending on the public nature of the demand shock rather than on the earnings it generates. This effect hinges on forced rather than voluntary exits,[4] which are instead unaffected by the winning of procurement contracts.

To investigate the implications of our results and understand which features survivors have in the wake of public demand, we explore two additional set of outcomes. On the one hand, we replicate the RD analysis to examine the role of productivity, a relevant aspect to consider because of its aggregate impact (Baier, Dwyer Jr, and Tamura, 2006)—especially for an economy like Italy with its sluggish and increasingly dispersed productivity (Calligaris et al., 2016)—and its role in predicting survival in the private market (Ugur and Vivarelli, 2021). We find no 'public procurement premium' (i.e., *ex post*), since we estimate no meaningful difference in lead labor productivity levels between winners and runners-up. Accordingly, public demand helps firms survive longer, but it does not make them more productive.

On the other hand, we investigate whether public demand shocks improve financial dynamics. We match our data with the Central Credit Register, which contains detailed monthly records on all bank-to-firm loans in Italy above €30,000. We start with compelling evidence that firms that win a public contract improve their credit performance compared to close losers. First, winners receive more credit through uncollateralized loans after being awarded the contract. This is attributed to the security of earnings from public contracts, as evidenced in the literature (Goldman, 2020; Giovanni et al., 2022). Second, the winners also witness a decline in low-performing loans in their accounts, which relaxes their financial constraints. This difference in creditworthiness, initially similar between winners and losers, becomes apparent immediately after the award and lasts for at least 36 months. Crucially, this improved credit stock and quality performance is associated with better survival prospects, particularly when spotlighting financially distressed firms. After securing the contract, the latter display a persistently lower exit rate than the losers, who continue to struggle with financial distress. The effect is milder when comparing unconstrained winners and losers. This evidence highlights the role of the alleviation of credit constraints induced by public procurement in promoting business survival. Winners leverage public sales to circumvent the cycle of tightening financial constraints and credit restrictions that ultimately may push losers out of the market.

Related Literature. By examining the government's role in firm survival, this paper joins a long-standing debate on the effectiveness of fiscal policies. Most of the existing evi-

[4] Throughout the paper, the 'forced exits' consist of bankruptcies and forced asset liquidations.

dence comes either from innovation and investment subsidies (Cerqua and Pellegrini, 2014; Criscuolo et al., 2019) to firms or from place-based policies (Becker, Egger, and Ehrlich, 2010; Kline and Moretti, 2014). Little is known about the implications of demand-based policies on firm performance. To contribute to this scholarship, we add to the more general empirical literature studying the effect of a demand shock on firms' outcomes (Foster, Haltiwanger, and Syverson, 2016; Pozzi and Schivardi, 2016), which hinges on solid theoretical predictions (Drozd and Nosal, 2012; Gourio and Rudanko, 2014; Arkolakis, Papageorgiou, and Timoshenko, 2018). In particular, we are interested in public demand shocks channeled to the private sector through procurement markets. Across different contexts, exposed firms—*conditioning on survival*—are found to experience a persistent boost in revenues and employment growth with evidence from Austria (Gugler, Weichselbaumer, and Zulehner, 2020), Brazil (Ferraz, Finan, and Szerman, 2015), Ecuador (Fadic, 2020), and South Korea (Lee, 2017).[5] A positive public demand shock is also found to induce more capital investment (Hebous and Zimmermann, 2021), easier access to external borrowing (Goldman, 2020; Giovanni et al., 2022), and more innovation (Czarnitzki, Hünermund, and Moshgbar, 2020). If the shock is negative, firms consistently respond by cutting capital (Coviello et al., 2021). Goldman (2020) documents how US federal contractors benefited from government purchases across these dimensions, using the 2008–2009 financial crisis as a natural experiment. Barrot and Nanda (2020) find that the speed of payments to these contractors significantly affects their employment growth. Our paper complements this empirical literature by focusing on survival, productivity, and credit as additional firm-level outcomes affected by procurement contracts. Moreover, we do not restrict our attention on small businesses.

Our work also directly advances the scholarship that studies the drivers of firm survival. Theoretical predictions and empirical evidence stress that the marginal survival probability increases with age and size (Evans, 1987a, 1987b; Hall, 1987; Dunne, Roberts, and Samuelson, 1989; Clementi and Hopenhayn, 2006). Yet the relationship between growth and the likelihood of survival is not as simple as it appears at first glance. For example, Agarwal and Audretsch (2001) shows that the variance of realized growth rates is found to decrease with size, conditioning on survival. The empirical evidence provided by the authors suggests that the association is shaped by technology and the stage of the industry life cycle. While the likelihood of survival for small entrants is generally less than that of their larger counterparts, the relationship does not hold for mature product life cycle stages or in technologically-intensive products. In mature industries that are still technologically intensive, entry may be less about radical innovation and more about filling strategic niches, negating the impact of entry size on the likelihood of survival. In short, increased scale is not necessarily associated with increased survival odds. Our results on the revenues composition (instead of scale) effect driving firm survival boost confirm this result. The forces affecting survival can be more generally divided into industry characteristics (Zingales, 1998), geography (Choi et al., 2021), macroeconomic conditions (Byrne, Spaliara, and Tsoukas, 2016), product life

[5] This effect is found to be relevant for domestic firms only in a cross-country analysis in Sub-Saharan Africa performed by Hoekman and Sanfilippo (2018).

cycle (Esteve-Pérez, Pieri, and Rodriguez, 2018), exposure to trade (Kao and Liu, 2022) and shocks (Brata, De Groot, and Zant, 2018), all of which interact with those arising from the idiosyncratic characteristics of the firm (Audretsch and Mahmood, 1995; Ortiz-Villajos and Sotoca, 2018).

To explain firm survival, less attention has been paid to institutional features in general (Byrne, Spaliara, and Tsoukas, 2016; Cevik and Miryugin, 2022) and demand constraints in particular (Syverson, 2011; Foster, Haltiwanger, and Syverson, 2016; Pozzi and Schivardi, 2016). We contribute to this scholarship by spotlighting the role of public demand. Consistent with existing contributions by De Silva, Kosmopoulou, and Lamarche (2009), De Silva, Kosmopoulou, and Lamarche (2017), and Kosmopoulou and Press (2022), we find that public procurement promotes firm survival. Our paper differs from these studies in three ways. First and foremost, we show that a public contract award *per se* affects survival probability—the above works focus on the role of subcontracting or reserve price information disclosure. Also, unlike Kosmopoulou and Press (2022), we make causal claims. Second, our analysis goes beyond entrant firms. Third, our data span a construction market country-wide and consider multiple levels of government, construction types, and auction formats.

The rest of the paper unfolds as follows. Section 3.2 describes the data and sketches stylized facts. Section 3.3 presents the identification strategy. Section 3.4 displays the results, which are discussed in Section 3.5. Section 3.6 concludes.

3.2 Data

We gather and combine data on firms and public procurements at the most detailed level available in Italy. The source for the former is the Company Accounts Data System (CADS), a yearly collection of individual balance sheets covering the quasi-universe of limited liability companies in Italy. We complement these using administrative data on the firms' market entry and—if applicable—exit date, with the exit reason as provided by the Chambers of Commerce (i.e., the official business register, *Infocamere* from now on). As for the procurement side, we employ the full list of tender and associated contract records provided by ANAC (i.e., the *OpenANAC* database). The two databases are matched via the winning firms' tax code. To determine the analysis sample, we complement the contract data with two additional data sources on the bids and bidders' records on construction procurement auctions. In this subset of data, we are able to merge procurement data with losing participants, making our firm-procurement dataset unique for Italy.

Finally, we also add credit information coming from a confidential dataset (i.e., the *Central Credit Register*) administered by the Bank of Italy which includes the universe of bank-to-firm loan records at the monthly level, including major features of lending channels and, crucially, the exposure amount per credit type (e.g., self-liquidating or upon-maturity) and the quality of credit (e.g., impaired or expired loans). We will present this data content in further detail in Section 3.5.2.

3.2.1 Firm-level Data

CADS. Produced and distributed by the Cerved Group, the CADS is a proprietary repository of balance-sheet and income-statement data.[6] It covers the population of limited liability companies in Italy—except for the finance and agriculture sectors—accounting for around 70% of the total yearly business turnover in the country. The dataset reports revenues, employment, financial debt stock, capital, and other information at the firm-year level.

Infocamere. The Chambers of Commerce gather data on the universe of active businesses (irrespective of their legal form) in the country, record their registration date (i.e., entry) as well as their de-registration date (i.e., exit), if applicable, including information on the exit reason.[7] From these census data, we use the exit records to build the 'survival' variables that we use as outcomes for the empirical analysis. In particular, we set the record to 'missing' whenever we observe that the firm de-registers due to a merger or relocation, as the de-registration does not involve market exit, and we cannot track future performance. By contrast, we label a firm de-registration as an exit for all other reported reasons, notably bankruptcy. Once a de-registration is labeled as an exit, we categorize such exits into three main types: *forced liquidations* (i.e., creditor-enforced), *bankruptcies*, and *voluntary* exits. The first two are considered 'forced' exits—i.e., events not driven by the owner's choice—while the latter encompasses decisions by owners to cease operations for various reasons, such as a strategic shift in business focus or retirement. Also, from the year of registration, we can retrieve firm age at each point in time.

Procurement versus Non-Procurement Firms. Table 3.1, Panel A, reports a selection of firm characteristics for the full 2008–2018 firm sample. Despite our paper only using a subsample of these firms (see below), it is useful to compare non-procurement(i.e., those that *only* operate in the private market) and procurement (i.e., those that *also* sell to the government.) firms across sectors to display structural differences and supplement the average survival rates displayed in Section 3.1. Overall, we observe 5.86 million unique non-procurement firm-year pairs and about 0.64 million procurement firm-year observations. The former tend to be younger (13 years old on average compared with about 17 for procurement firms) and of much smaller scale in terms of the number of employees (9 vs. 47), revenues (€2.63 vs. 16.43 million), but also in terms of capital and debt stocks. Labor productivity—which we obtain by dividing the value-added by the number of employees—is also higher for procurement firms. This difference in observables characteristics is important to be considered for the selection issues and associated identification concerns for our empirical strategy described in Section 3.3.1. Procurement firms win 6.66 auctions in the pooled sample—i.e., 0.6 contracts per year on average.

[6] www.cerved.com.
[7] www.infocamere.it.

3.2.2 Contract-level Data

OpenANAC. Since September 2020, ANAC has published a large amount of previously privately retained data on Italian public procurement. The OpenANAC database constitutes the single largest source of this type of data ever available in the country.[8] The data includes all tenders solicited by any public authority above €40,000 as a reserve price—a monetary value much lower than any sector-specific publicity thresholds for EU law—all the awarded contracts linked to them, and, importantly, the winner's identity and tax code.

The data report records of (i) the tender—e.g., the category of purchase, the reserve price, the awarding mechanism and the contracting authority; (ii) the award—e.g., the winning discount to the reserve price and number of bidders; and (iii) the post-awarding phase—e.g., contract duration. Among the many other pieces of information reported, the OpenANAC dataset allows us to identify whether the winning firms are part of a temporary partnership of firms (i.e., a consortium), which are typically created with the sole purpose of participating in single tenders and are either immediately dismantled if failing to win the auction, or persist until the contract expiration date. Through this information, we are able to assign the correspondent share of amount of the contract to the firm participating in a consortium.

The full sample (see Table 3.1, Panel B) comprises 1,274,979 contracts totaling a cumulative yearly value of €156 billion—representing about 9% of GDP and 90% of total procurement spending.[9] The mean contract amounts to €1.36 million, receives 4.4 bids, and lasts 585 days; medians are €130,000, 1, and 299, respectively, thus highlighting the skewed distributions typical of public contract data. We report summary statistics for the overall sample, and for constructions. Construction contracts are relevant in terms of the overall procurement market: Throughout the 11-years period covered by our full data, approximately 40% of procurement firms were awarded at least one construction contract, representing around 60% of the cumulative 1.73 trillion euros of public procurement spending tracked by OpenANAC. Consortia represent 6% of the winners. About 20% of the contracts are awarded via auctions—the awarding mechanisms we focus on in the rest of the paper and the setting for our identification strategy.

Additional Sources. The openly available dataset *Banca Dati Amministrazioni Pubbliche* (BDAP) allows us to retrieve one additional but crucial piece of information for this work, which was not available in OpenANAC at the time of the selection of the PDF for the bid-extraction process (see next subsection for details).[10] In particular, for the subset of tenders covered in BDAP—i.e., the work contracts between 2012 and 2017—we sourced data on the identity of all participants in the auctions along with their tax codes (but, notably, not their bids). In order to complement the information on the bidding process, we rely on proprietary data. More specifically, we purchased from *Telemat* the scanned version of tender documents for public works contracts solicited and auctioned off between 2012 and

8 https://dati.anticorruzione.it/opendata.
9 The downloaded dataset dates back to first release of OpenAnac in the fall 2020.
10 https://openbdap.rgs.mef.gov.it/.

Table 3.1. Summary Statistics: Full Sample 2008-2018

	Panel A: CADS Summary Statistics on Firms					
	Non-procurement			Procurement		
	Mean	Median	sd	Mean	Median	sd
Age (Years)	12.92	9.00	12.11	17.41	14.00	13.29
# Workers	8.96	2.58	319.86	47.16	8.76	459.88
Revenues (€,000)	2,634.64	391.00	55,798.56	16,387.86	1,387.00	261441.45
Capital (€, 000)	796.66	43.89	39,802.45	7,192.64	121.52	340169.26
Labor Productivity	84.49	38.33	32,322.42	141.02	49.58	13,635.45
Financial Debt (€, 000)	1,048.84	77.00	14,818.75	7,896.19	248.00	292722.13
Observations	5,859,034			645,723		
Unique Firms	1,046,930			74,399		

	Panel B: OpenANAC Summary Statistics on Contracts					
	Overall			Construction		
	Mean	Median	sd	Mean	Median	sd
Amount (€, 000)	1,357	130	82,705	1,411	151	52,463
# Bidders	4.44	1.00	48.82	13.06	4.00	36.57
Duration (Days)	585	364	1,049	326	231	484
Direct Award	0.27	.	0.44	0.12	.	0.32
Open Procedure	0.19	.	0.39	0.21	.	0.41
Negotiated Procedure	0.32	.	0.46	0.46	.	0.50
Consortium	0.06	.	0.24	0.09	.	0.28
Observations	1,274,979			324,533		

Notes: Panel A: The table reports summary statistics of the 2008–2018 CADS dataset for both non-procurement and procurement businesses. Only *Age* is sourced from Infocamere. Labor productivity is defined as value-added divided by employment.The observation is at the firm-year level and we report the corresponding unique number of firms. Panel B: The table presents summary statistics for the cross-section of OpenANAC data. The level of observation is a contract awarded between 2008 and 2018. The *Overall* column refers to the entire dataset, while the second columns refer to construction.

2017, when available.[11] Through Telemat data, we can link OpenANAC contracts data to the tender documentation by the unique tender ID (i.e., the *CIG* code). In a subset of these documents, alongside the identity of the bidders, the contracting agency reports the individual bids submitted—be it a discount in the case of price-based auctions, or the points obtained in scoring auctions. We extract this information to create a bid-level dataset by merging the bids with the firm-level information from CADS/Infocamere and the contract-level data from OpenANAC/BDAP. We refer the reader to Appendix 3.D for the details on the extraction process of digitalized tender-document records.

3.2.3 The Analysis Sample

We focus on the 11,078 contracts available both in BDAP and Telemat and, for the subset of those with available documentation—i.e., 1,896 contracts—we reconstruct the bid distribution. We define it as the *analysis sample*. We also drop the contracts when (i) they do not include the amount of the winning bid, or (ii) we cannot identify the winner. Our final working sample comprises 1,247 contracts. We merge the extracted bid data with contract-

[11] Telemat is a corporate division of DBInformation S.p.A.—a private company that provides multimedia services to Italian companies to support their development. One of its activities is collecting, scanning, and providing the digitalized version of official documents of the Italian public procurement tender—which are publicly available but only in paper format. See https://www.telemat.it/.

level data (i.e., OpenANAC) and firm-level data (i.e., CADS) via the CIG code, building a bid-level dataset featuring the full distribution of bids alongside the indication of winners as well as the business history of all participants.

Table 3.2. *Population* vs. *Analysis Sample* of Firms and Contracts

Panel A:
CADS vs. Analysis Sample – Firms Winning Construction Auctions

	CADS		Analysis		
	Mean	Median	Mean	Median	t-test
Age (Years)	18	15	18	14	0.323
# Workers	66	11	25	11	0.008
Revenues (€, 000)	29,376	2,037	10,705	1,843	0.394
Capital (€, 000)	6,724	192	1,105	157	0.431
Labor Productivity	109.28	52.30	68.89	50.47	0.261
Financial Debt (€, 000)	17,928	450	3,564	395	0.420
Public Revenues (€, 000)	3,769	523	2,349	610	0.379
# Awards	7	3	6	3	0.752
Share Public Revenues	0.45	.	0.48	.	0.066
Share Direct Award	0.07	.	0.05	.	0.000
Observations	22,806		881		

Panel B:
OpenANAC vs. Analysis Sample – Construction Contracts (Auctions Only)

	OpenANAC		Analysis		
	Mean	Median	Mean	Median	t-test
Reserve Price (€, 000)	2,184	450	1,512	406	0.069
Amount (€, 000)	1,398	308	1,388	310	0.986
Duration (Days)	429.73	308.00	388.36	305.00	0.071
# Bids	44.53	20.00	37.02	17.00	0.000
Observations	30,757		1,247		

Notes: The Panel A reports the mean value, median as well as the p-value for the conducted t-test, for different firm characteristics for the CADS dataset and the analysis sample. We compare the analysis sample of winners with the original sample of winners appearing in CADS. The observation is at the firm-year level. Note also that for CADS, for comparability, we consider only the years 2012 to 2017 for this table, as this is the time span for the analysis sample. We also restrict our attention to construction firms, as the analysis sample includes only public works. We label firms in CADS as construction firms by means of the *NACE* code.

