Can Audits Backfire?

Evidence from Public Procurement in Chile*

Maria Paula Gerardino† Stephan Litschig‡ Dina Pomeranz§

October 2017

Abstract

Audits are generally intended to monitor compliance with existing rules. However, audits can also create unintended effects and incentives through the specific protocol by which they are executed. In particular, audits can discourage the use of complex administrative procedures with more rules for auditors to check. This paper investigates the effects of procurement audits on public entities' choice of purchase procedures in Chile. While the national procurement legislation tries to promote the use of more transparent and competitive auctions rather than discretionary direct contracts for selection of suppliers, auctions are significantly more complex and the audit protocol mechanically leads to more scrutiny and a higher probability of further investigation for auctions than for direct contracts. Using a regression discontinuity design based on a scoring rule of the National Comptroller Agency, we find that audits lead to a decrease in the use of auctions and a corresponding increase in the use of direct contracts. In order to further test the underlying mechanism, we develop a new approach to conduct subgroup analysis in regression discontinuity designs while holding other observables constant.

*We thank Alberto Abadie, Yoichi Arai, Abhijit Banerjee, David Cutler, Eduardo Engel, Ben Faber, Lawrence Katz, Gary King, Wojciech Kopczuk, Juliana Londono Velez, Luke Miller, Mushfiq Mobarak, Paul Novosad, Ben Olken, Zhuan Pei, Stéphane Straub and participants from conferences and seminars for helpful comments and discussions. We are grateful for excellent collaboration to the public procurement agency ChileCompra and the national comptroller Contraloría General de la República and for generous support to the Spanish Ministry of Economy and Competitiveness through the Severo Ochoa Programme for Centres of Excellence in R&D (SEV-2011-0075), the Development Bank of Latin America (CAF) and the Abdul Latif Jameel Poverty Action Lab's Governance Initiative.

†Inter-American Development Bank, mariage@iadb.org.

‡National Graduate Institute for Policy Studies, s-litschig@grips.ac.jp.

§University of Zurich and NBER, dina.pomeranz@econ.uzh.ch.

1 Introduction

Audits are pervasive in government, corporations and other institutions. The economics literature usually considers them as neutral information extraction tools to monitor rule compliance (Becker, 1968). In practice however, the details of how the audits are designed and implemented may play an important role, and the incentives they create may be more complex than a simple probability of detection. This paper looks into the black box of the audit process and investigates the effects of audits on public procurement practices in Chile. We find that the audit design can create incentives that go against the goals of the national public procurement legislation that the audits are intended to enforce.

In particular, since audits typically verify rule compliance, more complex processes with more steps tend to lead to more checks during an audit. If agents run the risk of making a mistake in any given step of the process, procedures involving more steps will mechanically lead to a higher probability of being found to be incompliant. Unless penalties for infractions are lower for processes involving more steps, audits will lead to higher expected costs of using more complex processes. This distortionary incentive can go counter to underlying policy goals, unless the regulator actually intends to discourage the more complex process. In many cases however, the reverse is true.

In the case of public procurement, many governments promote the use of public auctions in place of less transparent and less competitive direct contracting. The emphasis on auctions is consistent with both theory and evidence suggesting that auctions tend to increase competitiveness of procurement (e.g. Bulow and Klemperer, 1996; Lalive et al., 2015; Litschig and Zamboni, 2016). However, auctions are also much more complex and involve many more steps and rules than direct contracting procedures.

We empirically analyze these issues in the context of public procurement contracting in Chile. Public procurement represents an important share of the economy (about 12.1% of GDP and 29% of total general government expenditures in the OECD (OECD, 2015)), and the government is the largest buyer in many countries. Free and fair competition for government contracts plays a key role both for the quality and cost of government purchases, and to create a level-playing field for new entrepreneurs and suppliers, who can benefit from such contracts (Ferraz et al., 2016). Most governments therefore monitor compliance with public procurement regulation through external auditing by an independent comptroller agency.

Using a fuzzy regression discontinuity design (RDD), we find that external audits of public procurement in 2011-2012 lead to a sizeable decrease in the use of auctions and a corresponding increase of direct contracting - a result which surprised representatives of the auditing agency and goes against the overall goals of the procurement regulation. The shift leads to a sizeable reduction in competitiveness measured by the number of competing suppliers. There is a particularly strong shift away from auctions with more than 3 bidders and towards the type of direct contracts that only require one quote. Given that officers may underuse auctions, the national procurement legislation requires an explicit justification for the use of direct contracting. The justification that is known to be particularly flexible is "emergency".[1] We find that the strongest increase in direct contracting is through this emergency justification, accounting for almost half of the overall effect.

In order to study the specifics of the auditing process and understand what mechanism might lead to this result, we worked with the Comptroller Agency to conduct additional audits. The audit protocols were structured in the same way as any standard procurement audit, except that more information was collected in the process. While auditors in Chile typically only record detected infractions, in these additional audits the auditors also recorded information on checks that were conducted where no infractions were found, as well as information about which purchases were audited.

Results from these additional audits reveal that contracts made through auctions undergo about 2.5 times as many checks and lead to twice as many detected infractions than contracts made through direct contracting. These figures remain essentially unchanged when controlling for contract characteristics such as the amount and type of purchase. Consistent with the hypothesis that this is driven by the difference in the contract awarding modality, these differences arise almost exclusively from aspects related to the contract creation stage, rather from aspects related to the execution of contracts. We also investigate the alternative hypothesis that after having been audited, entities become less adherent to procurement rules because of low likelihood of being audited twice in a row. If anything however, there is an increase in the audit probability in the year following an audit. And even if there had been a temporary reduction in the subsequent audit probability, current contracts could still be audited during audits several years later.

We complement these findings with a subgroup analysis to shed further light on the

[1] For this reason, "emergency" is the only justification for which the regulation includes a personal liability by the head of the entity in case of overuse (Procurement Law art.8, letter c).

mechanisms. For this, we develop a novel empirical approach to overcome a well-known challenge in regression discontinuity designs: how to conduct valid subgroup analyses holding other observables constant. Using propensity score weighting in the spirit of Abadie (2005), we weight the sample to make the subgroups similar to each other in terms of other covariates. Analyzing the differential treatment effect in the weighted sample helps isolating the difference due to the subgroup characteristic of interest from other observable dimensions.[2]

One subgroup of interest in our case are entities with a high share of direct contract purchases in the pre-audit period vs. those with a low share. Entities with a high share of direct contracts may be more likely to learn that abuse of direct contracting is rarely detected or punished and therefore to respond more strongly. Weighting on the propensity score essentially eliminates the originally large covariate imbalances across these subgroups. While the estimated impact of the audit is indeed larger for entities with a high pre-treatment share of direct-contracts compared to those with below-median use of direct contracts, this difference is not statistically significant.

In order to conduct these analyses, the paper combines detailed administrative data on public purchases from the Chilean public procurement agency ("ChileCompra") with both historical and newly collected auditing data of the Comptroller Agency. ChileCompra, provided information on all procurement processes conducted on their platform. The Comptroller Agency ("Contraloría General de la República") provided data on audits as well as on a scoring rule and scores used in the selection of entities to be audited during 2011-2012, which allow us to implement the RDD.