Table 3.2 Panel A reports the mean, median, and p-values for the t-test on differences across the CADS and *analysis* sample of construction firms. We classify firms as construction firms if *winning* construction auctions and being awarded correspondent contracts in the full dataset. On top of scale variables and age, we augment the firm comparison with procurement-specific metrics. First, *Public Revenues* reports the cumulative yearly amount of contracts awarded.[12] Second, # Awards reports the yearly number of contracts awarded. Third, we construct the variable *Share Public Revenues* as the ratio of *Public Revenues* over Revenues.[13] Finally, we define *Share Direct Award* as the share of public contracts awarded

[12] The amounts of multi-year contracts are assigned to the firm-year as follows. We assume that a multi-year contract value is uniformly split into the years of contract duration. For instance, the contract i is assigned at t and ends at $t+2$. The corresponding yearly contribution to the winner's Public Demand equals $amount/3$. This mechanic also applies to consortia.

[13] We emphasize again the distinction between public money flowing to private firms in the form of public contracts (i.e., public procurement) and public *subsidies*, whether in the form of investment programs, direct transfers, or tax cuts. We consider only the former as counterparts to private demand. See Cingano et al. (2022) for a recent overview of the impact of subsidies on firm outcomes.

through a direct award (i.e., without auctions or negotiations) relative to the total amount of contracts awarded by the firm in a given year. In Panel B, we compare the OpenANAC versus the *analysis* sample of auctioned construction contracts (reserve price, award amount, duration, bids received).

As for the balance-sheet and administrative data, differences in means of firm variables are found to be not statistically significant at the 95% significance level, with the exception of number of workers. The average firm-procurement metrics are comparable for Public Revenues and # Awards but tend to statically differ in terms of Share Public Revenues and Share Direct Awards: Winners in the analysis sample tend to rely slightly more on public sales and slightly less on direct awards in contracting. We find that contracts in the analysis sample have a similar amount and duration. However, differences between the two datasets arise for competition intensity as the analysis sample features fewer bids on average.

3.3 Empirical Analysis

In this section, we outline our RD methodology after presenting the identification concerns in our empirical setting and the necessary institutional background.

3.3.1 Identification Concerns

The link between the survival of firms and their access to public contracts is not trivial, in particular when dynamic considerations are included. A naive approach would be to project a firm-level indicator function for survival after k periods onto an indicator of contract(s) recipience—or, equivalently, the amount of public revenue received in levels or as a share of revenues—partialling out a variety of firm characteristics and fixed effects.[14] However, we would be overlooking the main feature of the public procurement market, which is firms' decision to participate in a public auction and the probability of being awarded the contract (conditional on participation). These two dimensions are strongly correlated with other firm characteristics, both observable—e.g., firm size and location—and unobservable—e.g., management quality and political connections. In fact, firms sequentially evaluate two elements when deciding to join the public procurement market. The first is the expected benefits of winning a public contract (in terms of, e.g., their survival chances) and the expected costs of participation.[15] Second, they evaluate the expected competition in the auction to assess the ex-ante probability of winning. Hence, at each point in time, the population of firms is composed of those who (i) self-select to participate in the public procurement market (who also choose which and how many auctions to participate in), whose group is in turn split into those who win and those who ultimately do not win, and (ii)

[14] In Appendix 3.B, we report the results of this exercise. The estimates show a positive correlation between the probability of survival and public revenues, and a smaller effect of typical business predictors for firms that receive procurement contracts.

[15] Firms approaching the public procurement market face both fixed and variable costs in the form of investments required to gather knowledge about the bureaucratic processes involved, analyze auction-specific documents, or build political connections (see, e.g., Akcigit, Baslandze, and Lotti, 2023).

compete only in the private market. In developing an identification strategy, we must necessarily consider such sequence of self-selection decisions to overcome resulting endogeneity problems.

We list three examples. First, demand shocks in the private market might affect the public market's participation rate. Indeed, because of capacity constraints, firms might be temporarily more (less) inclined to bid for government contracts if their private-sector demand gets weaker (stronger). In our setting, this type of selection bias might hold even after controlling for private-market revenues, given that procurement firms are intrinsically different from those that decide not to participate in the public procurement market, as shown in Section 3.2.1. Second, following the analysis in Akcigit, Baslandze, and Lotti (2023), we know that politically connected firms are more likely to be awarded a contract irrespective of their productivity and they also survive longer. As hinted at in Section 3.1, the procurement firms in our sample also survive longer. If the degree of firms' political connection evolves over time, omitting this information would yield upward-biased estimates for the parameter of interest. Third, participation decisions may be driven by the struggle to survive. Consider the case of limited liability firms facing the risk of bankruptcy: *because* they are likely to exit the market, they may decide to engage in public auctions and bid aggressively (see, e.g., Board, 2007 and Calveras, Ganuza, and Hauk, 2004). Such 'bidding for resurrection' effect might downward-bias the estimates.

We cannot rule out these sources of endogeneity in a nonexperimental context unless we assume that participation decisions and procurement contracts are randomly distributed across firms. In this ideal experimental scenario, we could simply contrast the survival rates of procurement and non-procurement firms at both the extensive and intensive margins. Due to data limitations, and the endogenous source of selection in the market and auctions, we cannot conduct such an analysis. A possible alternative strategy is to assign certain projects as unexpectedly assigned to winners instead of close-losers by exploiting the regulatory framework and auction design. This would additionally allow us to control for the fact that winners may be structurally different from losers within the same auction. To this end, we focus on the subset of public works contracts awarded through open auctions for which we observe both winning and losing bidders and the full distribution of bids, i.e., the *analysis* sample. With the bid-level data, we can compare auction-by-auction winners and losers (i.e., the runners-up and the third-ranked). In this way, we account for the decision of firms to participate in the market and in the particular auction; zooming in around the most competitive bids, firms have the same ex ante probability of winning and were awarded the contract quasi-randomly. To quantify the impact of the winning bid, we use a RD analysis whose main elements are tailored to the Italian legal framework.

3.3.2 Institutional Background: Bid Definition and Auction Mechanisms

Our analysis covers two auction formats: price-based and quality-based. In price-based auctions, participants submit bids as percentage-point rebates on the reserve price and are ranked from the highest to the lowest rebate. In quality-based auctions, firms offer a package that includes a discount on the reserve price, technical documentation for the public work to be performed, and possibly a project plan. The contracting authority rates this package on a quality scale of 0-100 and ranks the bidders from the highest to the lowest value. It is important to note that bids in both auction systems are on a 0-100 scale and the contract is awarded to the bidder with the highest rebate or value. Table 3.A.2 presents summary statistics on the distribution of bids, distinguishing between price-based and quality-based auctions.

From 2012 to 2017, for price-based auctions, Italian contracting authorities were required to select contractors through sealed-bid auction contests, which could feature the automatic exclusion of anomalous bids via an algorithm (average-bid auctions or ABAs) or award the contract to the lowest bid (first-price auctions or FPAs).[16] In both cases, the contracting agency announces a project description and a reserve price; then, firms submit sealed bids with discounts on the reserve price.

The idea underlying the ABAs is that, in the context of auctions with several participants, some bids are 'too-good-to-be-true'—i.e., can be associated with underbidding or poor quality bidders and later poor performance—and therefore contracting authorities would be better off by selecting more expensive bidders. The algorithm underlying the ABA procedure essentially eliminates all offers above a mechanically calculated threshold close to the average bid and awards to the highest discount in the interval. Figure 3.2 offers a visual representation of the ABA mechanism in a fictional 20-bid auction. The winner is determined as follows: (i) bids are ranked from the lowest to the highest discount; (ii) a trimmed mean (TM1) is calculated excluding the 10 percent highest and the 10 percent lowest discounts; (iii) a second trimmed mean (TM2) is calculated as the average of the discounts strictly above TM1; (iv) the winning bid is the highest discount strictly lower than TM2.[17] The regulatory default format is the FPA; however—even though not compulsory—public buyers *could* choose to employ an ABA (and hence exclude anomalous offers) when they receive more than ten offers, or the reserve price is below the EU statutory threshold. The auction format is not known in advance by bidders nor perfectly predictable.

3.3.3 Identification Strategy

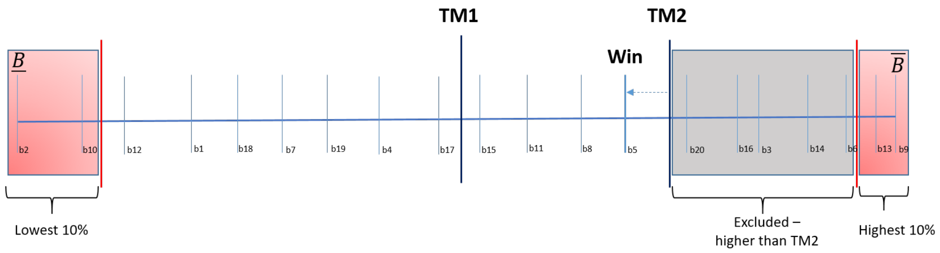
In order to ensure the identification of the public demand effect, we exploit the logic of quasi-random allocation of a contract to firms in the vicinity of the winning bid in an RD fashion. The idea is to compare the outcomes of winning and losing bidders under the assumption that—except for the fact that the former have been awarded a public contract—the two groups are ex ante identical (Cattaneo, Titiunik, Vazquez-Bare, et al., 2020). To do that, we propose an RD framework that pools together multiple auctions with a cutoff just to the right of the runner-up bids.[18]

[16] Contracting authorities can also use scoring rule auctions to select the winner—up to 100 points are assigned to most economically advantageous offer in terms of 'quality' and price. We consider scoring rule auctions as FPAs because their award mechanics is equivalent for the sake of our econometric analysis: the firm obtaining the highest score (instead of the highest discount) is awarded the contract.

[17] We refer to Conley and Decarolis (2016) for a thoughtful discussion of the Italian ABA mechanism.

[18] The idea that similar firms in the same auction hints at similar unobserved costs and/or similar information regarding the auction is not new in the literature. For example, Kawai et al. (2022) leverage the logic of RD

Figure 3.2. Visual Representation of the ABA Mechanism



Notes: This is an example of ABA with 20 bids, where bids are reported in increasing order between $\underline{B}$ and $\bar{B}$. Red areas represent the tails of the bid distribution (±10%), which are excluded to compute the average TM1. Focusing on bids higher than TM1, a second average is computed (TM2). The winning bid is the *nearest* but *lower* bid to TM2 (b5 in the example).

Despite that, and provided that there are no observable variables that influence the treatment probability, units with values of the running variable just below the cutoff (i.e., losers) can be used as a control group for treated units with values at or just above the cutoff (i.e., winners) to estimate the (local) treatment effects on the outcomes of interest. In the rest of this section, we discuss the characteristics and the assumptions of our RD design.

The Cutoff. To make our argument formal, consider a *sharp RD* setting, with a forcing variable B_i and a cutoff $B^\star$ which informs the running variable $X_i = B_i - B^\star$. In this framework, only subjects with positive values of the running variable—i.e., if $X_i > 0$—are treated. This is equivalent to claiming that the probability of treatment (i.e., $Pr(D_i)$) is one whenever B_i strongly exceeds the cutoff level—i.e., $Pr(D_i = 1|B_i > B^\star) = 1$. In the context of procurement auctions pooled together, there is no 'fixed' cutoff like $B^\star$ to be used in the definition of the running variable, as long as the discount of the winning bids differs depending on the bid distribution, the contract amount, the local market conditions, and so forth. Hence, we use a normalized, auction-level cutoff ($B_a^\star$) with the same characteristics as the one above, namely:

$$Pr(D_{i,a} = 1|B_{i,a} > B_a^\star) = 1, \tag{3.1}$$

and change the definition of the running variable accordingly ($X_{i,a} = B_{i,a} - B_a^\star$).[19] We define the auction-level cutoff by leveraging the institutional features presented in Section 3.3.2

design to distinguish allocation patterns reflecting cost differences across firms from patterns reflecting non-competitive environments. Kong (2021) employs the same strategy to isolate synergy from affiliation effect in sequential auctions. In both cases, the running variable is expressed as $\Delta_{i,a} = b_{i,a} - \wedge \boldsymbol{b}_{-i,a}$ where bids are normalized in percentages of the reserve price. Moreover, multiple cut-off RDs are typical in the education literature for estimating the effect of school quality on different pupils' outcomes. For instance, Sekhri (2020) exploit a threshold that is year, college and stream specific. Similarly, Pop-Eleches and Urquiola (2013) use a cutoff score that is school and track specific for the admission into secondary education. Finally, Lucas and Mbiti (2014) utilize as a threshold the score of the last student admitted at a school-district level.

[19] Cattaneo et al. (2016) present a class of RD models with multiple cutoffs close to ours, and discuss three common applications in the empirical literature: running variables informed by vote shares, population, and test scores. Applications encompass close call elections (Cerqua and Pellegrini, 2014) and school admissions (Hoekstra, 2009).

and our bid-level data. More specifically, for each auction, we rank the bids and pinpoint the winning (i.e., B_a^1), runner-up (i.e., B_a^2), and higher-order bids ($B_a^3, .., B_a^N$). Consider the case of FPAs: conditional on the observed bid distribution up to the runner-up's, any discount exceeding B_a^2 wins the contest—in formula:

$$Pr(D_{i,a} = 1 | B_{i,a} > B_a^2) = 1. \tag{3.2}$$

An immediate comparison between Equations (3.1) and (3.2) reveals that a straightforward choice of the auction-level cutoff is $B_a^\star = B_a^2 + \varepsilon$, where ε is a value smaller than the slightest auction-specific difference between a runner-up and a winner in our data.[20]

When it comes to ABAs, once excluding the tails and computing the trimmed averages, all bids in the TM2-TM1 interval are treated as in FPAs, and the winner is the one offering the largest discount (see Figure 3.2). Therefore, conditional on the observed bid distribution, and focusing on the TM2-TM1 interval only, we rank the bids from the highest to the lowest discount ($B_{a,TM}^1, B_{a,TM}^2, .., B_{a,TM}^N$) and define the cutoff as $B_a^\star = B_{a,TM}^2 + \varepsilon$. Note that, in defining such cutoff, we are implicitly modifying the definition in Equation (3.1) to reflect the fact that a winning firm should overbid the runner-up discount, but not exceed TM2—expressed in a formula: $Pr(D_{i,a} = 1 | B_{i,a} > B_{a,TM}^2 \forall B_{i,a} < TM2) = 1$.[21] Finally, the peculiarities of ABA auctions generate cases in which the absolute distance between the winning and the runner-up bid (as defined above) is larger than the absolute distance between the winning and the nearest absolute excluded bid. In a robustness check, we define the cutoff using the nearest bid with unaltered findings.

The Running Variable. The running variable takes up the following values: $X_{i,a} = 0$ for the runner-up, $X_{i,a} > 0$ for the winners, and $X_{i,a} < 0$ for all other losing bidders. In other words, a positive value of $X_{i,a}$ implies that bidder i won auction a, and a zero or negative value of $X_{i,a}$ implies that bidder i lost auction a. The running variable equaling $0 + \varepsilon$ marks the threshold between winning and losing the auction. Considering A auctions, we observe one point per auction (i.e., totaling A) to the right of the cutoff representing the winning bids (i.e., positive scores), A points massed at zero, and $N_l = \sum_{i=1}^{A} N_i$ to the left (i.e., negative scores)—where N_i is the number of losers in the auction i—representing losing bidders other than the runner-up.[22]

The RD Sample. In the spirit of Gugler, Weichselbaumer, and Zulehner (2020), we argue that the comparison between the winner (i.e., treated) and the runner-up plus the third-ranked bidder (i.e., controls) provides a valid counterfactual to estimate the effect of winning a procurement auction on firms' outcomes. There are two reasons for restricting our

[20] The slightest difference is 0.0000916 p.p., and we set the threshold value at a negligible 0.0000190 p.p.—We define the latter as such to have all the winners to the right of the cutoff and all losers to the left .

[21] We stress that our interest is in the ex-post analysis of bid distribution, hence we can safely condition our analysis on the observed bids and ignore the fact that different values of $B_{i,a}$ would modify TM1 and TM2 and move the very definition of runner-up with its relative cutoff.

[22] It is important to note again that our analysis sample includes both price-based auctions (FPAs and ABAs) and quality-based auctions. Bids in both auction systems are expressed a 0-100 point scale, as explained in Section 3.3.3. This approach ensures a consistent definition of our running variable across auction types.

sample up to the third-ranked bidder: on the one hand, it provides us with firms that are very similar to the winners not only in terms of bid distance but also in terms of the underlying characteristics. Some of the firm characteristics around the cutoff are no longer similar (and are jointly different) if we keep the full spectrum of bids (see Appendix 3.A). On the other hand, the choice of keeping only up to the third-ranked bid better balances the number of observations on both sides of the cutoff—as long as adding losing bids would only inflate the sample to the left of the threshold.

The RD Model. After defining the cutoff, the running variable, and the sample of bids we can implement a sharp RD by pooling the auction-specific scores. The regression model reads

$$Y_{i,a(t)} = \alpha + \tau D_{i,a(t)} + f_l(B_{i,a(t)} - B^*_{a(t)}) + D_{i,a(t)} f_r(B_{i,a(t)} - B^*_{a(t)}) + \epsilon_{i,a(t)}, \qquad (3.3)$$

where $Y_{i,a(t)}$ is the outcome of interest—e.g., in the baseline analysis it is an indicator for survival after the award ($Surv_{i,t}^{t+m}$ and $m = [12, 24, 36]$ months).[23] More specifically, we look at the probability of a firm i being alive 12, 24, and 36 months after participating in the auction a (at a specific point in time t). The variable $f_k(B_{i,a} - B^\star_a)$ stands for a second-degree polynomial function, which we let vary on the left and right side of the cutoff ($k \in \{l, r\}$). $B_{i,a}$ is the bid submitted by firm i in auction a, $B^\star_a$ is the auction-specific cutoff value, $D_{i,a}$ is an indicator function for winning the contract—i.e., $D_{i,a} = \mathbb{I}[B_{i,a} > B^\star_a]$—and τ is the estimand treatment effect. Given the time spanned by the data and in order to rely on the same sample of observations for all outcomes, we limit our analysis to 36 months.

Testing the RD Assumptions. The first identification assumption is that agents cannot manipulate the contract assignment around the cutoff. Therefore, the main confounding factor to the causal interpretation of the model from Equation (3.3) is the possibility that bidders change their score strategically and are assigned to their preferred treatment condition (McCrary, 2008). In our context this is not the case, as firms participating in the auctions cannot *perfectly* control their distance to the runner-up and therefore their ranking, which is the key ingredient for the definition of the cutoff. This is especially true in our sample, which features, on average, 37 bids in competitive contests (see Table 3.2). The second key element is the randomization assumption, namely that the regression functions $E[Y_i(0)|X_i = x]$ and $E[Y_i(1)|X_i = x]$ are continuous in x at $B^\star_a$.

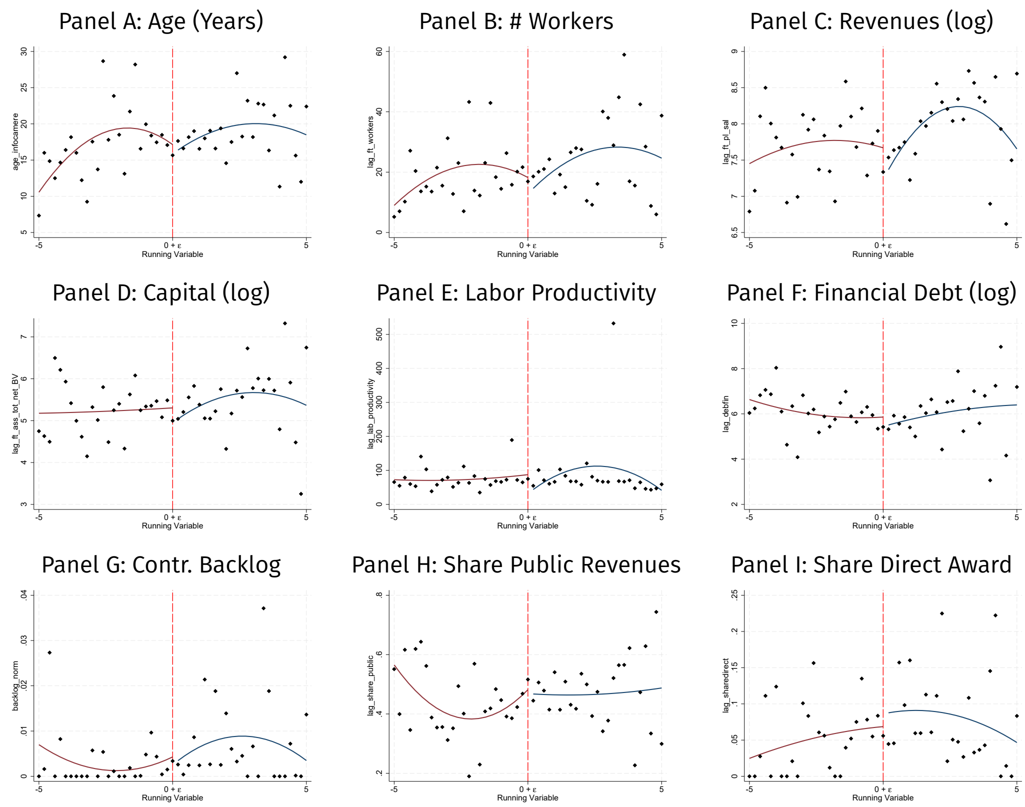
Appendix 3.C discusses the validity of these two hypotheses. On the one hand, a potential concern arises from the possibility of collusive behaviors by cartel members, who may manipulate their bids—and their ranking—even in the proximity of the cutoff and affect our notion of competition. We provide evidence that the results do not suffer from the risk of collusion. On the other hand, we propose a placebo exercise that does not falsify the randomization assumption.

[23] For the rest of the section we omit the subscript (t) given that there is a one-to-one mapping between the specific auction a and the time t.

The third condition is that there is a (sharp) discontinuity in the treatment probability at the cutoff. This condition is ensured by construction for auctions that assign the treatment (i.e., the contract) to bidders with the lowest bid.

Fourth, the groups are assumed to be exchangeable around the cutoff. In other words, treated and control firms are supposed to be ex ante identical, differing only by treatment status, in the absence of which they would exhibit the same dynamics of outcome variables. Hence, any difference between the average response of treated and control units around the cutoff is fully attributed to the (local) average effect of the treatment. This assumption is usually tested by looking at the continuity of the relevant characteristics before the event for firms around the cutoff. More specifically, we graphically compare the pre-event variables of winners and losers in Figure 3.3 where we plot the mean values of several characteristics the year prior to the auction.

Figure 3.3. Firm Characteristics: Winners and Marginal Losers at $t-1$



Notes: Firm-level characteristics for winners (blue line) and marginal losers (maroon line, include runners-up and third-ranked) prior to the contract award. The running variable is rescaled to reflect the distance from the runner-up bid ($X_i = B_{i,a} - B_a^2$). All variables are lagged one year except contracting backlog, which is measured on the award day, as it is a snapshot of firm backlog at the daily level. Balance-sheet variables are transformed in natural logarithms. Each point represents the average of the covariates for a given non-overlapping bin.

We test the continuity of firms' (lagged) characteristics that correlate most likely with the probability of both winning procurement contracts and surviving according to the literature—i.e., age, employment, and labor productivity. We also include other scale controls such as revenues, capital and financial-debt stocks.[24] In addition, we include metrics for behavior in public procurement. This exercise allows us to mitigate the risk of capacity constraints, corruption, and firms' connections biasing our results, as argued in Section 3.3.1. *Contracting backlog* is the residual backlog of ongoing contracts at the exact date of the award normalized by the revenues. It accounts for firms that rely more on public procurement and therefore are more likely to win, either because of experience or because of political connections. Notably, in the spirit of Kawai et al. (2022), its discontinuity at the cutoff can be indicative of bid rigging. We also look at the share of direct awards received. This measure proxies the degree of political connectedness and might signal the presence of relational contracts with buyers (Calzolari and Spagnolo, 2009; Albano, Cesi, and Iozzi, 2017), both cases in which firms are more likely to receive direct awards.

Winners and close-losers do not display significant differences along any of the above dimensions. All in all, the plots confirm the lack of systematic difference between winners and losers at the cutoff. As we consider several covariates, some discontinuities could be statistically significant (or close to) by chance. Therefore, to test against a continuity pattern jointly, we perform an auxiliary exercise inspired by Lee and Lemieux (2010). We execute seemingly unrelated regressions, where each regression features one of the nine covariates considered above. Specifically, we regress a binary model with an indicator equal to one when the observation is treated, i.e., if it lies above the threshold, on each of the nine covariates from Figure 3.3. We then perform a χ^2-test for the estimated coefficients being jointly equal to zero. We cannot reject the null, which corroborates that observable characteristics are jointly continuous at the cutoff. Winning and losing is therefore 'as-good-as-random' conditional on close bids.

Altogether, our empirical design bolsters a causal interpretation of the RD results, which we present in the next subsection.

3.4 Results

In this section, we present the baseline results of our RD model and the tests for robustness. Before that, we show the short-run impact of public contract awards on winner activity as a first-stage for analysis.

3.4.1 Short-run Responses: 'First-stage' Effects

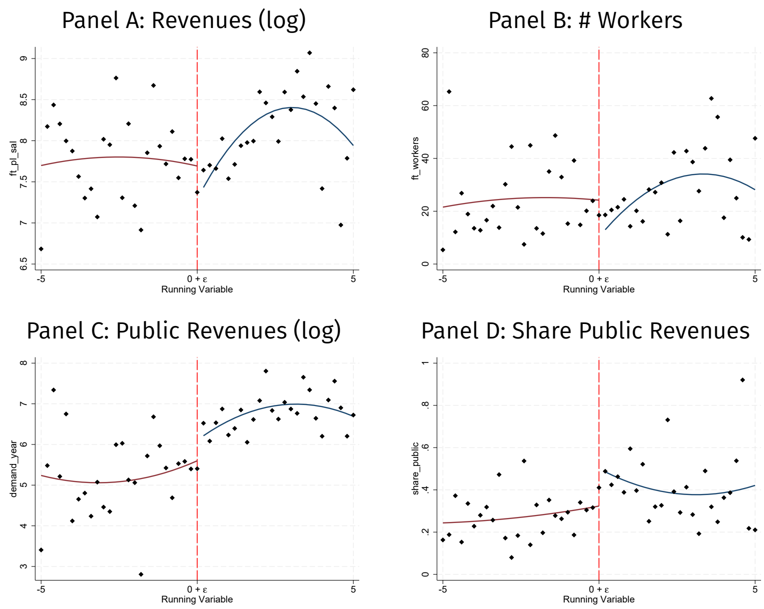
Winning a public contract secures a source of earnings while taking up part of the existing productive input. In response, a firm in the short run (i.e., within the award year t) can either expand its activity in order to keep its exposure to the private market unchanged or react to the congestion by reducing its private commissions, thereby substituting them with public revenues. Distinguishing the two strategies allows the correct interpretation of our

[24] A similar age between winners and close-losers is particularly important for identification in this context as entrant firms tend to bid more aggressively and win with significantly lower bids compared to incumbents (De Silva, Dunne, and Kosmopoulou, 2003; De Silva, Kosmopoulou, and Lamarche, 2009). We want therefore to compare firms with similar experience in the construction market.

RD estimates. Indeed, they quantify the 'gross' impact of public demand on survival probability which combines, and potentially conflates, both a *scale effect* (i.e., additional revenues coming from the contract award) and a *composition effect* (firms' rebalancing sources of income toward public money). To test whether winning firms expand their business right after a procurement award, we replicate the comparison of winners versus marginal losers from Figure 3.3 but at time t and for variables related to firm scale and business decisions.

In Figure 3.4 we plot the visual effect for revenues (Panel A) and employment (Panel B) to observe winners' short-term response compared to close-losers. The former provides direct information on whether the additional income from public contracts adds to the bulk of income or tends to crowd out private activities. At the same time, an increase in employment would signal the presence of a scale effect. Neither measure, however, shows any significant shifts in the award year, suggesting a zero-scale effect. Panels C and D present Public Revenues and Share Public Revenues to further explore the strategic response of firms in terms of revenue reallocation. They show significant jumps in public revenues ($\approx +70\%$) and share ($\approx +20\%$) when firms receive a public contract, regardless of the size. All in all, this evidence suggests that (i) being awarded a public contract induces a strategic response and (ii) a composition effect seems to be at play: firms absorb the higher public demand by shifting some of their sales from private to public customers with no apparent scale adjustment. Therefore, we can interpret the increase in the survival effect shown below as being related to the *nature* of the demand shock (i.e. public *vis-à-vis* private) rather than to the revenue it generated.

Figure 3.4. Contemporary Effects



Notes: Visual representation of the RD estimate of being awarded a public contract on revenues (Panel A), number of workers (Panel B), public revenues (Panel C) and the share of public revenues (Panel D), all measured at t.

3.4.2 Medium-run Responses: The Survival Premium

We report the results in the probability of surviving 12, 24, and 36 months after the award in Table 3.3, which show the results of the baseline specification.[25] The estimates are local as resulting from winners compared to the two closest losers with bids that are no more than 5 p.p. of discount on the reserve price away from the runner-up threshold.[26] We observe an accruing effect over time. Part of the effect mechanically reflects the contract duration. However, as the median contract in the RD sample lasts approximately ten months—or 300 days, as shown in Table 3.2—the boost to survival goes well beyond it.[27] Hence, public awards positively affect the survival probability of the winning bidders in the medium run compared to control bidders.

The estimates show that being awarded a public contract has a positive effect on the probability of surviving both 24 and 36 months after the award date. Survival increases by 1.9 and 3.4 p.p. from baseline values of 97.7 and 95.7 p.p., respectively. Looking at the 36-month survival rate, winning a contract allows a firm in our sample to reduce its exit rate from approximately 4% to 1%, corresponding to an 85% reduction in market exit odds.[28]

Table 3.3. RD Regressions—Baseline

			Survival		
Window	Polynomial	Kernel	m+12	m+24	m+36
± 5 p.p.	Quadratic	Triangular	0.008	0.019	0.034
			(0.005)	(0.008)	(0.011)
			0.991	0.977	0.957
			2,532	*2,532*	*2,532*
			0.231	*0.226*	*0.157*

Notes: The RD coefficients (first row of each panel, in bold) are bias-corrected and the robust standard errors are in parentheses (second row). We also report the mean of the dependent variable (third row), as well as the number of observations (fourth row). The fifth row reports the optimal bandwidth size in p.p. The observation is at the auction-bid level. Given our selection, the number of auctions in each regression corresponds to one-half to one-third of the observations, depending on the share of auctions with two participants (winner and runner-up only) or more (third-ranked also). We use the bandwidth minimizing the MSE.

We conclude this section by summarizing the insights from our robustness checks analyses, presented without a specific order of relevance. First, we use alternative model spec-

[25] We employ a triangular kernel and a second-order degree polynomial in the focal specification. On the one hand, the chosen kernel gives more weight to observations close to the cutoff. On the other hand, the chosen polynomial allows us to account for non-linearities in the scores on both sides of the cutoff. We refrain from using higher-order polynomials as they can lead to noisy weights and poor confidence intervals (Gelman and Imbens, 2019).

[26] To include at least one within-auction comparison always, we exclude in Table 3.A.1 those auction observations where the runner-up's bid exceeds 5 p.p. distance from the winner. Our findings are robust to this sample selection.

[27] For contracts above €150K, OpenANAC also provides information on renegotiations and delays so that we are able to compute the real duration of the contracts in our analysis sample. 90% of the contracts that we are able to merge appear to be on time.

[28] We employ heteroscedasticity-robust standard errors. Yet it is common in the empirical literature using RD studies to define standard errors as clustered by the running variable (Kolesár and Rothe, 2018). This means that observations with the same realization of the running variable are defined as members of the same cluster. A cluster-robust procedure is then used to estimate the variance of the estimator. Accordingly, in an auxiliary analysis, we cluster the standard error at the auction level with virtually unchanged results.

ifications of the RD to address concerns that the RD outcomes might be sensitive to the definition of the functional form. Second, we tackle the issue of potential contamination in both the treatment and control groups, given that any bidder might pursue subsequent contract opportunities following auction a. Third, we confirm that the findings are consistent to the changes in auction rules. Fourth, we ease the concerns of contract assignment manipulation, particularly bid rigging. Fifth, applying a placebo cutoff for treatment allocation yields non-significant estimates, suggesting that the effects observed in our regressions arise from the exogenous contract assignment informing a demand shock rather than other confounding variables. Sixth, adding covariates into our RD models does not alter the estimates, corroborating the correct specification of our RD model. Altogether, these results validate our findings across multiple dimensions. We refer to Appendix 3.C for a detailed discussion on each of these exercises.

3.5 Discussion

In this section, we discuss our findings. First, we present an effect decomposition on exit type. Results indicate that survival is exclusively driven by the reduction in forced exits. Second, to explore possible drivers of our results, we study the effect of public awards on the evolution of two key determinants of survival at the firm level: productivity and credit performance. We pin down the relevance of the latter as a mechanism for the results on survival based on forced exits.

3.5.1 Which Exits Drive the Survival Effect?

We employ the baseline RD methodology, utilizing the mutually exclusive exit classifications outlined in Section 3.2.1—namely, forced liquidations, bankruptcies, or voluntary exits—as separate outcome variables.[29] The analysis is instrumental to underpinning the mechanism behind the survival effect that we estimated. In fact, while forced liquidations and bankruptcies are externally imposed, voluntary liquidations can emerge from strategic firm decisions, such as participating in auctions as an ultimate survival strategy (e.g., bid-to-resurrect) and choosing to liquidate after the loss.

In Figure 3.5, we present the RD point estimates for each of the three exit outcomes at every $t+m$. We stress that we shift the analysis from survival to exit probability in order to streamline the interpretation of the results. Consequently, the sign of the estimated exit parameters is plotted in reverse relative to the baseline analysis for consistency in visual representation. Forced liquidations are depicted in blue, bankruptcies in gray, and voluntary exits in red. To enhance comparability with the gross survival effect, we include in black the baseline gross survival estimates and confidence intervals from Table 3.3.

No exit type presents significant effects 12 months following the award, consistent with the baseline null effect on survival in the first year post-award. Yet, after 24 months, esti-

[29] In particular, we generate binary indicators for each type of exit and respective time lead. These are denoted as $\mathbb{I}(Exit)^{type}_{i,t+m}$ for $m \in [12; 24; 36]$ and $type = e \in [forced\ liquidations; bankruptcy; voluntary]$.

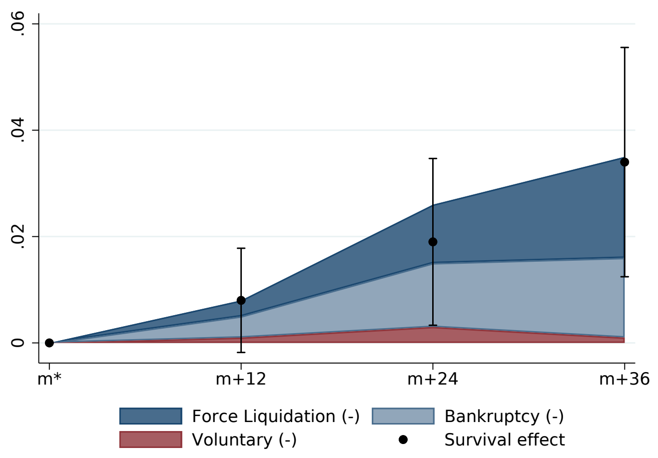
mates indicate a notable decrease in the probability of forced exits (around 2 p.p. combined), extending to more than 3 p.p. combined at *m+36*. We underscore that the forced exit parameters sum up to our baseline survival estimates, capturing the cumulative impact that procurement has on survival. We detect no discernible impact on voluntary exit at any point in time.