Our paper contributes to multiple strands of literature. First, it presents to the best of our knowledge the first causally identified analysis of the impact of an audit on subsequent procurement behavior. The most closely related paper by Ferraz et al. (2017) finds that audits reduce future corruption among Brazilian local governments. In the case of taxes, the evidence on impacts of audits is mixed. DeBacker et al. (2015a) find that U.S. firms reduce tax payments following an audit, while DeBacker et al. (2015b) and Kleven et al. (2011) find an increase in tax payments for the case of the individual income tax in the U.S. and Denmark respectively. Our results also complement previous studies that analyze an increase in audit *risk* on corruption in road construction (Olken, 2007) and public procurement (Litschig and Zamboni, 2016), tax evasion (e.g. Kleven et al.,

[2]We developed a Stata command that implements this approach, rddsga, which is available for download in the Stata repository.

2011; Pomeranz, 2015) or compliance with environmental regulation (Duflo et al., 2014).

Second, our paper highlights that when governments scrutinize certain processes more carefully, they may inadvertently discourage the use of these processes altogether. These findings are consistent with a number of papers showing that in the face of increased scrutiny, agents substitute activities to less well monitored margins (e.g. Olken, 2007; Carrillo et al., 2017; Niehaus and Sukhtankar, 2013; Yang, 2008). Our results indicate that such substitution may not only happen if certain activities are explicitly subject to higher audit probabilities than others, but can also result from differences in the specific audit protocol applied to different activities.

Third, our paper contributes to a rapidly growing literature on public procurement design by providing new insights into the factors that can affect the choice of procurement modality. The finding that the shift towards direct contracting reduces the competitiveness in terms of the number of involved suppliers is consistent with Lalive et al. (2015) who find that auctions lead to lower prices and higher quality than direct contracting in the case of rail services in Germany. Banerjee et al. (2015) find that increased competition through additional suppliers reduced costs without sacrificing quality of local public rice procurement in Indonesia. The way in which procurement contracts are awarded has also been found to impact quality-adjusted prices by Bandiera et al. (2009) for the case of framework agreements and by Lewis-Faupel et al. (2016) for the case of e-procurement. These results also relate to recent empirical work that investigates how aspects of auction design affect efficiency in public procurement (e.g. Tran, 2010; Decarolis, 2014; Coviello and Mariniello, 2014; Coviello et al., 2015).

Finally, our finding that the role of audits goes beyond a simple information extraction tool and that it is key to think carefully about (unintended) incentives of audit design are consistent with Duflo (2017) which stresses the importance of studying the details of policy implementation. Diving into the "plumbing" details of policy design often reveals a much more nuanced structure of incentives.

The remainder of the paper proceeds as follows. Section 2 provides institutional background on public procurement and on external audits of public entities in Chile. Section 3 describes the data. Section 4 discusses the empirical strategy. Section 5 shows the main results and Section 6 presents analyses of underlying mechanisms. Finally, Section 7 concludes.

2 Background

2.1 Public Procurement in Chile

ChileCompra is the government agency in charge of managing the public procurement system and the online platform on which most public procurement in Chile takes place.[3] Since its inception in 2003, the platform has grown to serve practically all public entities in Chile, with more than 100,000 firms supplying goods and services. During our study period, purchases conducted through the ChileCompra platform represented about 4% of GDP (ChileCompra, 2013).[4] Contraloría is the National Comptroller Agency in charge of monitoring all public entities, including ministries, municipalities, public services, and state-owned enterprises. Contraloría's primary monitoring activity consists of audits of different types of activities by public entities.

There are four main purchasing modalities in Chile, each with distinct implications for the extent of transparency and competitiveness. The two key modalities analyzed in this paper are auctions and direct contracting. Historically, almost all purchases were made through direct contracting, but over recent years ChileCompra has succeeded in moving a large share of purchases to be done through auctions. All else equal, public auctions are preferred by the regulation because of their higher transparency and competition compared to direct contracting. For such auctions, all information about the bidding process is publicly available online and selection criteria are specified explicitly and ex-ante by the purchasing entity. The mean number of bidders in auctions in our study sample is around 10 (see Table 1).

In order to use direct contracting, public entities need to provide a justification, explaining and documenting the need to do so. Possible justifications include cases in which only one supplier exists, emergencies, cases in which organizing an auction would represent a disproportionate cost, or cases where the total sum is below about 700 USD.[5] Depending on the type of justification for using direct contracting, procurement officers are required to get one or three quotes from suppliers. About 60% of direct contracts in

[3]There are a few exceptions for transacting outside the platform, such as for purchases by the armed forces.

[4]Large public works such as the construction of an airport or highway are not included in the ChileCompra procurement system and are handled by a different agency.

[5]The specific threshold is 10 UTM (Unidad Tributaria Mensual, an inflation-adjusted Chilean unit of account). In December 2011, 10 UTM were approximately 750 USD.

our study sample require only 1 quote, while 40% require 3 quotes, leading to an average number of quotes of 1.8 for direct contracts.

For products that are used by many public entities, such as office supplies, ChileCompra established framework agreements, in which entities can simply order products for a pre-established price from a website, called the "procurement supermarket". When a product is available in the "supermarket", the entity does not need to resort to the organization of an auction nor to direct contracting. In our study period, about 16% of the value of purchases was made through framework agreements. Finally, small purchases below about 200 USD can be made outside of the electronic procurement system altogether (about 0.4% of the total value of purchases).

2.2 Audit Selection Process

During our 2011-2012 study period, the Comptroller Agency selected entities for the most common type of audit using a scoring system.[6] In this system, the audit probability depends on an entity's "relative importance score". This score is calculated based on measures such as the entity's budget size, transfers to the private sector, etc. Public entities are ranked according to their relative importance score. Within each internal control department of the Comptroller Agency, entities are then divided into three groups: high, medium, low relative importance. The classification is done separately each year, so an entity can be in different cells in different years. Content of the factors going into the scoring rule, as well as even the existence of such a score were not publicly known and maintained secret within a small team at the central control office. Cutoffs are neither known ex-ante nor published ex-post, making manipulation around the cutoffs by entities virtually impossible.

For entities that the Comptroller Agency considers to be of medium risk for non-compliance,[7] the audit probability depends on their relative importance classification. Within the group of medium risk entities, those with high relative importance have a high audit probability, those with low relative importance have a low audit probability, and for those with medium relative importance it depends on the available auditing resources. (For entities classified as high risk, the audit probability is generally high, and for entities

[6] These audits involve a general examination of different areas of the entity's operations, which includes a strong focus on the entity's procurements.

[7] The risk classification includes elements such as low compliance in previous audits, complaints from civil society and government priorities. It also varies from year to year and is not publicly known.

classified as low risk, it is generally low, independent of the relative importance score). We can therefore use the cutoffs between high and medium, and between medium and low relative importance score for an RD design among the entities generally classified as medium risk. There are 720 and 953 medium risk entities in 2011 and 2012 respectively, which represents 59% and 63% of the total number of entities.

The scores and cutoffs are determined separately within 91 strata, which are defined by year, responsible control department ("UCE") and type of entity (national, regional or municipal). The reason for this is that auditors and corresponding auditing manpower are predetermined at the stratum level. Within each stratum, the range of values for each score is divided into three equally-sized parts that determine the cutoffs. For example, if the score ranges from 20 to 80, the cutoff between low and medium would be at 40 and the cutoff between medium and high would be at 60.

In addition to the relative importance score, other more subjective factors are involved in the determination of which entities get audited in a specific year. So there is not a one-to-one determination from an entity's classification to whether it is audited. However, there is a significant discontinuity in the probability of being audited around the cutoff. Small differences in the relative importance score can affect the entity's probability of being audited. This allows us to use fuzzy regression discontinuity analysis. In our data, there are two sizeable and significant discontinuities in the audit probability. In 2011, the discontinuity in the probability of audit occurs between low and medium levels of relative importance, while in 2012, it occurs between medium and high importance. We therefore focus our analysis on entities classified as medium risk and in the vicinity of the relevant discontinuity for each year.[8]

3 Data

We combine administrative data from the procurement agency ChileCompra's online procurement platform with audit data from the Comptroller Agency.[9] To complement this data, we worked with the Comptroller Agency to conduct additional audits to investigate what happens during procurement audits and shed light on the mechanisms underlying the shift from the use of auctions towards direct contracting.

[8]Including additional smaller discontinuities would lead to weak instrument problems (Feir et al., 2016).

[9]Our online data Appendix explains the construction of the dataset in detail.

3.1 Data from the Procurement Agency ChileCompra

The ChileCompra data include information on all purchase orders conducted on the online platform. Each auction or direct contract results in one or several purchase orders. For each purchase order, we have data on the purchasing entity, the purchase modality (auction, direct contract, or online supermarket), date of the purchase, product code of each item in the order (up to 8-digit codes), verbal description of each item, name of the seller and amount of the purchase.

Table 1 Panel A presents summary statistics for the universe of purchase orders issued by medium risk public entities during our estimation period of 2011-2012. Auctions make up about 51% of purchase orders and 66% of dollars spent. Direct contracts make up around 14% of purchase orders and 17% of dollars spent. Framework agreements represent almost 28% of purchase orders, but only 16% of total dollars spent. This is because framework agreements are most commonly used for relatively low-cost purchases, such as office supplies and cleaning materials. About half a percent of the value of purchases conducted through the platform was in the category of small purchases, for which use of the electronic procurement platform is optional. When we focus on public entities in our estimation sample (i.e., those relatively close to the cutoff of the regression discontinuity design) the numbers are quite similar (see Panel B).

In addition to the ChileCompra purchase order data, we also analyze a separate auctions database with information on the number of bidders in each auction. We use this information to analyze the share of purchases originating from auctions with high or low numbers of bidders.

3.2 Data from the Comptroller Agency Contraloría

Data on Entities and RDD Audits

The Comptroller Agency provided data both on characteristics of the public entities in the study, the audits, and the relative importance score used for the RDD. Information about the entities includes the following: the entity type (national, regional or municipal entity), the responsible control department in charge of auditing a given entity,[10] the score

[10]In 2011-2012 there were 23 control departments (*Unidades de Control Externo*, UCE): one for each of the fourteen non-metropolitan regions of Chile, one for municipalities and eight related to specific ministries.

and cutoff of the relative importance indicator and the level of the risk classification. We also construct a variable of political affiliation of each entity based on public information from the Chilean Electoral Service (Chilean Electoral Service, 2014) (right wing coalition, left wing coalition, independent).[11]

Data on the audits include which entities were audited and the audits' start and end date. In the estimation sample during 2011-2012, 260 entities were audited out of a total of 1002.

Additional Audits

In order to analyze what happens in detail during an audit, the Comptroller Agency agreed to undertake a number of additional audits. These audits were conducted in the same way as typical public procurement audits, with the key difference that auditors recorded more information than usual. Usually, auditors only report detected infractions, but not which purchases were audited and which types of checks were conducted. In these additional audits, auditors recorded which contracts were audited and which checks were being conducted, in addition to any detected infractions. This information was recorded at the purchase level. In regular audits, findings from several purchases are usually grouped together, so that it is not possible to study audit results by procurement modality.

The additional information recorded in these audits allows us to examine differences in the way auctions and direct contracts are audited, and to compare the frequency of detected infractions by purchase modality. For each purchase, we are able to see how many checks were carried out, and in how many of these checks infractions were found. In addition, auditors record whether the seriousness and type of the detected infraction warrant a follow-up visit (to verify whether the entity has introduced corrective measures) or an investigation (which can lead to punitive legal action).

While the audits examined in the RDD analysis happened 3-4 years earlier than these additional audits, there had been no substantive changes in the way the Comptroller Agency executes such audits. The audits analyzed with the RDD happened in 2011-2012 and the additional audits were implemented in 2015 (see Figure 1 for a timeline). The additional audits took place in two waves, half of them starting in July and the other

[11] National and regional entities are assigned to the right-wing coalition since they are part of the right-wing coalition government that was in office at the national level at that time. The affiliation of municipal entities is assigned according to the affiliation of the mayor at the time.

half in September. 18 entities were selected randomly within strata that had remaining auditing capacity. In each entity, Contraloria audited three purchases of goods and three purchases of services, for a total of 108. Purchases were selected mainly based on their value in order to justify the cost of auditing them. Similar to the study sample described in Table 1, here too about a third of the audited purchases originated from direct contracts, while the other two thirds stemmed from auctions.

The audit protocol involved 95 different checks, most of which referred to aspects of either the contract creation or the contract execution stage.[12] The contract creation stage includes everything that precedes contract signing, such as choosing the procurement modality, writing the specifications for auctions, requesting quotes for direct contracts, evaluating the bids or offers, etc. The contract execution stage refers to all activities following contract awarding, such as the timing of delivery, quality of the product or service, and whether contract specifications are met, etc.[13] This allows us to analyze the audit results both by type of purchase modality and by type of check.

4 Empirical Strategy

4.1 Identification Strategy for the RDD

Based on the audit-selection process described in 2.2, we can use a fuzzy regression discontinuity design to estimate the impact of the audits. As described above, within the group of medium risk entities, the audit probability is affected by which category of relative importance the entity falls in. Relative importance is determined by a continuous score with clear cutoffs. This allows us to use an RDD by comparing entities directly above and directly below the thresholds separating high from medium or medium from low relative importance. The intuition behind this RDD is that those directly below and above a relative importance threshold have practically the same relative importance. At the same time, the probability of being audited jumps discontinuously.

As discussed above, the cutoffs are determined separately in each stratum. In order

[12]These audits were also used to pilot an intervention in which the Comptroller Agency varied what letters they sent to the entities to inform them that they would be audited on the contract creation or creation and execution stages. The audit protocol was independent of the letter, and our analysis in this paper exploits variation within entities rather than across.

[13]Our online appendix shows the audit protocol with its 95 checks, and their classification according to whether they relate to the creation or the execution stage.

to be able to pool the data for the analysis across the different strata, we look at the normalized score of each entity, following the approach of prior literature (e.g. Hastings et al., 2014; Kaufmann et al., 2013; Pop-Eleches and Urquiola, 2013). This involves setting the cutoff for each stratum to zero, so that the normalized score represents the distance from the cutoff.

Formally, let Y_{ij} denote an outcome for public entity i in stratum j; D_{ij} the indicator for being audited; X_{ij} the relative importance score normalized with respect to cutoff c_j, $\mathrm{I}[X_{ij} \geq 0]$ an indicator for an importance score above the cutoff in stratum j; τ the effect of an audit; π the effect of crossing the cutoff on the audit probability; $f(X_{ij})$ and $g(X_{ij})$ polynomials in the importance score. Additionally, S_j is a full set of stratum dummies, and W_{ij} is a vector of entity characteristics, including past outcomes. Finally, U_{ij} and V_{ij} capture the error terms.