This outcome decomposition suggests that the primary drivers behind the estimated survival effects are forced (non-)exits. Winners of procurement contracts experience a buffer against externally imposed exit pressures compared to losers, emphasizing the 'protective' nature of public contracts. In the following subsection, we establish that this protective effect is significantly mediated by the availability and quality of credit, which are enhanced by contract awards and act as a channel for the survival effect of procurement.

3.5.2 How Do Winners Survive Longer?

Productivity Dynamics. Winning firms could increase their probability of survival by improving their productivity. In our data, labor productivity correlates with survival (see Appendix 3.B) and existing evidence posits a causal relationship between the two (Ugur and Vivarelli, 2021). We replicate our RD analysis to investigate whether procurement induces a productivity premium. Table 3.4 reports the estimated parameters for labor productivity one, two, and three years after the award. The level of the observation is the firm-year given that, unlike for survival, we measure productivity once a year through the balance sheet data. Essentially, we compare the future values of the productivity of firms receiving contracts at any point in t with those of the runners-up and the third-ranked. We can run this exercise only for firms active in the market in each lead—that is to say, all results hold conditioning on survival. Hence, the estimated difference between the groups includes both the effect due to the evolution of labor productivity and the one driven by the differential in survival rates.

Figure 3.5. Visual dynamic representation of the RD estimates by type of exit



Notes: RD estimated coefficients and standard errors (95% confidence level) of survival boost m months after the participation in auction a (black dots and lines). The blue shaded areas correspond to the coefficients for forced exits, namely forced liquidations (dark blue) and bankruptcies (light blue); the red shaded area reports the RD parameter estimated for voluntary exits.

We detect no meaningful effect of public contracts on the labor productivity dynamics of survivors. Lead productivity does not seem to be affected by shocks in public sourced demand. This, in turn, implies that the estimated increase in survival rate is not channeled through an increase in productivity. This result is consistent with several non-competing explanations offered both in the policy practice and in the academic literature. First, there is evidence that government-linked firms invest less in intellectual capital (Cohen and Malloy, 2016). Second, the government may have incentives to protect inefficient firms through public contracts and shield them from market competition because their existence meets policy goals or dynamic considerations—e.g., this is the case, for example, with 'set-aside' programs in the US, where about a quarter of the federal government procurement budget is allocated to support small, disadvantaged or local firms (Cappelletti and Giuffrida, 2022). Yet neither the requirements nor the explicit goals for these programs take into account the impact on business productivity.

Table 3.4. RD Regressions—Productivity

	Outcome		
	t+1	t+2	t+3
Labor Productivity	-15.888	-6.943	-1.128
	(20.632)	(4.425)	(5.846)
	61.038	61.094	61.364
	2,154	*1,829*	*1,505*

Notes: Table 3.3 is replicated using labor productivity as an outcome. Labor productivity is value-added divided by the number of workers

Credit Dynamics. Winning firms could experience an increase in their probability of survival by improving their credit position. Four elements set the ground for the empirical investigation of this mechanism. First, it is broadly established that an improvement (worsening) in a firm's financial performance is a strong predictor of medium- to long-term survival (exit) (see, e.g., Blattner, Farinha, and Rebelo, 2023 and Schiantarelli, Stacchini, and Strahan, 2020). Second, a lower chance of forced exits drives the survival effect in our data—and financial distress may eventually result in a forced closure. Third, the Italian government is a reliable payer—that is, it invariably honors its financial commitments with suppliers.[30] In turn, this implies that from a bank's perspective, a firm winning a public contract has a more secure future cash flow. Fourth, recent contributions stress the importance of procurement for a firm's financial performance. For example, Goldman (2020) finds that small firms receiving a procurement demand shock receive more bank credit. Recently, Giovanni et al. (2022) propose a procurement model in which financially constrained awardees pledge their future earnings to improve their financial position. In terms of our analysis, this mechanic would imply that winning firms survive longer *because* they improve their financial position.

[30] For details, see the Italian National Association for Constructors report 'Pagamenti della Pubblica Amministrazione' (Payments from the Public Sector) at https://ance.it/wp-content/uploads/archive/29547-ANCE-Report%20sui%20Pagamenti%20PA.pdf, in Italian.

In what follows, we propose a battery of empirical exercises that point to the relevance of the credit channel for our survival results. We match the RD sample with the monthly bank-to-firm credit records (i.e., the Central Credit Register) presented in Section 3.2.1. We provide suggestive evidence that improved credit position is a prominent transmission channel for survival dynamics by showing that i) credit stock and quality evolve differently for winners and losers after the award and ii) the variation in credit quality mirrors differential survival prospects for winners and losers.

Credit and Procurement. We look at monthly dynamics of different credit variables for winners as opposed to 'non-contaminated' losers—i.e., runners-up and third-ranked being awarded no public contracts 12 months before or after the award. Such a high frequency of data and selection of losers allows us to eyeball uncontaminated short-term dynamics triggered by the award itself. We examine a 12-month before and 36-month after time window around the award month to also eyeball pre-auction dynamics.

We start by documenting three facts linking credit and procurement in our data. First, winners obtain more credit. Second, winners obtain more earnings-based credit but not different collateralized credit, hinting at the role of the award rather than alternative factors in increasing the credit stock. Third, winners see an improvement in their credit quality.

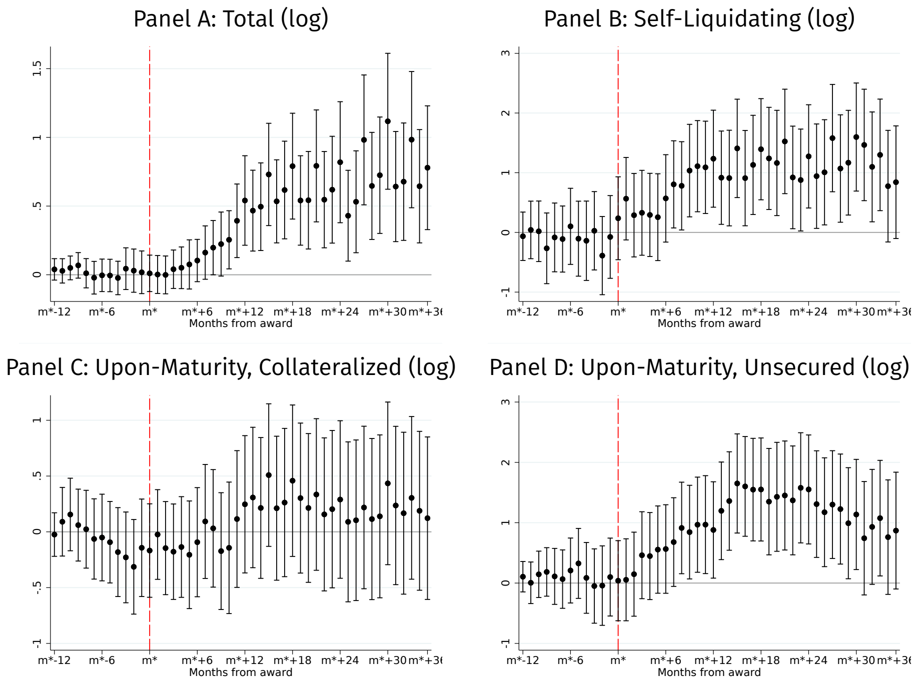
In Figure 3.6, we plot the estimated differences in total credit stock and across different credit stock classifications between the two groups of firms.[31] In Panel A, we display (log) total credit dynamics. While there is no difference before the event, the winner-loser spread begins widening significantly as early as 8 months after the award, with a clear upward trend up to 0.5% until 12 months, when the survival effect is not yet significant. The gap widens up to around 1% and remains significant until 36 months ahead.

The short-term response highlights that the effect on total credit kicks in before any survival differential is detectable. To get a first clue of the mechanism that links the awards to credit stock dynamics, we further break down the total credit into subcategories. First, we look at (log) *self-liquidating credits* (Panel B), which consists of bank loans typically granted for investments and whose amount is usually paid back with the profits generated by the investment itself—i.e., with the income generated by a secured contract in our case. The difference in self-liquidating credit exhibits a steeper pattern after the awarding date than total credits, already turning significant after 7 months (+0.8%) and increasing up to +1.2% (+1.5%) after 12 (36) months. This pattern suggests that winning firms leverage the advance invoices (or award notices) to borrow just after signing the contract. We also focus on upon-maturity loans and further distinguish between *collateralized* (Panel C) and *unsecured* (Panel D). The former consists of loans granted against a collateral—typically physical assets that firms pledge to secure the loan. However, future revenues from the award of a public contract cannot be collateralized in a strict sense; hence, the contract award should not have effects in this regard. On the other hand, the choice to accord unsecured credit does not depend on the availability of pledgeable assets but rather on a case-by-case assess-

[31] Specifically, we plot the coefficients estimated in a regression of the credit variable(s) on month fixed effects interacted with the binary indicator for winners.

ment of cash flow prospects. In this sense, being awarded a public contract could matter. While collateral-backed loans do not record any difference between winners and losers, unsecured credit stock for winners increases after 8 months and remains significantly higher (more than 1%) more than two years after the award.[32]

Figure 3.6. Credit Stock Dynamics



Notes: Estimated differences between bidders awarded and not awarded public contracts in terms of total credit (Panel A), self-liquidating credit (Panel B), and in credit upon-maturity which required (Panel C) and did not require collateral (Panel D). For each panel, we plot coefficients estimated in a regression of the credit variable(s) on month fixed effects interacted with the binary indicator for winners. Sample include winner plus losers up to rank 3.

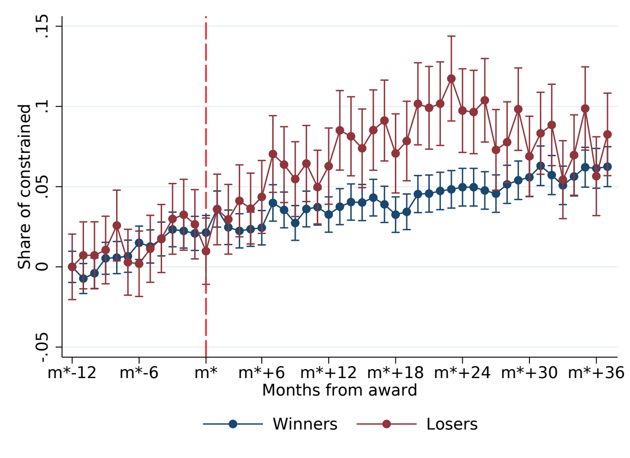
To complement our findings on credit stock performance improved by procurement, we check whether easier access to credit also helps winners overcome existing financial constraints. For this scope, we build the binary variable *credit constraint*, which indicates the monthly presence in the firm's account of any loans formally classified as bad, impaired, or expired by the lending bank. In Figure 3.7, we plot the dynamics of average credit constraint status for winners (in blue) and losers (in red).[33] The figure illustrates how a common rising pre-award trend in credit constraints is projected smoothly for the losers up to two years later while flattens for the winners. Hence, winning firms effectively leverage a contract to avoid tightening financial constraints.[34]

[32] Firm size could drive these trends. We verify the role of credit exposure relative to firm size (measured by total assets or current assets) on top of absolute credit stock in a robustness check (not reported). We retrieve very similar findings.

[33] For the sake of clarity, we present the series rescaled to be 0 at $m - 12$. We obtain comparable results using the (unreported) descriptive regression approach as in Figure 3.6.

[34] We stress further that in both Figures 3.6 and 3.7, we do not find evidence of outcome pre-trends: Winners and losers show similar dynamics of credit stock types and credit quality in the twelve months predating the auction. This is not only reassuring for a valid comparison of credit performance across firm types but also

Figure 3.7. Credit Quality Dynamics



Notes: Development of the average Credit Constraint, represented as a binary indicator flagging any bad, impaired, or expired credit status per firm monthly. We rescale the series to be 0 at $m^* - 12$.

Credit and Survival. After presenting empirical evidence linking procurement awards and improving credit performance, we provide evidence linking procurement, credit, and survival dynamics. Essentially, we provide evidence for credit as a channel for the survival premium induced by public procurement. We start by replicating our RD exercise for credit-constrained and credit-unconstrained winners—defined in the month before the auction—separately. The idea is to test whether financially constrained firms receive a higher boost in survival probability. In other words, does the improvement of credit performance induced by an award matter more for firms in financial distress? As reported in Table 3.5, Panel A, the constrained winners experienced a boost in survival odds already twelve months after the auction. The effect increases over time and consistently exceeds the baseline point estimates. Instead, the unconstrained winners do not experience significant boosts in survival for the first two years and only see a positive, smaller effect thirty-six months after the auction. Thus, financially constrained winners drive most of the average survival premium induced by procurement.

The causal interpretation of the exercise requires the pre-auction firms' credit status to be balanced across winners and close losers in order for the credit dimension to be locally orthogonal to the treatment.[35] Figure 3.8 confirms that the credit-constrained status in our data is indeed continuous across winners and losers around the cutoff.

serves as a diagnostic that further corroborates our identification assumption that the two groups are ex-ante balanced.

[35] An ideal exercise would require comparing constrained (unconstrained) winners and close losers (i.e., both runner-up and third-ranked) at the auction level, i.e., the first three ranked bidders share the same credit status at the time of the auction. However, this selection would considerably limit our sample of auctions with insurmountable power problems due to the sparse distribution of the credit constraint dummy in our data—the share of constrained firm-months pairs is 13%. See Appendix 3.E for an in-depth discussion and evidence on the survival effect heterogeneity.

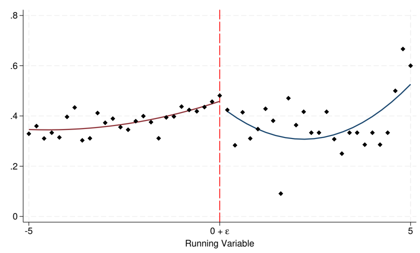
Table 3.5. RD Regressions by Winner's Credit Constraint Status

	Survival		
	m+12	m+24	m+36
Panel A: Credit Constrained	0.032	0.047	0.059
	(0.011)	(0.020)	(0.029)
	0.978	0.951	0.914
	689	*689*	*689*
Panel B: Credit Unconstrained	-0.001	0.010	0.023
	(0.006)	(0.007)	(0.010)
	0.996	0.990	0.967
	1,843	*1,843*	*1,843*

Notes: We replicate Table 3.3 splitting the sample in winners based on winner's credit constraints—binary indicator flagging any bad, impaired, or expired credit status for the firm—the month before auction ($m^{*} - 1$). In panel A, we report constrained firms (credit constraint = 1); in Panel B, we show results for unconstrained (credit constraint = 0).

Also, to further decompose the dynamics of credit constraints and survival across winners and losers, we plot the time series of credit constrainedness, conditioning for firms that faced the same condition at $m^{\star} - 1$. Figure 3.9 shows the average credit constraint dynamics of winners (in blue) and losers (in red) for credit-constrained (Panel A) and credit-unconstrained firms (Panel B). To ease the interpretation of the results, the panels feature histograms depicting survival rates for both groups, color-coded accordingly. In panel A, the credit-constrainedness series shows similar patterns and does not highlight significant differences between winners and losers; however, the mortality rate explains this apparent contradiction, as long as 30 percent of losers exited the market at the end of the sample, as opposed to less than 17 percent of winners (i.e., a remarkable 13 percentage point spread). Hence, firms awarded a public contract are much more able to cope with their credit-constrained status than their losing counterparts, which are instead driven out of the market. On the other hand, the takeaways for panel B are clear: losers face a significantly increased likelihood of becoming constrained, and this effect is immediate—as evidenced by the spread observable as early as $m + 1$. Moreover, losers are forced out of the market at twice the rate of winners within a 36-month period.

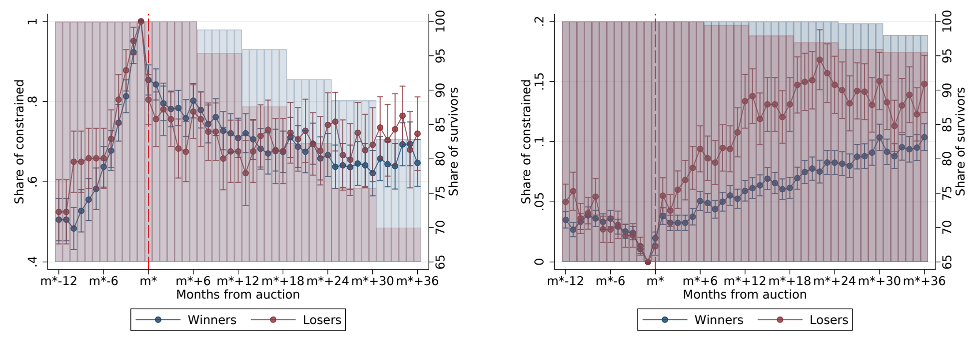
Figure 3.8. Firm Credit Constraints: Winners and Marginal Losers at $m - 1$



Notes: Firm-level characteristics for winners (blue line) and marginal losers (red line, include runners-up and third-ranked) prior to the contract award. The running variable is rescaled to reflect the distance from the runner-up bid ($X_i = B_{i,a} - B_a^2$). The variables is measured at $m - 1$, i.e., a month prior the award.

Taken together, these exercises suggest that credit stands out as a critical mechanism underlying the survival effect of public contracts, playing a role even before the award's survival effect kicks in. Winners receive more bank credit and experience a relaxation of their financial constraints because of the contract. Firms in financial distress benefit more from the award in terms of survival probability, and their improved credit quality is associated with more pronounced lower exit rates.

Figure 3.9. Credit Quality Dynamics and Survival



Notes: Credit Constraint are represented as a binary indicator flagging any bad, impaired, or expired credit status per firm monthly. In the panels, winners (blue) and losers (red) are contrasted based on identical credit issue dummies the month before auction ($m^* - 1$). Constrained firms (credit constraint = 1) are in Panel A; unconstrained (credit constraint = 0) in Panel B. The time series displays average credit constraint from $m^* - 12$ to $m^* + 36$, the values are reported on the y-axis on the left-hand side. Histograms depict exit rates for both firm groups, color-coded accordingly. The values are reported on the y-axis on the right-hand side.