The model is then as follows:

$$Y_{ij} = \tau D_{ij} + f(X_{ij}) + S_j + \gamma W_{ij} + U_{ij} \tag{1}$$

$$D_{ij} = \pi \mathrm{I}[X_{ij} \geq 0] + g(X_{ij}) + S_j + \theta W_{ij} + V_{ij} \tag{2}$$

The IV estimator using $\mathrm{I}[X_{ij} \geq 0]$ as an instrument for D_{ij} identifies the effect of an audit τ under two main assumptions. The first is the continuity of $\mathrm{E}[U_{ij}|X_{ij}]$. Intuitively, this requires that there are no other unobserved factors that change discontinuously at the relevant cutoffs. As shown in Lee and Lemieux (2010), a sufficient condition for continuity of unobservables is that the density of the variable determining treatment assignment is continuous.

In practice, this is fulfilled if there is no precise manipulation of entities to be on one side or the other of the cutoff. This is plausible, because public entities have at most imprecise control over their value of the relative importance score. Both the fact that the control department of the Comptroller Agency makes such a score and the details of how it is calculated are unknown to the public entities. In addition, the cutoffs are determined after the indicators comprising the relative importance index have been calculated, so nobody knows where the cutoffs are going to be at the time the indicators are calculated. While this assumption is not directly testable, it has testable implications, which we examine using the McCrary test (McCrary, 2008) in Section 5.1 below.

The second assumption is the exclusion restriction, i.e., the assumption that crossing the cutoff affects outcomes only through the increased audit probability, not through other channels. In our case, it is very unlikely that any other changes are happening at that threshold, given that this is an internal score of the central control department, not shared with any other areas of the Comptroller Agency, and different for every stratum and in every year.

4.2 Estimation Approach for the RDD

During our study period from 2011-2012, the normalized distance to the cutoff ranges from -62.5 to 38.9. However, observations that are very far from the cutoff are not very informative to estimate the RD-gap at the cutoff. Following Hahn, Todd and Van der Klaauw (2001) and Imbens and Lemieux (2008), we estimate local linear regressions in samples around the cutoffs. In accordance with Lee and Lemieux (2010), we also use a quadratic specification in a larger bandwidth. Based on visual inspection, the linear and quadratic specifications use bandwidths of ± 4 and ± 10, respectively. We use OLS with a rectangular kernel, which in effect amounts to giving higher weight to observations closer to a given cutoff c_j. In addition, we present results from outcome-specific bandwidths that are Mean Square Error (MSE)-optimal as proposed by Imbens and Kalyanaraman (2012) as well as the bias-corrected estimates and robust standard errors proposed by Calonico, Cattaneo and Titiunik (2014). Throughout the paper, we focus on reduced form estimates from Equation (4) below in order to maintain a close correspondence with the graphical reduced form evidence.

In particular, our linear specification for observations within a distance h of the cutoff between levels of relative importance within a given stratum is as follows:

$$D_{ij} = \pi \mathrm{I}[X_{ij} \geq 0] + \rho_0 + \rho_1 X_{ij} + \rho_2 X_{ij} \times \mathrm{I}[X_{ij} \geq 0] + S_j + \gamma W_{ij} + V_{ij} \tag{3}$$

$$Y_{ij} = \delta \mathrm{I}[X_{ij} \geq 0] + \beta_0 + \beta_1 X_{ij} + \beta_2 X_{ij} \times \mathrm{I}[X_{ij} \geq 0] + S_j + \theta W_{ij} + U_{ij} \tag{4}$$

where $\delta = \tau \times \pi$, X_{ij}, D_{ij}, Y_{ij}, S_j, W_{ij}, V_{ij}, and U_{ij} are as in Section 4.1.

We present specifications both with and without control variables. The control variables (W_{ij}) included are: 1) the shares of purchases the entity made through auctions and through direct contracting one and two years prior to the treatment year; 2) the log (+1) of the total amount purchased by the entity in dollars one and two years prior to the treatment year; 3) a dummy for being audited one year prior to the treatment;[14] and 4) dummies for the entities' political affiliation. In order to account for potential common shocks, we cluster the analysis at the stratum level.

4.3 New Approach to Subgroup Analysis in RDD

In order to investigate which hypotheses are supported by data, researchers often apply subgroup analyses to study the differential treatment effect of a certain type in the population. However, since subgroups may differ from each other on various dimensions, researchers would often like to compare two subgroups while holding other characteristics constant.

As an example, one subgroup of interest in our case are entities with a high share of direct contract purchases in the pre-audit period (let's call that group G_i=1) vs. those with a low share (G_i=0). The reasoning is that if procurement officers learn during the audit that abuse of direct contracting justifications is not detected or punished, one might expect entities with more direct contracts to be more likely to have the opportunity to learn this and therefore to respond more strongly to the audit. However, entities with a high share of direct contracts differ from those with a low share in other ways. They tend to be larger, more likely to be national as opposed to municipal and more likely to have been previously audited. The goal is to find a way to analyze the different responses to the audit by entities with a high vs low pre-treatment auction share, while holding such other characteristics constant.

The standard approach to investigate systematic treatment effect heterogeneity in many settings is to use specifications that include the subgroup of interest indicator and an interaction term of the treatment indicator with that subgroup indicator. This approach is very convenient because it easily accommodates additional covariates and their interactions with treatment status, thus holding other observables constant when testing for differential effects by subgroup.

However, as we discuss in Appendix B and show empirically in Tables A7 and A8, the simple treatment-subgroup-interaction approach is not generally valid in the RDD setting, unless the relationship between the outcome and the running variable is the same across subgroups. While this specification bias can be addressed by allowing for separate slopes

[14] Audit data is only available for one year prior to 2011.

by subgroup (i.e. full interaction of the running variable polynomial with the subgroup indicator), the problem remains that other characteristics may vary systematically across subgroups, thus making it difficult to interpret differential subgroup impacts. Simply including these additional covariates and their interaction with the treatment indicator would lead to the same specification errors discussed above. Another approach is to run separate RDDs in cells defined by specific covariate combinations, but this runs into weak instrument and sample size issues (Feir et al., 2016). An intermediary approach is to include interactions of the covariates with running variable polynomials, but this has not been formally investigated in the RDD setting as far as we know.[15] In practice, RDD papers that conduct subgroup analyses typically run separate RDDs for given values of the covariates (Hsu and Shen, 2016) and implicitly or explicitly assume that close to the cutoff other determinants of outcomes are not correlated with the subgroup of interest (Becker et al. (2013), assumption 3).

In what follows, we propose a new approach to analyze subgroup effects in RDDs without having to make additional assumptions . Section 6.3 shows an empirical application of this approach with the example mentioned above (comparing entities with high vs. low pre-treatment auction shares). We also developed a Stata command that implements a weighted binary subgroup analysis for RDDs, rddsga, which is available for download in the Stata repository.

Propensity Score Weighting

In order to estimate the differential impact on group $G_i = 1$ vs $G_i = 0$ in an RDD setting while holding other observable characteristics constant, we propose an approach based on propensity score weighting in the spirit of Abadie (2005). This involves weighting observations from each subgroup by the inverse of their conditional probabilities to belong to that subgroup given a set of covariates. Thus, observations from subgroup $G_i = 1$ with high estimated propensity scores and associated covariate characteristics will be down-weighted, while those with low estimated scores will be up-weighted, making the covariates of the weighted $G_i = 1$ sample more closely resemble the similarly weighted $G_i = 0$ subgroup sample. Running the RDD analysis separately within each weighted

[15] Calonico et al. (2017) aims to clarify the conditions under which adding covariates in the RDD identifies the average effect of treatment at the cutoff. An earlier version of their paper suggests that future work could study how the interaction approach might be used to investigate treatment effect heterogeneity at the cutoff.

subgroup eliminates potential confounding differences due to other observable factors that may vary systematically across (unweighted) subgroups. As is the case with any approach of controlling for observables, propensity score weighting cannot rule out any unobserved confounding factors.

To estimate the propensity score, we first restrict the sample to observations close to the cutoff using the same bandwidths as for the main results of the paper. We then follow the standard approach for propensity score weighting. Estimate a logit model in order to calculate a predicted probability to belong to subgroup G_i=1 (in our example the probability of having above-median share of purchases made by direct contracting):

$$P(G_i = 1|W_i) = \frac{e^{h(W_i)}}{1 + e^{h(W_i)}} = P(W_i), \tag{5}$$

where $h(W_i)$ is a starting specification that includes the covariates W_i as linear or interaction terms. Restrict the sample to the common propensity score support and weight observations by the inverse propensity score. Specifically, the weight attached to the i-th observation is

$$G_i \frac{p}{P(W_i)} + (1 - G_i) \frac{1 - p}{1 - P(W_i)}, \tag{6}$$

where p is the unconditional probability of belonging to subgroup $G_i = 1$.

Assessing Covariate Imbalance Reduction

After weighting, we can check whether this process removed imbalances in covariates between the two subgroups of interest by comparing mean differences in the unweighted and weighted samples. To assess the statistical significance, we use a t-test for individual coefficients and an F-test for overall balance, as typically done in balance tables. To assess economic or substantive balance we use standardized mean differences (SMD). To assess average balance across covariates we take a simple average of the SMDs in absolute terms.

Estimating the RDD by Subgroup

Once we have settled on a propensity score specification that eliminates or strongly reduces the imbalances of observables between $G_i = 0$ and $G_i = 1$, we can proceed to estimate the differential treatment effect. One aspect to consider before doing so is that the first stage (i.e. the degree to which the treatment probability jumps at the cutoff) can be different for different subgroups. For this reason, it is important to compare the estimated effects of the actual treatment (in our case, the impact of being audited, i.e. the IV regressions using Equations (3) and (4)) rather than the reduced form estimates (i.e. the effect of passing the cutoff). Another caveat is that conventional (robust or clustered) standard errors do not capture sampling variability coming from the fact that the propensity score is estimated. To deal with this issue, our Stata command (block) bootstraps standard errors and confidence intervals.

4.4 Identification Strategy for the Additional Study Audits

In analyzing the additional audits, we run OLS regressions of the number and type of checks and infractions on whether a purchase was done through auction or direct contracting:

$$Y_{ic} = \beta_0 + \beta_1 Auction_{ic} + \beta_2 X_{ic} + \varepsilon_{ic} \tag{7}$$

where Y_{ic} is the outcome variable of contract c in entity i, $Auction_{ic}$ is a dummy that equals 1 when the purchase originates from an auction and 0 when it originates from a direct contract, and X_{ic} are a number of covariates including the type of product or service, the month of the purchase, the amount of the purchase, as well as the month of the audit and the control department responsible for the entity in question. ε_{ic} is the error term. Standard errors are clustered at the entity level.

In order to hone in more precisely on the mechanism, we distinguish differential effects on the creation stage of the contract, where auctions and direct contracting differ strongly, and the contract execution stage, where there is no reason to believe that, all else equal, contracts awarded through auctions or direct negations should differ.

5 Results

5.1 RDD Internal Validity Checks

As discussed in Section 4.1, certain assumptions for the validity of the RDD are testable. First, we conduct a McCrary density test (McCrary, 2008) to analyze whether there is bunching of the mass of entities on one side of the cutoff. Then we analyze whether entities are balanced with respect to the observable characteristics at the cutoff.

Figure 2 shows the results of a McCrary density test in our study sample, entities with medium level risk in 2011 or 2012 in the ±10 range around the discontinuity cutoff. The null hypothesis of the McCrary test is that the density of the treatment-determining variable - in our case the relative importance score - is smooth around the cutoff. The dashed line estimates the density on either side of the cutoff, while the solid lines provide a 95% confidence interval. There is no statistical evidence against the null hypothesis that the density is smooth around the cutoff. This is consistent with the assumption that there is no manipulation of entities around the cutoff value.

Local random assignment of entities just above vs. just below the cutoff can also be indirectly tested by examining whether pre-treatment covariates are balanced above vs. below the cutoff, as shown in Table 2. In this balance table, we test whether there is a discontinuity for any of the covariates, by running an RDD with the covariates as an outcome, as in Equation (4). Columns (1) and (4) show comparison means in the ±4 and ±10 range, respectively. Columns (2) and (3) show linear estimates in the ±4 range without and with controls. Columns (5) and (6) show quadratic estimates in the ±10 range without and with controls.

The pre-treatment covariates include the share of purchases made by auction, direct contracting and framework agreement and the log (+1) of the total amount purchased in the two years prior to treatment, a dummy for being audited in the year prior to treatment and the political affiliation of the entities. While the p-values of the F-tests for joint significance of all variables do not indicate significant discontinuities at the cutoff, in one specification there is a pre-treatment outcome which is significantly different from zero at the 10% level. For this reason, we include specifications with lagged outcomes as well as lagged values of the other covariates in Table 2 as controls. We also show a graphical representation over time that confirms that the treatment effect indeed starts at the time of the treatment.

5.2 RDD First Stage: Effects on the Audit Probability

We now turn to the analysis of whether there is indeed a discontinuity in the audit probability at the cutoff, i.e. whether there is a significant first stage. Figure 3 presents the first stage results for the pooled estimation sample for the discontinuity of moving from low to medium relative importance in 2011 and from medium to high relative importance in 2012.[16] The x-axis represents an entity's normalized relative importance score. In Panel A, the y-axis shows audit probabilities. In Panel B, the y-axis represents the residual audit probability after controlling for stratum fixed effects and the control variables. Each dot represents a two-point wide bin. For example, the first dot to the right of the cutoff combines all entities with a relative importance score between 0 and 2 points above the cutoff. Linear and quadratic fitted lines are also included. The graphical evidence suggests that the probability of being audited increases discontinuously at the threshold.

Table 3 presents these results numerically based on Equation (3). The specifications include a linear or quadratic spline in entities' distance to the cutoffs. Columns (1) to (3) use a bandwidth of ± 4 and a linear spline whereas columns (4) to (6) use a bandwidth of ± 10 and a quadratic spline with varying inclusion of covariates. Column (7) employs the optimal bandwidth given by the MSE criterion proposed by Imbens and Kalyanaraman (2012). Column (8) uses the same bandwidth as in column (7) but with the bias-corrected RD estimate and robust standard errors proposed by Calonico, Cattaneo and Titiunik (2014). In our preferred specifications, which include stratum fixed effects and control variables (Columns (3), (6), (7) and (8)), the probability of being audited increases above the cutoff by 15.8 to 19.3 percentage points. All first stage estimates are statistically significant at the 5% or 1% level.

5.3 Effects on Purchase Modalities

Having established that there is indeed a significant first stage, we now turn to the analysis of our main outcome: the impact on procurement modalities. Figure 4 presents graphical evidence for the impact on the use of auctions and direct contracting. It shows discontinuities in the shares of the amount bought through auctions and direct contracts. The x-axis represents the normalized distance from the cutoff in the relative importance score. In the two figures on the left-hand side, the y-axis represents the share bought through auctions (Panel A) and direct contracts (Panel C). On the right hand side, the y-axis represents residuals from a regression of the share of purchases made through auctions (Panel B) and direct contracts (Panel D) on the stratum fixed effects and control

[16] The separate results for years 2011 and 2012 respectively can be found in Table A1 and Figure A1 in the Appendix.

variables. The vertical distance between the dots close to the discontinuity represents the intent-to-treat (or reduced form) effect, i.e. the impact of passing the threshold, and therefore having a higher audit probability, on the outcomes.