3.6 Conclusions

This paper quantifies the effect of public purchases on supplier survival dynamics. We construct a unique dataset on firms, contracts, and auctions and focus primarily on the survival likelihood—an aspect mostly overlooked in the extant literature, which instead tends to analyze the effects of procurement contracts on business performance *conditioning* on survival. In doing so, we inform the debate on how the public sector's monopsonistic power could be an effective fiscal measure for the government to intervene in the economy and affect business performance.

Our results indicate that winning a contract per se increases a firm's survival probability above and beyond the contract expiration. We show that this effect is associated with a recomposition of revenues from private to public customers rather than a pure scale effect induced by the award. Regardless of size, contracts that are long-lasting, awarded by decentralized buyers, or in industries for which the public sales are more relevant than private opportunities are more impactful for survival prospects. To explain the implications of these results, we examine the impact of public demand on different firm outcomes. Labor productivity is unaffected by the demand shocks. This result suggests that public procurement helps firms stay in the market longer but does not make them more productive. To get a better sense of why procurement firms are not necessarily forced to exit, we rely on evidence showing that public contracting revenues protect them from competing with more efficient firms in the private market (Akcigit, Baslandze, and Lotti, 2023). In addition, we find suggestive evidence that procurement firms survive longer by leveraging public contracts to gain easier access to new credit and improve their credit score. Thus, credit could be a transmission channel for survival dynamics.

Our result suggests that procurements remove a friction that causes firms to shut down: public awards relax credit constraints. Intentionally or not, winners are kept alive by the contract. However, we cannot argue that such government intervention is justified, as we find no evidence of efficiency, e.g., in terms of labor productivity dynamics. Moreover, we are agnostic about spillover effects and the possible crowding-out implications on unexposed companies (Barrot and Nanda, 2020). Our paper is a first step in understanding the impact of demand source for firm survival and paves the way for further research to measure the welfare effects of procurement spending on firm dynamics.

References

Abrantes-Metz, Rosa M, Luke M Froeb, John Geweke, and Christopher T Taylor. 2006. "A variance screen for collusion." *International Journal of Industrial Organization* 24 (3): 467–86. [159]

Agarwal, Rajshree, and David B Audretsch. 2001. "Does entry size matter? The impact of the life cycle and technology on firm survival." *Journal of Industrial Economics* 49 (1): 21–43. [123]

Agarwal, Rajshree, and Michael Gort. 2002. "Firm and Product Life Cycles and Firm Survival." *American Economic Review* 92 (2): 184–90. http://www.jstor.org/stable/3083399. [119]

Akcigit, Ufuk, Salomé Baslandze, and Francesca Lotti. 2023. "Connecting to Power: Political Connections, Innovation, and Firm Dynamics." *Econometrica* 91 (2): 529–64. https://doi.org/https://doi.org/10.3982/ECTA18338. eprint: https://onlinelibrary.wiley.com/doi/pdf/10.3982/ECTA18338. [129, 130, 146, 167]

Albano, Gian Luigi, Berardino Cesi, and Alberto Iozzi. 2017. "Public procurement with unverifiable quality: The case for discriminatory competitive procedures." *Journal of Public Economics* 145: 14–26. https://doi.org/https://doi.org/10.1016/j.jpubeco.2016.11.004. [136]

Arkolakis, Costas, Theodore Papageorgiou, and Olga A Timoshenko. 2018. "Firm learning and growth." *Review of Economic Dynamics* 27: 146–68. [123]

Audretsch, David B, and Talat Mahmood. 1995. "New firm survival: New results using a hazard function." *Review of Economics and Statistics,* 97–103. [124]

Baier, Scott L, Gerald P Dwyer Jr, and Robert Tamura. 2006. "How important are capital and total factor productivity for economic growth?" *Economic Inquiry* 44 (1): 23–49. [122]

Barrot, Jean-Nöel, and Ramana Nanda. 2020. "The Employment Effects of Faster Payment: Evidence from the Federal Quickpay Reform." *Journal of Finance* 75 (6): 3139–73. https://doi.org/https://doi.org/10.1111/jofi.12955. [123, 147]

Bartelsman, Eric, John Haltiwanger, and Stefano Scarpetta. 2009. "Measuring and Analyzing Cross-country Differences in Firm Dynamics." In *Producer Dynamics: New Evidence from Micro Data,* 15–76. University of Chicago Press. http://www.nber.org/chapters/c0480. [119]

Becker, Sascha, Peter Egger, and Maximilian von Ehrlich. 2010. "Going NUTS: The effect of EU Structural Funds on regional performance." *Journal of Public Economics* 94 (9-10): 578–90. https://EconPapers.repec.org/RePEc:eee:pubeco:v:94:y:2010:i:9-10:p:578-590. [123]

Blattner, Laura, Luisa Farinha, and Francisca Rebelo. 2023. "When Losses Turn into Loans: The Cost of Weak Banks." *American Economic Review* 113 (6): 1600–1641. https://doi.org/10.1257/aer.20190149. [141]

Board, Simon. 2007. "Bidding into the Red: A Model of Post-Auction Bankruptcy." *Journal of Finance* 62 (6): 2695–723. [130]

Brata, Aloysius Gunadi, Henri LF De Groot, and Wouter Zant. 2018. “Shaking up the firm survival: Evidence from Yogyakarta (Indonesia).” *Economies* 6 (2): 26. [124]

Byrne, Joseph P, Marina-Eliza Spaliara, and Serafeim Tsoukas. 2016. “Firm survival, uncertainty, and financial frictions: Is there a financial uncertainty accelerator?” *Economic Inquiry* 54 (1): 375–90. [123, 124]

Calligaris, S., Massimo Del Gatto, F. Hassan, Gianmarco Ottaviano, and Fabiano Schivardi. 2016. “Italy’s productivity conundrum. A study on resource misallocation in Italy.” European Commission Discussion Papers 30. https://EconPapers.repec.org/RePEc:euf:dispap:030. [122]

Calveras, A., J.J. Ganuza, and E. Hauk. 2004. “Wild bids. Gambling for resurrection in procurement contracts.” *Journal of Regulatory Economics* 26 (1): 41–68. [130]

Calvino, Flavio, Chiara Criscuolo, and Carlo Menon. 2016. “No Country for Young Firms? Start-up dynamics and national policies.” *OECD Science, Technology and Industry Policy Papers, No. 29,* no. 29. [119]

Calzolari, Giacomo, and Giancarlo Spagnolo. 2009. “Relational Contracts and Competitive Screening.” Working paper DP7434. CEPR Discussion Papers. [136]

Cappelletti, Matilde, and Leonardo M Giuffrida. 2022. “Targeted Bidders in Government Tenders.” Discussion Paper 22-030. ZEW-Centre for European Economic Research. [141]

Cattaneo, Matias D, Rocio Titiunik, Gonzalo Vazquez-Bare, et al. 2020. “The regression discontinuity design.” *Sage Handbook of Research Methods in Political Science and International Relations,* 835–57. [131]

Cattaneo, Matias D, Rocío Titiunik, Gonzalo Vazquez-Bare, and Luke Keele. 2016. “Interpreting regression discontinuity designs with multiple cutoffs.” *Journal of Politics* 78 (4): 1229–48. [132]

Cerqua, Augusto, and Guido Pellegrini. 2014. “Do subsidies to private capital boost firms’ growth? A multiple regression discontinuity design approach.” *Journal of Public Economics* 109: 114–26. [123, 132]

Cevik, Serhan, and Fedor Miryugin. 2022. “Death and taxes: Does taxation matter for firm survival?” *Economics & Politics* 34 (1): 92–112. [124]

Chassang, Sylvain, Kei Kawai, Jun Nakabayashi, and Juan Ortner. 2022. “Robust screens for noncompetitive bidding in procurement auctions.” *Econometrica* 90 (1): 315–46. [158–160, 163]

Choi, Yeol, David W. Marcouiller, Hyun Kim, Jaesong Lee, and Sungho Park. 2021. “What drives survival of urban firms? An asset-based approach in Korea.” *International Journal of Urban Sciences* 25 (4): 574–92. [119, 123]

Cingano, Federico, Paolo Pinotti, Enrico Rettore, and Filippo Palomba. 2022. “Making Subsidies Work: Rules vs. Discretion.” Working paper 2207. Centre for Research and Analysis of Migration (CReAM), Department of Economics, University College London. [128]

Clementi, Gian Luca, and Hugo A Hopenhayn. 2006. “A theory of financing constraints and firm dynamics.” *Quarterly Journal of Economics* 121 (1): 229–65. [119, 123]

Cohen, Lauren, and Christopher J. Malloy. 2016. “Mini West Virginias: Corporations as Government Dependents.” MIMEO. [141]

Conley, Timothy G., and Francesco Decarolis. 2016. “Detecting Bidders Groups in Collusive Auctions.” *American Economic Journal: Microeconomics* 8 (2): 1–38. [131, 157]

Coviello, Decio, Immacolata Marino, Tommaso Nannicini, and Nicola Persico. 2021. “Demand Shocks and Firm Investment: Micro-Evidence from Fiscal Retrenchment in Italy.” *Economic Journal* 132 (642): 582–617. [123]

Criscuolo, Chiara, Ralf Martin, Henry G. Overman, and John Van Reenen. 2019. “Some Causal Effects of an Industrial Policy.” *American Economic Review* 109 (1): 48–85. https://doi.org/10.1257/aer.20160034. [123]

Cull, Robert, and Lixin Colin Xu. 2005. “Institutions, ownership, and finance: The determinants of profit reinvestment among Chinese firms.” *Journal of Financial Economics* 77 (1): 117–46. [167]

Czarnitzki, Dirk, Paul Hünermund, and Nima Moshgbar. 2020. "Public Procurement of Innovation: Evidence from a German Legislative Reform." *International Journal of Industrial Organization* 71: 102620. https://doi.org/https://doi.org/10.1016/j.ijindorg.2020.102620. [123]

De Leverano, Adriano, Robert Clark, and Decio Coviello. 2020. "Complementary Bidding and the Collusive Arrangement: Evidence from an Antitrust Investigation." *ZEW-Centre for European Economic Research Discussion Paper,* nos. 20-052. [159]

De Silva, Dakshina G, Timothy Dunne, and Georgia Kosmopoulou. 2003. "An empirical analysis of entrant and incumbent bidding in road construction auctions." *Journal of Industrial Economics* 51 (3): 295–316. [136]

De Silva, Dakshina G, Georgia Kosmopoulou, and Carlos Lamarche. 2009. "The effect of information on the bidding and survival of entrants in procurement auctions." *Journal of Public Economics* 93 (1-2): 56–72. [120, 124, 136]

De Silva, Dakshina G, Georgia Kosmopoulou, and Carlos Lamarche. 2017. "Subcontracting and the survival of plants in the road construction industry: A panel quantile regression analysis." *Journal of Economic Behavior & Organization* 137: 113–31. [124]

Decarolis, Francesco, Giancarlo Spagnolo, and Riccardo Pacini. 2016. "Past Performance and Procurement Outcomes." Working Paper No. 22814. National Bureau of Economic Research. https://doi.org/10.3386/w22814. [158]

Drozd, Lukasz A, and Jaromir B Nosal. 2012. "Understanding international prices: Customers as capital." *American Economic Review* 102 (1): 364–95. [123]

Dunne, Timothy, Mark J Roberts, and Larry Samuelson. 1989. "The growth and failure of US manufacturing plants." *Quarterly Journal of Economics* 104 (4): 671–98. [119, 123]

Esteve-Pérez, Silviano, Fabio Pieri, and Diego Rodriguez. 2018. "Age and productivity as determinants of firm survival over the industry life cycle." *Industry and Innovation* 25 (2): 167–98. https://doi.org/10.1080/13662716.2017.1291329. eprint: https://doi.org/10.1080/13662716.2017.1291329. [124]

Evans, David S. 1987a. "Tests of alternative theories of firm growth." *Journal of Political Economy* 95 (4): 657–74. [119, 123]

Evans, David S. 1987b. "The relationship between firm growth, size, and age: Estimates for 100 manufacturing industries." *Journal of Industrial Economics,* 567–81. [119, 123]

Fadic, Milenko. 2020. "Letting luck decide: Government procurement and the growth of small firms." *Journal of Development Studies* 56 (7): 1263–76. [123]

Ferraz, Claudio, Frederico Finan, and Dimitri Szerman. 2015. "Procuring Firm Growth: The Effects of Government Purchases on Firm Dynamics." Working Paper, Working Paper Series 21219. National Bureau of Economic Research. https://doi.org/10.3386/w21219. [120, 123]

Foster, Lucia, John Haltiwanger, and Chad Syverson. 2016. "The Slow Growth of New Plants: Learning about Demand?" *Economica* 83 (329): 91–129. https://doi.org/https://doi.org/10.1111/ecca.12172. eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/ecca.12172. [119, 120, 123, 124]

Gelman, Andrew, and Guido Imbens. 2019. "Why high-order polynomials should not be used in regression discontinuity designs." *Journal of Business & Economic Statistics* 37 (3): 447–56. [138]

Giovanni, Julian di, Manuel García-Santana, Priit Jeenas, Enrique Moral-Benito, and Josep Pijoan-Mas. 2022. "Government procurement and access to credit: Firm dynamics and aggregate implications." CEPR Discussion Paper No. 17023. CEPR. [120, 122, 123, 141]

Goldman, Jim. 2020. "Government as customer of last resort: The stabilizing effects of government purchases on firms." *Review of Financial Studies* 33 (2): 610–43. [120, 122, 123, 141]

Gourio, Francois, and Leena Rudanko. 2014. "Customer capital." *Review of Economic Studies* 81 (3): 1102–36. [123]

Gugler, Klaus, Michael Weichselbaumer, and Christine Zulehner. 2020. "Employment behavior and the economic crisis: Evidence from winners and runners-up in procurement auctions." *Journal of Public Economics* 182: 104112. [120, 123, 133]

Hall, Bronwyn H. 1987. "The Relationship Between Firm Size and Firm Growth in the US Manufacturing Sector." *Journal of Industrial Economics* 35 (4): 583–606. Accessed July 4, 2023. http://www.jstor.org/stable/2098589. [119, 123]

Hebous, Shafik, and Tom Zimmermann. 2021. "Can government demand stimulate private investment? Evidence from U.S. federal procurement." *Journal of Monetary Economics* 118: 178–94. https://doi.org/https://doi.org/10.1016/j.jmoneco.2020.09.005. [123]

Hoekman, Bernard, and Marco Sanfilippo. 2018. "Firm performance and participation in public procurement: Evidence from Sub-Saharan Africa." RSCAS Working Papers 2018/16. European University Institute. https://ideas.repec.org/p/rsc/rsceui/2018-16.html. [123]

Hoekstra, Mark. 2009. "The effect of attending the flagship state university on earnings: A discontinuity-based approach." *Review of Economics and Statistics* 91 (4): 717–24. [132]

Horrace, William C., and Ronald L. Oaxaca. 2006. "Results on the bias and inconsistency of ordinary least squares for the linear probability model." *Economics Letters* 90 (3): 321–27. https://doi.org/https://doi.org/10.1016/j.econlet.2005.08.024. [155]

Hsu, Yu-Chin, and Shu Shen. 2019. "Testing treatment effect heterogeneity in regression discontinuity designs." *Journal of Econometrics* 208 (2): 468–86. [166]

Imbens, Guido W, and Thomas Lemieux. 2008. "Regression discontinuity designs: A guide to practice." *Journal of econometrics* 142 (2): 615–35. [166]

Imhof, David, Yavuz Karagök, and Samuel Rutz. 2018. "Screening for Bid Rigging—Does It Work?" *Journal of Competition Law & Economics* 14 (2): 235–61. [159]

Johnson, Simon, and Todd Mitton. 2003. "Cronyism and capital controls: Evidence from Malaysia." *Journal of financial economics* 67 (2): 351–82. [167]

Jovanovic, Boyan. 1982. "Selection and the Evolution of Industry." *Econometrica* 50 (3): 649–70. Accessed July 4, 2023. http://www.jstor.org/stable/1912606. [120]

Kao, Erin Hui-Chuan, and Jin-Tan Liu. 2022. "Extensive margins of trade and firm survival." *Economics Letters* 218: 110716. https://doi.org/https://doi.org/10.1016/j.econlet.2022.110716. [124]

Kawai, Kei, Jun Nakabayashi, Juan Ortner, and Sylvain Chassang. 2022. "Using Bid Rotation and Incumbency to Detect Collusion: A Regression Discontinuity Approach." *Review of Economic Studies* 90 (1): 376–403. https://doi.org/10.1093/restud/rdac013. eprint: https://academic.oup.com/restud/article-pdf/90/1/376/48523410/rdac013.pdf. [131, 136, 160]

Khwaja, Asim Ijaz, and Atif Mian. 2005. "Do lenders favor politically connected firms? Rent provision in an emerging financial market." *Quarterly Journal of Economics* 120 (4): 1371–411. [167]

Kline, Patrick, and Enrico Moretti. 2014. "People, Places, and Public Policy: Some Simple Welfare Economics of Local Economic Development Programs." *Annual Review of Economics* 6 (1): 629–62. https://doi.org/10.1146/annurev-economics-080213-041024. eprint: https://doi.org/10.1146/annurev-economics-080213-041024. [123]

Kolesár, Michal, and Christoph Rothe. 2018. "Inference in Regression Discontinuity Designs with a Discrete Running Variable." *American Economic Review* 108 (8): 2277–304. https://doi.org/10.1257/aer.20160945. [138]

Kong, Yunmi. 2021. "Sequential auctions with synergy and affiliation across auctions." *Journal of Political Economy* 129 (1): 148–81. [132]

Kosmopoulou, Georgia, and Robert Press. 2022. "Supply side effects of infrastructure spending." *Economics Letters,* 110642. https://doi.org/https://doi.org/10.1016/j.econlet.2022.110642. [124]

Lee, David S, and Thomas Lemieux. 2010. "Regression discontinuity designs in economics." *Journal of Economic Literature* 48 (2): 281–355. [136]

Lee, M. 2017. "Government Purchases, Firm Growth and Industry Dynamics." Mimeo. [123]

Leuz, Christian, and Felix Oberholzer-Gee. 2006. "Political relationships, global financing, and corporate transparency: Evidence from Indonesia." *Journal of Financial Economics* 81 (2): 411–39. [167]

Lucas, Adrienne M, and Isaac M Mbiti. 2014. "Effects of school quality on student achievement: Discontinuity evidence from Kenya." *American Economic Journal: Applied Economics* 6 (3): 234–63. [132]

McCrary, J. 2008. "Manipulation of the running variable in the regression discontinuity design: A density test." *Journal of Econometrics.* [134]

OECD. 2019. *Health at a Glance 2019: OECD Indicators.* Paris: OECD Publishing. https://doi.org/10.1787/4dd50c09-en. [120]

Ortiz-Villajos, José M., and Sonia Sotoca. 2018. "Innovation and business survival: A long-term approach." *Research Policy* 47 (8): 1418–36. https://doi.org/https://doi.org/10.1016/j.respol.2018.04.019. [124]

Pop-Eleches, Cristian, and Miguel Urquiola. 2013. "Going to a better school: Effects and behavioral responses." *American Economic Review* 103 (4): 1289–324. [132]

Pozzi, Andrea, and Fabiano Schivardi. 2016. "Demand or productivity: What determines firm growth?" *RAND Journal of Economics* 47 (3): 608–30. https://doi.org/https://doi.org/10.1111/1756-2171.12142. eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/1756-2171.12142. [119, 123, 124]

Ramey, Valerie A. 2019. "Ten years after the financial crisis: What have we learned from the renaissance in fiscal research?" *Journal of Economic Perspectives* 33 (2): 89–114. [119]

Schiantarelli, Fabio, Massimiliano Stacchini, and Philip E Strahan. 2020. "Bank quality, judicial efficiency, and loan repayment delays in Italy." *Journal of Finance* 75 (4): 2139–78. [141]

Sekhri, Sheetal. 2020. "Prestige matters: Wage premium and value addition in elite colleges." *American Economic Journal: Applied Economics* 12 (3): 207–25. [132]

Syverson, Chad. 2011. "What Determines Productivity?" *Journal of Economic Literature* 49 (2): 326–65. https://doi.org/10.1257/jel.49.2.326. [119, 124]

Ugur, Mehmet, and Marco Vivarelli. 2021. "Innovation, firm survival and productivity: The state of the art." *Economics of Innovation and New Technology* 30 (5): 433–67. [119, 122, 140]

Zingales, Luigi. 1998. "Survival of the Fittest or the Fattest? Exit and Financing in the Trucking Industry." *Journal of Finance* 53 (3): 905–38. [119, 123]

Appendices to Chapter 3

Appendix 3.A Additional Tables and Figures

Table 3.A.1. RD Regressions–LATE

	Window	Polynomial	Kernel	Survival m+12	Survival m+24	Survival m+36
Panel A: Baseline	± 5 p.p.	Quadratic	Triangular	0.008	0.019	0.034
				(0.005)	(0.008)	(0.011)
				0.991	0.977	0.957
				2,532	*2,532*	*2,532*
Panel B: LATE specification	± 5 p.p.	Quadratic	Triangular	0.010	0.021	0.024
				(0.005)	(0.008)	(0.012)
				0.991	0.977	0.958
				2,221	*2,221*	*2,221*
Panel C: LATE specification	± 1 p.p.	Quadratic	Triangular	0.006	0.015	0.032
				(0.005)	(0.009)	(0.011)
				0.993	0.982	0.968
				1,706	*1,706*	*1,706*
Panel D: LATE specification	± 9 p.p.	Quadratic	Triangular	0.008	0.015	0.024
				(0.005)	(0.007)	(0.012)
				0.991	0.976	0.955
				2,490	*2,490*	*2,490*

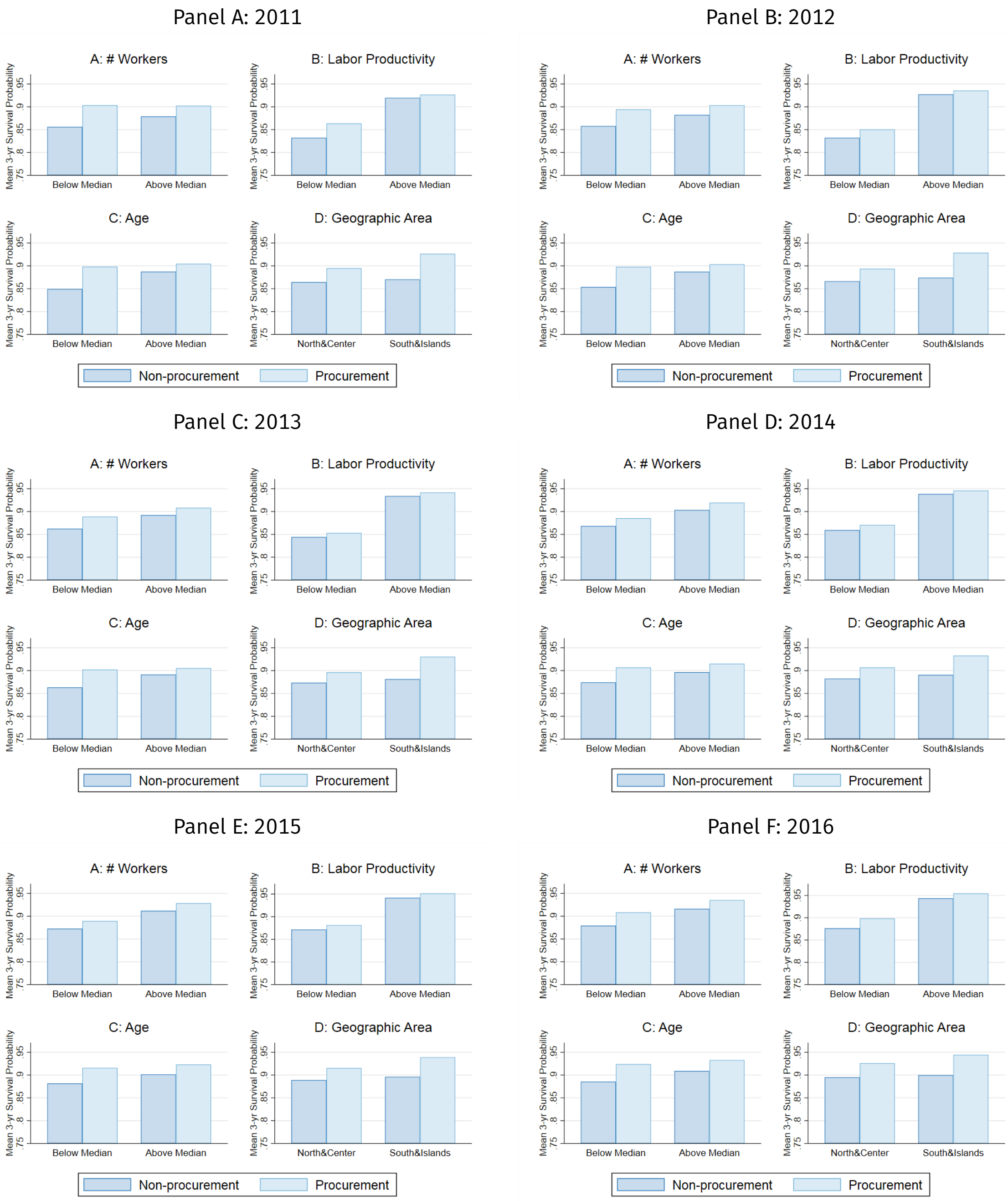
Notes: The baseline estimates of Table 3 in the paper (Panel A) and the estimation of the Local Average Treatment Effect obtained by excluding runners-up and third-ranked in the auctions where the winner is ruled out because beyond the window specific to the regression (Panel B). Panels C and D report the same specification with 1 and 9 percentage points window, respectively.

Table 3.A.2. Analysis Sample - Price and Quality Auctions

	Price Auctions		Quality Auctions	
	Mean	Median	Mean	Median
Average Bid	24.91	26.42	66.61	67.33
St.Dev.	3.13	2.38	14.89	13.11
Minimum	17.68	17.17	47.10	48.26
Maximum	28.16	28.68	85.32	88.50
ABA Share	0.82	1.00	0.00	0.00
Observations	888		359	

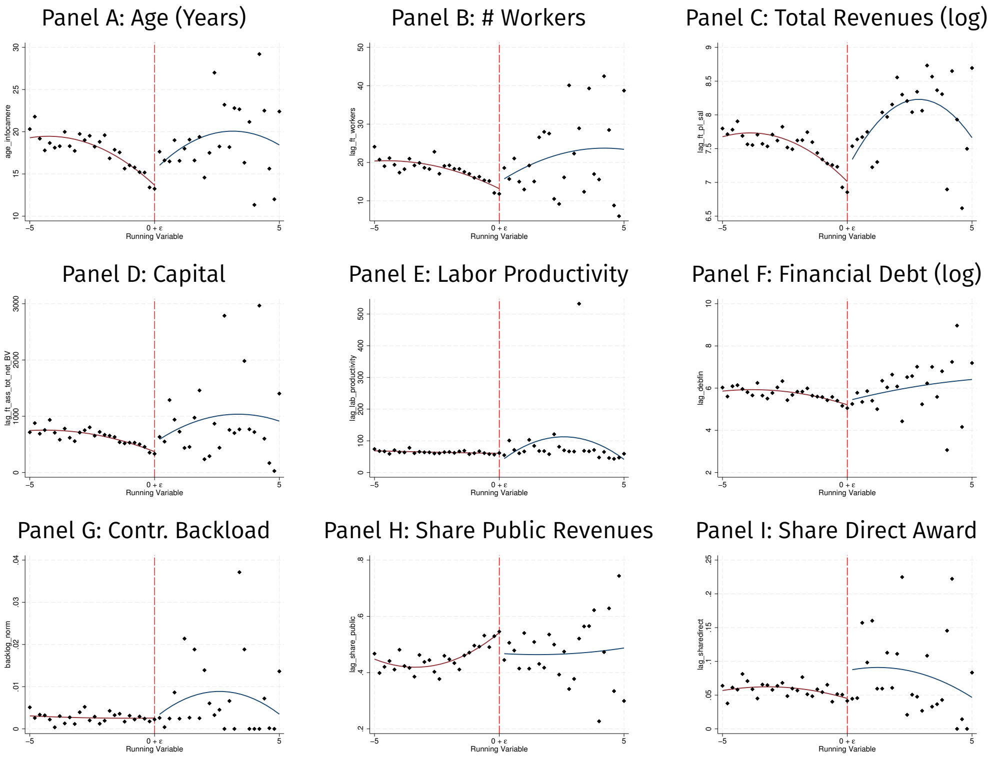
Notes: For the baseline RD sample, we report the mean and the median of the average, standard deviation, minimum and maximum bid across price (columns 1 and 2) and quality auctions (columns 3 and 4), alongside the share of ABAs (vs. FPAs) for the price-based auctions.

Figure 3.A.1. Procurement vs. non-Procurement Firms – Average Survival Rate



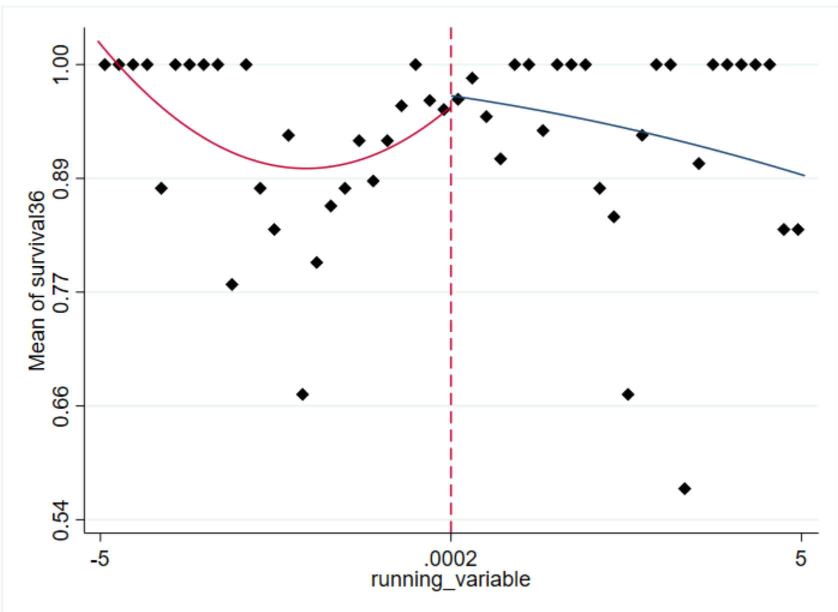
Notes: We report the average three-year survival rate for procurement and non-procurement firms over time. We define the former as those firms that received at least one procurement contract in the three years prior to year *t*, and the latter as the complement. In each subpanel, we partial out a predictor of survival: the number of workers (A), the labor productivity (B), firm age (C), and geographic area (D). Labor productivity is defined as value-added divided by the number of workers. Panel F reports Figure 3.1.

Figure 3.A.2. (Discontinuous) Firm Characteristics': Winners and Marginal Losers at $t-1$ (All Bidders)



Notes: We replicate Figure 3.3 including all bidders.

Figure 3.A.3. Visual Representation of the RD estimates



Notes: This is a visual representation of the baseline RD estimates for the survival at $m+36$ (Table 3.3). Accordingly, bin size (i.e., the bandwidth minimizing the MSE: approximately 0.2 p.p.) and window (i.e., 5 p.p. on either side of the cutoff) are implemented. Note that the threshold of the RD is set at '0+ ε', which is here reported in the graph at 0.0002.

Appendix 3.B Regression-based Evidence

We exploit the wealth of our data to check whether the association of survival and public demand is confirmed in a correlation fashion in the full sample of firms and contracts. We run a static, regression-based exercise that leverages the wealth of our data and looks at the *conditional* survival probability. More specifically, we regress an indicator function for firm staying in the market j years ahead—$\mathbb{I}(Surv)_{i,s,t+j}$ where $j \in [2;3]$—against an indicator variable for any contract awards in the year (i.e., $\mathbb{I}\left\{PubWinner_{i,t}\right\}$), plus observables. The resulting linear probability model reads:

$$\begin{aligned}\mathbb{I}(Surv)_{i,s,t+j} \;=\; & \alpha + \beta_1 \mathbb{I}\left\{PubWinner_{i,t}\right\} + \\ & + \beta_2 LabProductivity_{i,t} + \beta_3 \#\ \text{Workers}_{i,t} + \zeta_i + \zeta_{t,s} + \epsilon_{i,s,t},\end{aligned} \tag{3.B.1}$$

where $LabProductivity_{i,t}$ is labor productivity while ζ_i and $\zeta_{t,s}$ are firm and year-sector fixed effects, respectively.[36] Let β_1 be the parameter of interest capturing the effect of being awarded at least a contract at t on the probability of staying in the market, conditional on firm size, productivity, and all sector- and local-related characteristics captured by the battery of fixed effects.[37]

In Table 3.B.1 column (1), we report the results for two years of survival: $\hat{\beta}_1$ is positive and strongly significant, meaning that awards make less likely for the same firm to exit the market. Its effect amounts to 2.4 p.p.—i.e., about half of the mean two-years exit probability—and appears to be a major driver of business survival.

Even though interpreting linear probability parameters as marginal effects is a challenging exercise for a number of reasons (e.g., Horrace and Oaxaca, 2006), our results indicate that, all else equal, being awarded at least one public contract in a given year is associated with a boost in survival probability corresponding to that of expanding the employment by around 2,400 employees. The table shows another remarkable fact: Firm scale—proxied by employment—and productivity predict survival but matter more when the firm is exposed to public procurement contracts (column 2 versus column 3). Columns (4) to (6) replicate the same analysis for a three-year survival with overlapping takeaways.

Appendix 3.C Robustness Analysis

We propose multiple empirical exercises to corroborate the robustness of our baseline findings. We report the mentioned tables and figures in the bottom of this appendix.

Model specifications. We run the analysis with alternative model specifications to test whether the results are sensitive to arbitrary choices on the functional form. In particular,

[36] The procurement sector is represented by the 2-digits CPV code.

[37] Our fixed-effect model makes age collinear with firm fixed effects as age mechanically increases by one every year for every firm and we omit it from this model.

Table 3.B.1. Firm Survival – Static Regressions, Full Sample

	Two-Years			Three-Years		
		Proc	Non-proc		Proc	Non-proc
	(1)	(2)	(3)	(4)	(5)	(6)
1{PubWinner}	0.024			0.022		
	(0.001)			(0.001)		
# Workers (000)	0.001	0.014	0.001	0.001	0.021	0.001
	(0.000)	(0.003)	(0.000)	(0.000)	(0.004)	(0.000)
Labor Productivity (000)	0.003	0.007	0.003	0.018	1.386	0.018
	(0.003)	(0.051)	(0.003)	(0.010)	(0.771)	(0.010)
Observations	4,544,345	192,917	4,320,525	4,068,993	169,433	3,869,487
Unique Firms	738,141	42,749	726,836	696,572	39,800	685,197
Mean Y	0.95	0.98	0.95	0.92	0.95	0.92
Year*Sector FE	✓	✓	✓	✓	✓	✓
Firm FE	✓	✓	✓	✓	✓	✓

Notes: Columns 1 reports the results of Equation (3.B.1) on the full sample using two and three years as the survival horizon. Results are replicated for procurement firms (columns 2 and 5) and non-procurement firms (columns 3 and 6), respectively. The observations are at the firm-year level. All models feature year-sector and firm fixed effects. Standard errors are clustered at the year-area-sector level.

we show the results when using either a local linear regression on each side of the cutoff (Table 3.C.1, Panel B) or an alternative non-parametric specification of the kernel (i.e., Epanechnikov, Panel C). In both cases, we obtain very similar results, both qualitatively and quantitatively. We also change the window of scores around the threshold: In Panels D and E, we restrict and extend the window by 4 p.p. on both sides of the cutoff, respectively. The more we zoom in on the score space around the cutoff, the more we keep auctions where the first three bids are very close, the more the number of observations decreases relative to the baseline. However, the smaller the window, the larger and more significant (but also more local) the estimated effect. When using a ±1-p.p. window, the results on +36 months increase in magnitude, indicating a stronger survival boost. When we expand the window to ±9 p.p around the cutoff, or when we do not impose any window in the running variable space and employ a robust bias-corrected RD—as in Panel F—the estimates hold comparable nonetheless. The inclusion of less comparable firms does not seem to affect our estimates but does indeed affect their validity. We then keep our preferred window specification of ±5 p.p. to maximize the trade-off between the locality of estimates and their validity.