The graphical evidence shows a discontinuous decrease at the cutoff in the share of purchases made through auctions and an increase in the share made through direct contracts. The similar magnitudes and opposite directions of the jumps suggest that entities increase the use of the direct contracts at the expense of auctions.[17] Figure A5 in the Appendix shows no change in the total amount of purchases.

Table 4 displays these results in regression form, following Equation (4). Column specifications are the same as in Table 3.[18] The results are quite robust across the different specifications, even though the level of significance varies. In the specifications including the control variables, the estimates of the reduction in the share bought through auctions range from 6.9 to 8.9 percentage points and the estimates for the increase in the share of direct contracts range from 6.1 to 7.7 percentage points.[19] The IV regression that combines the first stage with the reduced form shows an impact of being audited of 34 to 41 percentage points for the shift towards direct contracts and 38 to 48 percentage points for auctions. However, these specific point estimates have to be interpreted with caution both because the standard errors are large and because the first stage estimates may be too weak to provide reliable IV estimates due to weak instruments problems.

We also look at how the treatment effect develops over time. Appendix Figure A7 shows quarterly treatment effects for two years before to two years after the beginning of year t. While the quarterly results are relatively noisy, the effect size seems to increase over the course of year t and decrease over the course of year $t+1$. Table A6 displays regression results for year $t+1$. Although the estimates are again negative for auctions and positive for direct contracts, they are smaller than in the year of the audit and not statistically significant. This suggests that the impacts decrease over time. However, this

[17]Figure A2 in the Appendix shows no apparent discontinuities in the shares of purchases made through framework agreements or small purchases. Separate figures by year are shown in Figures A3 and A4.

[18]Table A2 in the Appendix shows the regressions for purchases done through framework agreements and small purchases. The effects are small and insignificant. Separate estimates for 2011 and 2012 are shown in Tables A3 and A4 respectively.

[19]As a robustness check, we calculate these effects including the interaction between stratum dummies and distance to the cutoff. This allows for the possibility that the relationship between the outcome and the running variable could be different within each stratum. Figure A6 and Table A5 show these estimates graphically and in regressions. The results are quite similar (8.3 to 9.5 percentage points for auctions and 5.6 to 7.2 for direct contracting).

interpretation has to be made with caution, because the confidence intervals are wide, and the effects are not statistically significantly different from year t.

Beyond analyzing whether entities use direct contracting or auctions, we can also analyze the impact on the type of direct contracting used. There are 21 different justifications or reasons for the use of direct contracting instead of auctions.[20] The most frequent ones are: unique supplier (only one supplier in the market), emergency, purchase is for less than $720, trust in a particular supplier only, and disproportionate cost of organizing an auction.

Out of these 21 justifications, the emergency justification is generally known as being prone to overuse and is for this reason the only one for which the regulation explicitly stipulates that the head of an entity is personally liable in case of overuse. In part, this is because it is difficult to monitor ex-post whether some purchases were indeed urgent and because if a buyer waits long enough, almost any purchase becomes justifiable on emergency grounds.

Table 5 shows RDD estimates for the shares of purchases made through direct contracts using each of the five most frequent justifications and through any of the other justifications grouped together. We see that the emergency justification accounts for the biggest increase in the share of total purchases made. This type of direct contract increased by 4.2 to 4.9 percentage points, from a base share of 0.8 to 1.4 percentage points significant at the 1% or 5% level. These results suggest that emergency purchases represent almost half of the aggregate increase in direct contract purchases caused by the audits. Purchases with unique supplier justifications start at a higher base share of 2.5 percentage points and increase by 1.2 to 1.8 percentage points (only marginally significant). The coefficients for the other four categories are close to zero and not statistically significant. Figure 5 presents graphical evidence of these results. Overall, these results suggest that the increase in direct contracting is driven to a large degree by the most flexible emergency justification.

5.4 Effect on Competitiveness

One reason that policy makers tend to promote auctions over direct contracting is that auctions tend to be more competitive in the sense that a larger number of suppliers

[20]The 21 justifications coming from Chilean law number 19,886 are listed in Appendix C.

compete for the contract. However, it is not a priori clear that a move from auctions towards direct contracting necessarily implies a decrease in the number of bidders. Some auctions attract only a small number of bidders, and certain types of direct contracts also require the procurement officer to obtain 3 quotes from different firms.[21]

In this section, we therefore examine whether the effects described above indeed lead to a reduction in the number of suppliers involved. In particular, we would like to rule out that the reduction in auctions comes mainly from auctions with a small number of bidders, while the increase in direct contracting comes mainly from cases requiring three quotes, in which case the number of suppliers involved may not actually change even though we see a reduction in the share of auctions.

Table 6 presents regression results measuring the impact separately on the share of auctions and direct contracts with high and low number of involved suppliers. Columns (1) and (2) show the impact on the share of auctions with 3 or fewer bidders and with more than 3 bidders. The difference between the coefficients of Columns (1) and (2) is not statistically significant, but if anything, the reduction is larger for auctions with a large number of bidders. Columns (3) and (4) show the impact on direct contracts that require 1 quote or 3 quotes. Again, the difference between the two is not statistically significant, but if anything, the effects appear larger for direct contracting that requires only one quote. Overall, these results show that the shift from auctions to direct contracting primarily stems from competitive auctions with more than 3 bidders to direct contracting types such as emergency, which require only one quote.

6 Mechanisms

This section presents evidence on potential mechanisms that might explain the observed increase in the use of direct contracting and corresponding decrease in the use of auctions. First, we discuss results from the additional audits that allow for contract-level analysis. Second, we check whether the audit probability decreases following an audit, which could have led entities to temporarily relax compliance with procurement regulation. Finally, we discuss the results of an application of our new approach to conduct subgroup analysis

[21] The most frequent justifications that require 1 quote are unique supplier, emergency, trust in suppliers and disproportionate cost. On other hand, the most common justifications that require 3 quotes are: unsuccessful auction, unfinished contract due to breach of contract from original supplier, contracting with foreign legal persons and cost of less than 720 USD.

in an RDD setting, comparing public entities with high vs. low pre-audit shares of direct contract purchases.

6.1 Auctions Receive More Scrutiny Than Direct Contracts

In order to measure the way in which auditors treat purchases through auctions and direct contracts, we use data from the additional audits, described in Section 3.2. These audits were implemented in the same way as standard audits, with the only difference that more detailed information about the process was collected.

Table 7 presents regression analysis of the number and type of checks and detected infractions by purchase on whether that purchase was done through auction or direct contracting. Panel A shows results without controls, and panel B includes purchase-level controls for the amount of the purchase, the responsible control department (UCE), the month of the purchase, the product code, and the month of the audit.

Column (1) shows that purchases from auctions were subject to 31 more checks than purchases from direct contracts, which were only subject to 21 checks. This corresponds to almost 2.5-times as many checks for auctions. The point estimate remains practically unchanged when including the controls. Column (4) shows that purchases made through auctions were marked for 1.6 more infractions without controls, and 2.7 more infractions with controls, compared to 2.5 detected infractions for purchases made through direct contracts. Including controls, the probability of an infraction being detected is therefore twice as high for purchases made through auctions rather than through direct contracts, holding the size of the purchase, the type of product, etc. constant.