Contamination. A concern for the identification assumption of as-good-as-random award relates to contamination. Although losers and winners are similar in terms of pre-treatment exposure to public procurement, the contamination problem is underpinned by the fact that the longer the period after the award event, the greater the chances for both losers and winners to win other contracts (control and treatment contamination, respectively). In this scenario, the comparison between losers and winners could become increasingly contaminated over time.

In Table 3.C.1, Panels G-K, we propose a series of exercises to show the robustness of our estimates to this contamination problem. Panel G shows the scenario with no contamination in the control group—i.e., excluding *all* runners-up and third-ranked that do not receive a contract starting at t through the following three calendar years. To perform this exercise, we use the entire OpenANAC data to make the firm selection independent of the analysis sample of contracts. As expected, the survival coefficient is much stronger—as we compare the winners (who may receive more contracts) to the 'never winners'—but the sample size decreases dramatically; never-winners are indeed few—i.e., 16.8% of the firms in the analysis sample). However, these point estimates should be taken with caution: although they show the remarkable robustness of our baseline results, the increase in the parameter value could be due to the adverse selection of poor quality controls. By restricting attention to firms that do not receive a future contract we could boost the treatment parameter 'endogenously.' In Panel H, we propose a mirror approach, i.e., we exclude only winners that received other contracts ('no-more-winners'). As expected, the mechanical absence of contamination in the treatment but the presence of contamination in the control quickly causes the comparison of no-more-winners to 'winning-losers' to become nonsignificant.

These findings speak for themselves: there is a risk of contamination both in treatment and control groups and the issue becomes bigger as long as we focus on longer outcome leads. If we jointly remove control and treatment contamination, we are left with no sufficient power for estimation; yet, even if we could, we would be including two-layers of suppliers selection which would make the comparison biased in an unpredictable way.

An alternative approach is to *control for contamination*. In fact, if there is a risk of contamination but alike in both control and treated firms, our baseline RD would be estimating the local average treatment effect in an unbiased manner as comparing firms with similar future exposure to procurement contracts *excluding* the contract under analysis. In Table 3.C.2 we replace the survival outcome variable with a binary indicator signalling at least one award after t and up to $t+m$, for all leads, and could not estimate any significant effect. In other words, the probability of obtaining contracts after time t is the same for control and treated firms, and such zero-effect confirms that winners and losers around the cutoff are in fact fully comparable, also in terms of exposure to public procurement sales after auction a. Thus, despite the risk of contamination, our RD design is capable of estimating the effect of public demand at t for winners.

Auction rules. The next robustness tests pertain to the variation in auction rules. Given that our auction sample includes diverse award (e.g., FPA or ABA) and selection mechanisms (e.g., price or price-quality), there is potential for such auction rule heterogeneity to influence our findings, raising concerns about their validity. In Panel I of Table 3.C.1, we exclude the 2017 auctions from the sample as the rules of ABA changed slightly in May 2017. From then on, before opening the sealed bids, the buyer proceeded with a random draw among five criteria to assess an offer as anomalously low and some criteria were not coherent with the definition of the ABA mechanism discussed in Section 3.3.2.[38] In Panel J, we use an alternative definition of the runner-up for the ABAs. The ABA mechanism can yield situations where the absolute distance between the winning and the runner-up bid is

[38] The details on the 'new' ABAs are presented by Conley and Decarolis (2016).

larger than the absolute distance between the winning and the nearest excluded bid.[39] We define the cutoff using the absolute-nearest bid instead of the baseline ABA's runner-up bid. The further specification does not induce a different pattern in the results. As an additional robustness check for the winner and runnerup-definition, we exclude in Panel K consortia from the sample of winners and losers and we obtain similar results, both qualitatively and quantitatively. These exercises confirm the robustness of the baseline findings against the risks for the validity of the RD associated with auction rules.

Manipulation. To corroborate the as-good-as-random treatment assumption, we need to prove that firms do not manipulate the assignment around the cutoff—i.e., firms behave competitively. More specifically, since bidders' ranking is key to selecting treated and control bidders, we require that firms do not agree on manipulating their ranking strategically. If collusive agreements are at play, bidders are more likely to change their bid and ranking strategically and be assigned to their preferred treatment condition. The presence of cartels in our sample of auctions could be an issue depending on the interplay between a bid-rigging strategy around the threshold and the award mechanism. If the manipulation only occurs among losing bidders, though, this would not undermine the correct identification.

Ideally, we would like to exclude from the sample all auctions in which bidders are found to be part of collusive agreements. In the absence of such ideal records, we propose a series of empirical exercises to corroborate the lack of manipulation. Considered altogether, these exercises suggest that our findings are robust against manipulation concerns. We structure our argument in three complementary parts.

Regression-based exercises. The stability of a cartel is arguably more likely when 'the cake is shared', that is, when all members are awarded a contract at some point in time. As a result, we would expect cartels' members to win at least one contract every year. In Panel A of Table 3.C.3, we repeat the baseline RD exercise excluding all auctions whose runners-up or third-ranked bidders win another contract in the same year of the award under analysis. To implement this, as for the contamination exercises above, we employ the entire OpenANAC data to make firm selection independent from the analysis sample of contracts. By excluding the 'winning losers' at time t, we exclude auctions potentially awarded to cartel members from the sample and only keep firms that participate in contests that are more likely to be competitive. The effects are stronger and more significant despite the halved sample size. Panel G of Table 3.C.1 presents the ideal exercise to exclude collusive practices in the auctions from the viewpoint of an eventual cartel's stability: we keep runners-up and third-ranked bidders that are never awarded a contract until $t + 3$, despite the multiple award opportunities over time. The effect holds stronger and the results are already commented above.

The second regression exercise we propose is inspired by the results of Decarolis, Spagnolo, and Pacini (2016) and Chassang et al. (2022), considered together. The former discuss

[39] For instance, see 'b20' in Figure 3.2.

how the risk of collusive behavior in Italian public procurement auctions is particularly relevant for ABAs, as they provide vigorous incentives to manipulate the bid distribution. Since the rules allow each firm to submit at most one bid, firms that submit multiple bids must game the system by creating shadow subsidiaries. Alternatively, a bidder may also seek to coordinate with other companies to form a bidding ring and pilot TM2 (see Section 3.3.2). For the strategy to work, cartel members must participate in a sufficient number. By contrast, non-coordinating firms do not have incentives to participate jointly. However, it is a safe strategy to focus only on FPAs where rigging bids do not entail manipulation of the average bid. We report the relative results in Panel B of Table 3.C.3: The medium-term effects are bigger in magnitude despite the much-restricted sample. Conversely, despite the larger sample, when focusing on ABAs only, the effect tend to dilute.

According to the collusion detection literature, a signal of bid-rigging in FPAs would be the variance of all bids (Abrantes-Metz et al., 2006), which is not necessarily located around the threshold.[40] To corroborate these results on the FPA sample, we propose below an empirical exercise based on the frontier collusion detection tool from Chassang et al. (2022), whose takeaway is no evidence consistent with the null hypothesis of collusion in the FPAs in the analysis sample. This is understandable as long as cartel members have the possibility of participating in ABAs, where bid-rigging was easier. Finally, Panel C splits the sample depending on the number of submitted bids below versus equal and above 10—the latter being a necessary (but not sufficient) condition for the procurers to opt for an ABA. In the case of more competed contracts—regardless the awarded mechanism—the effect on survival is weaker, consistently with the idea that firms more likely to manipulate procurement awards concentrate in auctions where the probability of ABA implementation—and the actual employment of collusive schemes—is positive.

Collusion detection algorithm. Figure 3.C.1 replicates the visual test for collusion proposed by Chassang et al. (2022). When a cartel participates in a first-price sealed-bid procurement auction, colluding firms designate a winner among themselves and have the other firms submit intentionally losing bids. To decrease the chance of error and increase the cost of betraying the cartel, especially in a very competitive market, Chassang et al. (2022) and Imhof, Karagök, and Rutz (2018) argue that the difference between the designated winning bid and others is typically larger than it would be in a collusion-free auction.

The idea is that colluding firms rig the planned-to-be-losing bids, but they might do so far away from the designated winning bid. This creates a suspicious drop in the density of the bid-to-bid distance around zero. Chassang et al. (2022) exploit this behavior to detect collusion by plotting the distribution of Δ, the proportional difference between each bid and the winning bid in that auction.[41] A fair and competitive auction will show increasing bid density as this difference approaches zero, while a colluded auction will exhibit missing mass near $\Delta = 0$.

[40] This pattern is observed in the field. De Leverano, Clark, and Coviello (2020) show that the collapse of a cartel in the road pavement market in Montreal after the start of the investigation caused the standard deviation of bid differences in auctions to increase dramatically.

[41] For the winning bid, the difference is from the second-lowest, this creating negative values of Δ'.

Unlike the results from Chassang et al. (2022) for Japanese auctions, and despite focusing on the public construction sector as well, we observe no missing bids near $\Delta = 0$—suggesting that the behavior in our sample of FPAs is not the same as in the auctions in Japan. Our data exhibit the highest bid density slightly above zero, suggesting that many auctions have one or more losing bids very close to the winner—inconsistent with the behavior seen in the source paper, where collusive firms arrange for intentional losing bids to be significantly higher than the designated winner's bid. The lack of missing mass near $\Delta = 0$ persists even if we only consider the subset of bids greater than $-0.10 < \Delta < 0.10$ of the reserve price, where the incentive to collude is highest. However, the distribution of bids is significantly wider in our context than in the data used in Chassang et al. (2022).

In the paper, the bulk of observations were contained in the interval $-0.05 < \Delta < 0.05$ p.p. of the reserve price. The authors note that this is usually associated with a very competitive market and one where a small change in bid is associated with a large change in expected profit. The distribution has higher kurtosis, due to heavier tails in our data, and we believe this has two implications. First, there would be less incentive to collude since an efficient firm could take advantage of low competition to increase profits without resorting to collusion. Second, if collusion *were* present, it would be less important that the cartel enforces a 'no-bid mass near zero' rule since the incentive to deviate is lower. Panel B of Figure 3.C.1 further examines the density falloff with a window three times larger than Chassang et al. (2022)'s (i.e., $-0.15 < \Delta < 0.30$ instead of $-0.05 < \Delta < 0.05$) with overlapping conclusions.

Pre-treatment firm characteristics. Continuous firm features in the vicinity of the cutoff exclude the presence of shill bidders created by cartels to better manipulate the allocation, particularly the average bid in the case of ABAs. A shill bidder is a firm created only for this illegal purpose and closed down afterward; therefore, it is hardly comparable with established 'real' firms. Following Kawai et al. (2022), observing a discontinuity at the threshold in the level of pre-award backlog—as defined in Section 3.3.3—can be also a sign of bid ridding. Indeed, the backlog proxies the costs of participation in the auction and, in the case of a cartel using bid rotation, all else equal, those with a higher backlog might be less likely to win in a given auction. Given that we find no sign of discontinuity at the cutoff, we can further rule out concerns on possibly colluding bidders.

Placebo cutoff. We run a battery of placebo RD regressions that replicate the baseline model and the functional-form robustness checks (i.e., Table 3.C.1, Panels A to F), by ruling out the winners from the sample and replacing them with the runners-up. The third-ranked in the original regression sample turn to be the runners-up in the placebo exercise, and so forth. The results, reported in Table 3.C.4, show that all of the coefficients, except for one, are no longer statistically significant, which advocates that the effect identified in our regressions are indeed triggered by the exogenous demand shock rather than by other confounding factors.

Adding covariates. As our RD design is correctly specified, we expect that adding covariates should not change the treatment effect estimate substantially but might reduce the standard errors. In Table 3.C.5, we add as controls in the RD model alternatively the nine firm-level variables measured at $t-1$, which we used to show pre-treatment similarity of bidders in Section 3.3.3. In Table 3.C.6, we instead include seven contract- and auction-level controls, namely contract duration, reserve price, reserve price over revenues, reserve price over employment, North-regions dummy (vs. rest of the country), type of contracting authority (i.e., central vs. local government), and construction category (i.e., buldings vs. other constructions). All these exercises but one—i.e, controlling for share of direct procurement awards, which restricts the sample to procurement winners in the year t mechanically—show robust estimates of τ over the time leads.

Table 3.C.1. RD Regressions—Specification, Contamination, and Auction-rules Robustness Checks

				Survival		
	Window	Polynomial	Kernel	m+12	m+24	m+36
Panel A: Baseline	± 5 p.p.	Quadratic	Triangular	0.008	0.019	0.034
				(0.005)	(0.008)	(0.011)
				0.991	0.977	0.957
				2,532	*2,532*	*2,532*
Panel B: Linear	± 5 p.p.	Linear	Triangular	0.008	0.017	0.032
				(0.005)	(0.007)	(0.011)
				0.991	0.977	0.957
				2,532	*2,532*	*2,532*
Panel C: Epanechnikov	± 5 p.p.	Quadratic	Epanechnikov	0.009	0.019	0.033
				(0.005)	(0.008)	(0.012)
				0.991	0.977	0.957
				2,532	*2,532*	*2,532*
Panel D: 1 percentage point	± 1 p.p.	Quadratic	Triangular	0.009	0.018	0.047
				(0.003)	(0.009)	(0.011)
				0.993	0.982	0.967
				2,166	*2,166*	*2,166*
Panel E: 9 percentage points	± 9 p.p.	Quadratic	Triangular	0.008	0.015	0.025
				(0.005)	(0.007)	(0.011)
				0.991	0.976	0.954
				2,681	*2,681*	*2,681*
Panel F: All percentage points (optimal bandwidth)	All p.p.	Quadratic	Triangular	0.008	0.018	0.044
				(0.004)	(0.008)	(0.012)
				0.989	0.973	0.948
				2,878	*2,878*	*2,878*
Panel G: No contamination (control)	± 5 p.p.	Quadratic	Triangular	0.025	0.067	0.131
				(0.015)	(0.024)	(0.041)
				0.996	0.986	0.969
				270	*270*	*270*
Panel H: No contamination (treatment)	± 5 p.p.	Quadratic	Triangular	-0.000	0.004	-0.008
				(0.013)	(0.026)	(0.036)
				0.991	0.975	0.953
				675	*675*	*675*
Panel I: Without 2017	± 5 p.p.	Quadratic	Triangular	0.009	0.017	0.024
				(0.005)	(0.008)	(0.011)
				0.991	0.978	0.958
				2,236	*2,236*	*2,236*
Panel J: Alternative runner-up	± 5 p.p.	Quadratic	Triangular	0.009	0.019	0.035
				(0.005)	(0.007)	(0.011)
				0.992	0.977	0.958
				2,607	*2,607*	*2,607*
Panel K: No Consortia	± 5 p.p.	Quadratic	Triangular	0.008	0.018	0.032
				(0.005)	(0.008)	(0.012)
				0.991	0.977	0.957
				2,446	*2,446*	*2,446*

Notes: Table 3.3 is replicated using different bandwidth, polynomials, kernels, and sample selections. For all specifications, we use the bandwidth minimizing the MSE.

Table 3.C.2. RD Regressions—Productivity

	Outcome		
	t+1	t+2	t+3
Winning a Contract	0.041	0.011	0.027
	(0.032)	(0.033)	(0.035)
	0.522	0.486	0.454
	2,532	*2,532*	*2,532*

Notes: Table 3.3 is replicated using the winning indicator variable as an outcome.

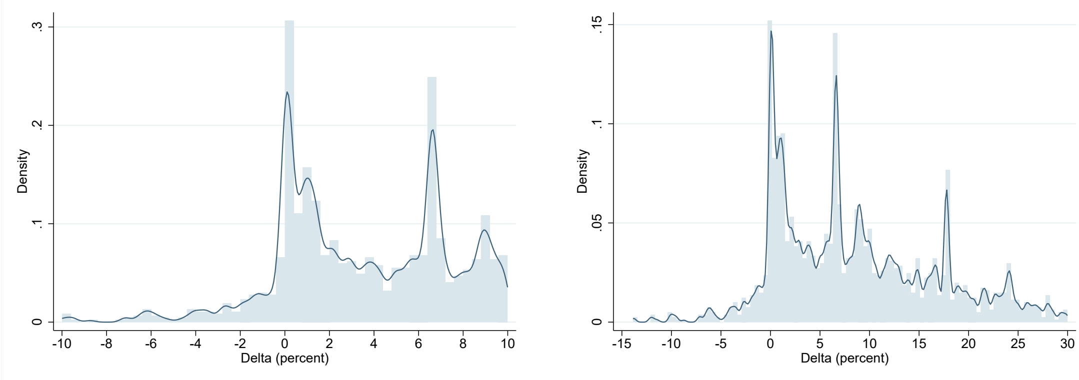
Table 3.C.3. RD Regressions – Collusion Checks

	Survival		
	m+12	m+24	m+36
Panel A: "Cake is shared"	0.019	0.025	0.049
	(0.009)	(0.012)	(0.018)
	0.983	0.966	0.943
	1,252	*1,252*	*1,252*
Panel B: Auction Type			
FPA	-0.000	0.009	0.036
	(0.000)	(0.009)	(0.017)
	0.984	0.966	0.922
	344	*344*	*344*
ABA	0.007	0.012	0.015
	(0.005)	(0.008)	(0.012)
	0.991	0.977	0.957
	1,672	*1,672*	*1,672*
Panel C: Number of Offers			
$<$ 10 bids	0.012	0.036	0.045
	(0.006)	(0.010)	(0.048)
	0.988	0.974	0.931
	731	*731*	*731*
$\geq$ 10 bids	0.008	0.015	0.026
	(0.005)	(0.008)	(0.012)
	0.992	0.978	0.959
	1,801	*1,801*	*1,801*

Notes: RD estimates are executed keeping auctions in which losers are not awarded other contracts at *t* (Panel A), splitting the sample of auctions conditioning on the award format (Panel B) and the number of bids received (below 10 versus 10 or above, Panel C). Given our selection, the number of auctions in each regression corresponds to one-half to one-third of the observations, depending on the share of auctions with two participants (winner and runner-up only) or more (third-ranked also). We replicate in each subsample Table 3.3.

Figure 3.C.1. Chassang et al. (2022)'s Visual Test for Collusion – Distribution of Bid Difference in FPAs

Panel A: $-0.10 < \Delta < 0.10$ Panel B: $-0.15 < \Delta < 0.30$



Notes: Distribution of bid difference Δ from the winning bid. The parameters of the estimated density match those of the original paper, with a smoothing width of 0.75%. Panel A includes all bids: Panel B focuses on bids within 10 p.p. discount from the threshold.

Table 3.C.4. RD Regressions—Placebos

	Window	Polynomial	Kernel	Survival m+12	m+24	m+36
Panel A: Baseline	± 5 points	Quadratic	Triangular	-0.004	-0.008	0.014
				(0.007)	(0.011)	(0.014)
				0.991	0.977	0.957
				2,417	*2,417*	*2,417*
Panel B: Linear	± 5 points	Linear	Triangular	-0.003	-0.008	0.014
				(0.007)	(0.011)	(0.014)
				0.991	0.977	0.957
				2,417	*2,417*	*2,417*
Panel C: Epanechnikov	± 5 points	Quadratic	Epanechnikov	-0.004	-0.008	0.014
				(0.008)	(0.011)	(0.014)
				0.991	0.977	0.957
				2,417	*2,417*	*2,417*
Panel D: 1 point	± 1 point	Quadratic	Triangular	0.009	0.009	0.036
				(0.007)	(0.012)	(0.015)
				0.993	0.982	0.966
				2,067	*2,067*	*2,067*
Panel E: 9 points	± 9 points	Quadratic	Triangular	-0.001	-0.003	0.017
				(0.007)	(0.010)	(0.013)
				0.991	0.976	0.954
				2,547	*2,547*	*2,547*
Panel F: All points (optimal bandwidth)	All points	Quadratic	Triangular	-0.002	-0.005	0.015
				(0.007)	(0.011)	(0.014)
				0.989	0.973	0.948
				2,686	*2,686*	*2,686*

Notes: We replicate Table 3.C.1 by dropping the winning bid from each auction, replacing it with the runner-up, establishing the third-ranked as the runner-up and the fourth-ranked as the third-ranked.

Table 3.C.5. RD Regressions—Pre-treatment Firm-level Covariates

	Survival		
	m+12	m+24	m+36
Panel A: Age (Years)	0.000	0.010	0.022
	(0.004)	(0.008)	(0.012)
	0.995	0.983	0.964
	2,264	*2,264*	*2,264*
Panel B: # Workers	0.001	0.011	0.022
	(0.004)	(0.008)	(0.012)
	0.995	0.983	0.963
	2,202	*2,202*	*2,202*
Panel C: Revenues (€,000)	0.001	0.012	0.024
	(0.004)	(0.008)	(0.012)
	0.995	0.983	0.963
	2,202	*2,202*	*2,202*
Panel D: Capital (€,000)	0.001	0.012	0.024
	(0.004)	(0.008)	(0.012)
	0.995	0.983	0.963
	2,202	*2,202*	*2,202*
Panel E: Labor Productivity	0.001	0.012	0.025
	(0.004)	(0.008)	(0.012)
	0.995	0.983	0.963
	2,183	*2,183*	*2,183*
Panel F: Financial Debt (€,000)	0.000	0.008	0.022
	(0.005)	(0.010)	(0.014)
	0.995	0.982	0.962
	1,754	*1,754*	*1,754*
Panel G: Contracting Backlog	0.002	0.012	0.025
	(0.004)	(0.008)	(0.012)
	0.995	0.983	0.964
	2,262	*2,262*	*2,262*
Panel H: Share Public Revenues	0.006	0.016	0.025
	(0.005)	(0.009)	(0.013)
	0.994	0.980	0.960
	2,165	*2,165*	*2,165*
Panel I: Share Direct Award	0.009	0.012	0.009
	(0.007)	(0.010)	(0.015)
	0.992	0.977	0.956
	1,784	*1,784*	*1,784*

Notes: We replicate Table 3.3 by separately adding firm covariates measured at $t-1$ from Figure 3.3.

Table 3.C.6. RD Regressions With Contract-level Covariates

	Survival		
	m+12	m+24	m+36
Panel A: Duration (Days)	0.008	0.012	0.024
	(0.006)	(0.008)	(0.013)
	0.991	0.977	0.957
	2,154	*2,154*	*2,154*
Panel B: Amount (€, 000)	0.008	0.019	0.035
	(0.005)	(0.008)	(0.011)
	0.991	0.977	0.957
	2,519	*2,519*	*2,519*
Panel C: Amount/Revenues	0.001	0.012	0.025
	(0.004)	(0.008)	(0.012)
	0.995	0.983	0.964
	2,251	*2,251*	*2,251*
Panel D: # Bids	0.009	0.017	0.024
	(0.005)	(0.008)	(0.011)
	0.991	0.977	0.957
	2,532	*2,532*	*2,532*
Panel E: Geographic Area	0.006	0.014	0.025
	(0.005)	(0.008)	(0.011)
	0.991	0.977	0.957
	2,514	*2,514*	*2,514*
Panel F: Buyer Type	0.009	0.019	0.034
	(0.005)	(0.008)	(0.011)
	0.991	0.977	0.957
	2,517	*2,517*	*2,517*
Panel G: CPV	0.008	0.019	0.034
	(0.005)	(0.008)	(0.011)
	0.991	0.977	0.957
	2,519	*2,519*	*2,519*

Notes: We replicate Table 3.3 by separately adding reported contract covariates.

Appendix 3.D Extraction Procedure for Tender Documents

In order to extract the information on the distribution of the bids—from the PDFs documentation provided by Telamat—we had to proceed in several steps. We started with downloading tenders' outcomes PDFs from Telemat's website using Python. In particular, we downloaded only those present both in the Telemat and BDAP database, as the latter data provided us with the name and tax number of auctions participants necessary for the merge with CADS-firm data. The merged data consisted of 11,079 unique contracts. As the documents were not standardized, we had to proceed in several steps. First of all, we had to select the documents containing the list of bids. Note that the downloaded PDFs were more than the number of contracts as, for each contract, more than a document can be produced by the contracting officer. Using Python, we searched among the over 16,000 downloaded documents (corresponding to 10,000 contracts) to select only those containing the list of participants, which BDAP provided. As the documents were not standardized, this was the only characteristic that all PDF documents with the distribution of bids have in common. Then, the 8,348 Python-selected documents for such contracts were inspected manually and with Python, and the bids placed by each auction participant were recorded to create a unique dataset. Given that placed bids appear in a table, we mainly used the package Camelot in Python to extract tables containing the bids from 3,686 machine-readable documents. We had to proceed with manual data extraction for about 4,580 PDFs, namely for those documents that were scanned PDFs and were therefore not machine-readable. However, not all the Python-selected documents reported the bids information, as many reported only the participants' list but not the placed bids. We were able to retrieve bids information for 1,896 contracts (about 16% of the sample for which we had participant information).

Appendix 3.E Which Contracts Matter for Firm Survival?

Our sample of construction contracts is highly heterogeneous in terms of size, content, and buyer characteristics. To investigate heterogeneous treatment effects, we split the data into a constellation of subsamples and separately estimate τ from Equation (3.3), providing estimates conditional on a specific set of observable orthogonal contract characteristics.

We perform a subsample-regression approach—in opposition to an interaction-term analysis conditioning on pre-treatment outcomes—in order to accommodate concerns over multiple testing and invalid inference on heterogeneity in the sharp RD framework (Imbens and Lemieux, 2008). Moreover, we consider the lack-of-power-versus-coarseness trade-off raising from subsampling in an RD framework. On the one hand, if the subgroups are too finely discretized, the subsample regression method can lose power. On the other hand, coarsely defining groups can let important information on treatment effect heterogeneity be lost (Hsu and Shen, 2019). We maximize this trade-off by splitting the sample into three or four groups depending on the source of variation. Specifically, in the case of continuous variables, we group observations in quantiles according to the median of the variable of interest (e.g., the reserve price or expected duration) and define four cross-subgroups (i.e., below reserve price median *and* above expected duration median). In the case of categorical variables, we typically assign groups based on three meaningfully selected elicited categories (e.g., government layers). We separately estimate the original regressions for these new subsamples across the three time leads and assess whether the estimated effects differ from one another and vis-à-vis the baseline's.

As reported in Table 3.E.1, we pin down heterogeneous effects on contract size *and* duration (Panel A), buyer type (Panel B), and construction category (Panel C). Regardless of the reserve price, the survival boost of awards is significant and stronger only for long contracts, while it is not significant for short small contracts and short large contracts. In other words, winners of contracts that are long and small (i.e., above the median expected duration and below the median reserve price) or long and large (i.e., above the median expected duration and above the median reserve price) have a survival advantage over auction losers. On the one hand, these results suggest that a firm survives because the awarded contract is active. However, we are not overly concerned about the estimated effect being mechanically driven by contract duration because i) we observe in our data that firms are awarded contracts quite regularly, and a pure mechanical effect would entail that they never exit the market, which is not the case; ii) the estimated survival effects after three years still far exceed the median contract duration *within* any 'long duration' subsample (i.e., 422 and 590 days for small and large long contracts, respectively); and iii) no legal constraint forces a firm to postpone declaring bankruptcy and exiting the market during the execution of a public contract until its end.

Most important for our work, concerns about mechanical effects are mitigated by the fact that contract size does not seem to matter. Hence, regardless of the size of the award, winners use contracts as a source of secured income to marginally improve their credit position—and thus indirectly their survival prospects. For example, in the event of a symmetric shock at the industry level, procurement firms could use their earnings-based collateral to access credit more easily and be more likely to survive. This would happen regardless of the income size and only because of the (public and therefore secured) nature of it.

Interestingly, contracts auctioned off by local buyers, irrespective of being on behalf of local government (i.e., region or municipalities) or the central government (e.g., universities) impact winner's survival; instead, contracts awarded by central administrations (e.g., ministries) are associated with an effect dissipating after the second year, despite awarding, on average, larger contracts. This suggests that 'geographical proximity' plays a role in the survival effect, as long as comparable contracts have longer-lasting (even though weaker) effects when awarded by local authorities. With the idea that firms are more likely to have political connections to local rather than central authorities and that political connectedness helps firms remove certain market frictions—importantly for our effect mechanism, strong evidence is available concerning credit access and financing (Johnson and Mitton, 2003; Cull and Xu, 2005; Khwaja and Mian, 2005; Leuz and Oberholzer-Gee, 2006)—this would reconcile such effect heterogeneity with Akcigit, Baslandze, and Lotti (2023) results that firms with political connections in Italy survive longer in the market also thorough relaxed credit constraints.[42]

Finally, we divide our sample of construction contracts according to the Common Procurement Vocabulary (CPV), which is adopted in Italy as well as other EU member states. In particular, we group contracts in Civil Works (i.e., CPV 452), Buildings (i.e., CPV 454), and Other Constructions. In this case, we signal the lack of effect for buildings. This finding can be motivated by the different weights of public customers for the total turnover in the civil work industry and building industry in Italy. In fact, the average share of public versus private spending in the public works construction market is higher than in the construction of buildings (i.e., 27 versus 23% respectively). The award (lack) of public contracts in the former case likely benefits (damages) firm business more than in the latter, as buildings companies have additional sale opportunities in the private building market and might replace a missed public with a new private customer more easily. Winners of public building contracts, therefore, tend to display no differential survival prospects compared to auction losers.

We have explored further subgroups analyses along other dimensions. The results are displayed in Table 3.E.2. For instance, we split the sample in terciles of contract reserve price distribution relative to winner's revenues (Panel A) and the same relative to employment (Panel B) to normalize contract size to suppliers size. In addition, in Panel C, we associate the contracts to the geographical area of the buyer to capture possible unobserved drivers related to local institutions and divided the country in northern regions (pooling NUTS1 Northwest Italy and NUTS1 Northeast Italy), central regions (NUTS 1 Central Italy, which includes the capital city of Rome), and southern and islander regions (NUTS1 South Italy plus NUTS1 Insular Italy) according to the subdivision of the country in adopted by the National Statistic Office and European Statistical Office. We find no detected heterogeneous effect along all these dimensions.

[42] Although we lack hard data to test the connection channel hypothesis, we can rule out two possible alternative explanations for this effect heterogeneity. First, a lack of power does not seem to explain the null effects of central government contracts. Indeed, the category Others in Panel C includes 201 observations (versus 317 for central government) but still shows a strong and significant effect. Second, the median length of contracts awarded by local agencies is only 80 days longer than those awarded by central agencies. This rules out the possibility that longer contracts explain the survival boost of local contracts.

Table 3.E.1. RD Regressions– Heterogeneity Analyses 1

	Survival		
	m+12	m+24	m+36
Panel A: Contract Duration and Size			
Short and Small	0.007	-0.004	0.008
	(0.005)	(0.015)	(0.020)
	0.994	0.981	0.961
	737	*737*	*737*
Short and Large	0.006	0.012	-0.019
	(0.006)	(0.009)	(0.041)
	0.994	0.984	0.968
	358	*358*	*358*
Long and Small	-0.000	0.012	0.053
	(0.000)	(0.009)	(0.019)
	0.995	0.984	0.968
	389	*389*	*389*
Long and Large	0.007	0.026	0.051
	(0.017)	(0.019)	(0.027)
	0.985	0.965	0.938
	670	*670*	*670*
Panel B: Buyer Type			
Central Government	0.027	0.042	-0.006
	(0.014)	(0.017)	(0.044)
	0.991	0.980	0.957
	317	*317*	*317*
Local Government	0.005	0.012	0.034
	(0.006)	(0.010)	(0.013)
	0.992	0.977	0.957
	1,677	*1,677*	*1,677*
Other Local	0.009	0.032	0.050
	(0.006)	(0.013)	(0.026)
	0.990	0.976	0.958
	523	*523*	*523*
Panel C: Construction Type (CPV)			
Civil Works	0.005	0.018	0.038
	(0.007)	(0.011)	(0.013)
	0.992	0.976	0.955
	1,820	*1,820*	*1,820*
Buildings	0.014	0.014	-0.004
	(0.008)	(0.008)	(0.024)
	0.992	0.986	0.969
	498	*498*	*498*
Others	0.023	0.033	0.112
	(0.016)	(0.020)	(0.035)
	0.989	0.971	0.945
	201	*201*	*201*

Notes: Subsamples of cross-terciles of reserve price (Panel A) and expected duration distributions (B) and different types of buyer (C). Contract size is defined based on auction reserve price, while contract duration on expected duration. We replicate in each subsample Table 3.3.

Table 3.E.2. RD Regressions–Heterogeneity Analyses 2

	Survival		
	m+12	m+24	m+36
Panel A: Reserve Price over Revenues			
Lower Tercile	0.009	-0.001	0.017
	(0.005)	(0.021)	(0.027)
	0.992	0.974	0.949
	661	*661*	*661*
Middle Tercile	-0.012	0.014	0.023
	(0.012)	(0.016)	(0.024)
	0.992	0.975	0.955
	635	*635*	*635*
Upper Tercile	0.015	0.022	0.010
	(0.007)	(0.009)	(0.016)
	0.991	0.982	0.968
	667	*667*	*667*
Panel B: Reserve Price over # Workers			
Lower Tercile	0.006	0.001	0.024
	(0.004)	(0.018)	(0.023)
	0.994	0.978	0.954
	645	*645*	*645*
Middle Tercile	-0.002	0.021	0.026
	(0.011)	(0.015)	(0.022)
	0.992	0.977	0.958
	661	*661*	*661*
Upper Tercile	0.016	0.023	0.005
	(0.008)	(0.009)	(0.019)
	0.989	0.977	0.961
	657	*657*	*657*
Panel C: North/Center/South&Islands			
North	-0.005	-0.008	-0.020
	(0.015)	(0.023)	(0.030)
	0.982	0.953	0.911
	1,035	*1,035*	*1,035*
Center	0.020	0.036	0.013
	(0.010)	(0.013)	(0.033)
	0.989	0.973	0.947
	453	*453*	*453*
South and Islands	0.003	0.004	0.014
	(0.002)	(0.006)	(0.011)
	0.994	0.985	0.971
	1,369	*1,369*	*1,369*

Notes: We split the sample of auctions depending on the distribution tercile of reserve price over firm revenues (Panel A), reserve price over firm employment (Panel B), geographical areas (Panel C). We replicate in each subsample Table 3.3.

Declaration

This dissertation is the result of my own work, and no other sources or means, except the ones listed, have been employed.

Matilde Cappelletti

Curriculum Vitae

Education

2020-2025	University of Mannheim Ph.D. in Economics
2017-2019	University of Mannheim M.Sc. in Economics
2013-2016	Free University of Bozen-Bolzano Laurea in Economics and Social Sciences

Work Experience

Since 2019	ZEW - Leibniz Centre for European Economic Research Department of Corporate Taxation and Public Finance Junior Research Group “Public Procurement" Researcher