In order to further investigate whether these differences stem indeed from the purchase modality and not from other unobserved differences between the purchases, we analyze the impact separately for the creation and the execution stage of the contract. The creation stage of a contract includes everything that precedes contract signing, such as choosing the procurement modality, writing the specifications for auctions, requesting quotes for direct contracts, evaluating the bids or offers, etc. This is where auctions and direct contracting differ the most. The contract execution stage refers to all activities following contract awarding, such as the timing of delivery, quality of the product or service, and whether contract specifications are met. In this stage, all else equal, contracts awarded through auctions and direct negations are quite similar.

Looking at Columns (3) and (4), we see that 90% of the higher number of checks for purchases made through auctions stem indeed from the creation stage. Columns (5) and (6) show that with regards to the additional detected infractions, without controls over 70% come from the creation stage, and with controls over 80%. The finding that most of the difference in numbers of checks and infractions occurs in the creation stage suggests that this is related to the different protocols for auctions and direct contracting, rather than some unobserved difference between purchases done through the two modalities. Finally, Column (7) shows the differences by modality in the probability that follow-up actions for serious infractions are taken. There is no significant effect. This suggests that the additional detected infractions are not due to serious malpractice, but rather due to more routine errors. However, even being detected for more routine infractions leads to significant consequences for the affected entities and the officers in question.[22]

These findings are consistent with the notion that running an auction mechanically involves a larger number of steps and complying with a larger number of rules and that auditors therefore have more steps to audit. Following each step of the regulation in this legalistic way may result in the unintended consequence of auditing purchases made by auctions more intensely, implicitly discouraging the use of auctions even though the government explicitly tries to promote them. On the other hand, it does not seem to be the case that entities are frequently detected or prosecuted for overuse of direct contracting. Focusing more on the justification step of the direct contracting process and reducing the number of checks conducted in routine auditing of auctions could potentially reduce this implicit disincentive for auctions.

6.2 Audit Risk Does Not Decrease the Following Year

An alternative mechanism could be that public agents relax after being audited because they think that the auditing selection rule makes it less likely that they will be audited again in the subsequent year. This mechanism is unlikely to drive the results, for two reasons.

[22]Following the audit, the Comptroller Agency issues a report detailing all the infractions of the agency in question, which is publicly available on the internet. The agency then needs to file a response explaining how the infractions happened and what they will do to remedy the issue. If the Comptroller Agency is not satisfied, additional reports are requested. During this process, officials who committed the errors in question typically need to explain themselves internally to their superiors, and qualitative interviews suggest that this is usually a painful process that can affect officials career opportunities.

First, as both procurement officers and auditors told us in a number of qualitative interviews, even if the audit probability were to temporarily dip in the year following an audit, that would not leave entities "protected" from scrutiny, as even audits conducted in later years will still be able to investigate purchases from several years back. So they told us that it would not be reasonable for agents to expect that their current purchases would be subject to less scrutiny.

Second, we find that in fact there is no such dip in the audit probability. Table 8 reports the estimates of the impact on the probability of being audited in the year after the audit.[23] Although the results are not statistically significant, if anything, there is a slightly increase in the probability of being audited in the following year.

6.3 RDD Subgroup Analysis Results

An additional way to shed light on mechanisms is often to conduct subgroup analyses to investigate whether treatment effects are larger for groups for which one would expect a larger effect when certain mechanisms are at play. One subgroup of interest in our case are entities that have a large share of direct contracting in their purchases prior to the audit. If public entities have a large share of direct contracting, they may be more likely to learn that overuse of direct contracting is not detected or punished. We analyze this issue by comparing entities with above-median share of direct contracting in the year prior to treatment.

However, entities with a high share of direct contracts differ from those with a low share in other ways. Tables A9 and A10 Columns (1) to (4) show that those with a high share of purchases made through direct contracting tend to be larger, more likely to have been previously audited, and more likely to be located in the central region, for example. (Table A9 shows the ± 4 range around the cutoff and Table A10 the ± 10 range). We would like to analyze the differential treatment effect between entities with high versus low shares of direct contracting while holding other covariates constant. However, as discussed in Section 4.3, using simple interactions to estimate the different effects by subgroup and to control for other differences is not generally valid in the RDD setting.[24]

[23] Figure A8 shows these results graphically.

[24] Tables A7 and A8 illustrate the specification bias that would arise in subgroup analyses if we used simple interaction terms in RDD. Panel A shows the analysis using an interaction term of the dummy for crossing the threshold with the dummy for having an above-median share of direct contracting in the pre-treatment year. The coefficients of this specification are clearly very different from the difference in

We therefore follow the propensity score weighting procedure laid out in Section 4.3. We start by constructing the propensity score for being in the group with above-median share of direct contracting in the ± 4 and ± 10 range. The propensity score is based on the following covariates: dummies for entity managed by left-wing or right-wing party in $t-1$ (omitted category is independent party), log (+1) of total amount of purchases in $t-1$ and $t-2$, dummy for being audited in $t-1$, dummy for being a municipal or national entity (omitted category is regional entities), year dummies and dummies for North and South of the country (omitted category is Central Region).

Columns (5) to (6) show how covariate balance improves with the inverse propensity score weighting. In both the ± 4 and the ± 10 range, the overall P-value for the test of joint significance of covariates changes from 0.000 to 1.000, and the average standardized mean difference is reduced from 0.148 to 0.016 in the ± 4 range and from 0.162 to 0.005 in the ± 10 range. This suggests that when running the RDD separately in each subgroup after weighting, any differential impact estimates we find are unlikely to be driven by differences in the characteristics shown in this table. As with any method that controls for other observable characteristics, this approach cannot rule out remaining differences in unobserved variables.

Since the first stage may not be identical for both subgroups, differences in the reduced form estimates (i.e. the impact of crossing the cutoff) between the subgroups might stem either from a differential treatment effect or from a difference in the strength of the first stage. To see the differential effect on each subgroup, we therefore run RDD IV regressions by subgroup as shown in Table 9 (impact of an audit on auction share) and Table 10 (impact of an audit on share of direct contracts). Panel A shows the results without weighting, and Panel B shows the results with propensity score weighting, which equalizes the observable covariates across the subgroups. The estimates of the difference in treatment effects are shown in row 3 of each panel.

Consistent with our hypothesis, the reduction in the auction share and the increase in the direct contracting share are larger for those with an above-median share of direct contracting prior to the intervention. However, the difference is not statistically significant. As we can see from Panel B in Table 10, with this sample size it is unlikely that we would ever find a significant difference between any two subgroups, unless one subgroup had a large positive and the other a large negative treatment effect.

the treatment effect for the two subgroups found by estimating the RDD separately within each subgroup in Panel B.

In sum, propensity score weighting essentially eliminates originally large covariate imbalances across these subgroups, allowing us to conduct separate RDD regressions in each group while holding other observables constant. Being able to do this reveals that there is no statistically significant difference in the impact between entities with above- and below-median direct-contracting share in the pre-treatment period.

7 Conclusion

This paper investigates the impact of audits by the Chilean Comptroller Agency Contraloría on subsequent public procurement practices, using a fuzzy RDD design based on a scoring rule that Contraloría used in 2011-2012 to allocate its audits. Being audited causes public entities to shift away from auctions to the less transparent and less competitive modality of direct contracting. This change in behavior goes against the goal of the Chilean procurement regulation to promote the use of auctions over direct contracting. The increase is particularly strong for direct contracts justified by emergency, which are notoriously prone for overuse and only require a quote from one firm. On the other hand, there is a large reduction in auctions with more than 3 bidders so the overall competitiveness of the procurement process is reduced.

In order to shed light on the potential underlying mechanisms, we worked with the Comptroller Agency to implement audits aimed at gathering additional data. Results from these audits show that holding the amount and type of purchase constant, audits undergo about 2.5 times as many checks as purchases made through direct contracting and lead to about twice as many detected infractions. The effects hold when controlling for other characteristics and are concentrated on the creation stage of the procurement process, where auctions and direct contracting differ the most. Learning that entities do not seem to get in trouble for overuse of direct contracting but on the contrary, are more likely to be called out for infractions when using auctions, may discourage procurement officers from using the already more work-intensive auctions that the government tries to promote. Overall, these results suggest that it is key not to think of audits merely as "neutral" verification and information extraction mechanisms, but to carefully consider potential impacts and incentives created by the audit design.

References

Abadie, Alberto, "Semiparamteric Difference-in-Differences Estimators," *Review of Economic Studies*, 2005, *72* (1), 1–19.

Bandiera, Oriana, Andrea Prat, and Tommaso Valletti, "Active and passive waste in government spending: evidence from a policy experiment," *The American Economic Review*, 2009, *99* (4), 1278–1308.

Banerjee, Abhijit, Rema Hanna, Jordan C Kyle, Benjamin A Olken, and Sudarno Sumarto, "Contracting out the Last-Mile of Service Delivery: Subsidized Food Distribution in Indonesia," Working Paper, National Bureau of Economic Research 2015.

Becker, Gary S, "Crime and punishment: An economic approach," in "The economic dimensions of crime," Springer, 1968, pp. 13–68.

Becker, Sascha, Peter Egger, and Maximilian von Ehrlich, "Absorptive Capacity and the Growth and Investment Effects of Regional Transfers: A Regression Discontinuity Design with Heterogeneous Treatment Effects," *Amercian Economic Journal: Economic Policy*, 2013, *5* (4), 29–77.

Bulow, Jeremy and Paul Klemperer, "Auctions Versus Negotiations," *American Economic Review*, 1996, *86* (1), 180–194.

Calonico, Sebastian, Matias Cattaneo, Max Farrell, and Rocıo Titiunik, "Regression discontinuity designs using covariates," Working Paper, University of Michigan 2017.

_ , Matias D. Cattaneo, and Rocıo Titiunik, "Robust Nonparametric Confidence Intervals for Regression-Discontinuity Designs," *Econometrica*, 2014, *82* (6), 2295–2326.

Carrillo, Paul, Dina Pomeranz, and Monica Singhal, "Dodging the taxman: Firm misreporting and limits to tax enforcement," *American Economic Journal: Applied Economics*, 2017, *9* (2), 144–164.

Chilean Electoral Service, "Elecciones Alcaldes 2004–2012," 2014.

ChileCompra, "16 de Octubre: Evento 10 Años de ChileCompra," 2013.

Coviello, Decio and Mario Mariniello, "Publicity Requirements in Public Procurement: Evidence from a Regression Discontinuity Design," *Journal of Public Economics*, 2014, *109*, 76–100.

_ , Andrea Guglielmo, and Giancarlo Spagnolo, "The Effect of Discretion on Procurement Performance: Evidence from a Regression Discontinuity Design," *Management Science*, 2015.

DeBacker, Jason, Bradley Heim, Anh Tran, and Alexander Yuskavage, "Legal Enforcement and Corporate Behavior: An Analysis of Tax Aggressiveness after an Audit," *Journal of Law and Economics*, 2015, *58* (2), 291 – 324.

_ , _ , _ , and _ , "Once bitten, twice shy? The lasting impact of IRS audits on individual tax reporting," *Journal of Financial Economics*, 2015, *117* (1), 122–138.

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

Duflo, Esther, "Richard T. Ely Lecture: The Economist as Plumber," *American Economic Review*, May 2017, *107* (5), 1–26.

_ , Michael Greenstone, Rohini Pande, and Nicholas Ryan, "The Value of Regulatory Discretion: Estimates from Environmental Inspections in India," Working Paper, National Bureau of Economic Research 2014.

Feir, Donna, Thomas Lemieux, and Vadim Marmer, "Weak identification in fuzzy regression discontinuity designs," *Journal of Business & Economic Statistics*, 2016, *34* (2), 185–196.

Ferraz, Claudio, Frederico Finan, and Dimitri Szerman, "Procuring firm growth: the effects of government purchases on firm dynamics," Working Paper, National Bureau of Economic Research 2016.

_ , _ , and Eric Avis, "Do Government Audits Reduce Corruption? Estimating the Impacts of Exposing Corrupt Politicians," *Forthcoming Journal of Political Economy*, 2017.

Hahn, Jinyong, Petra Todd, and Wilbert Van der Klaauw, "Identification and Estimation of Treatment Effects with a Regression-Discontinuity Design," *Econometrica*, 2001, *69* (1), 201–09.

Hastings, Justine S., Christopher A. Neilson, and Seth D. Zimmerman, "Are Some Degrees Worth More than Others? Evidence from College Admission Cutoffs in Chile," *NBER Working Paper w19241*, 2014.

Hsu, Yu-Chin and Shu Shen, "Testing for treatment effect heterogeneity in regression discontinuity design," Working Paper, Institute of Economics, Academia Sinica, Taipei, Taiwan 2016.

Imbens, Guido and Karthik Kalyanaraman, "Optimal Bandwidth Choice for the Regression Discontinuity Estimator," *Review of Economic Studies*, 2012, *79* (3), 933–959.

_ and Thomas Lemieux, "Regression Discontinuity Designs: A Guide to Practice," *Journal of Econometrics*, 2008, *142* (2), 615–635.

Kaufmann, Katja Maria, Matthias Messner, and Alex Solis, "Returns to Elite Higher Education in the Marriage Market: Evidence from Chile," Working Paper 489, IGIER (Innocenzo Gasparini Institute for Economic Research), Bocconi University 2013.

Kleven, Henrik, Martin Knudsen, Claus Kreiner, Sren Pedersen, and Emmanuel Saez, "Unwilling or Unable to Cheat? Evidence From a Tax Audit Experiment in Denmark," *Econometrica*, 2011, *79* (3), 651–692.

Lalive, Rafael, Armin Schmutzler, and Christine Zulehner, "Auctions vs Negotiations in Public Procurement: Which Works Better," Working Paper 209, University of Zurich 2015.

Lee, David S. and Thomas Lemieux, "Regression Discontinuity Designs in Economics," *Journal of Economic Literature*, 2010, *48* (2), 281–355.

Lewis-Faupel, Sean, Yusuf Neggers, Benjamin A. Olken, and Rohini Pande, "Can Electronic Procurement Improve Infrastructure Provision? Evidence from Public Works in India and Indonesia," *American Economic Journal: Economic Policy*, August 2016, *8* (3), 258–83.

Litschig, Stephan and Yves Zamboni, "Audit Risk and Rent Extraction: Evidence from a Randomized Evaluation in Brazil," Working Paper 554, Barcelona Graduate School of Economics 2016.

McCrary, Justin, "Manipulation of the Running Variable in the Regression Discontinuity Design: A Density Test," *Journal of Econometrics*, 2008, *142* (2), 698–714.

Niehaus, Paul and Sandip Sukhtankar, "Corruption dynamics: The golden goose effect," *American Economic Journal: Economic Policy*, 2013, *5* (4), 230–269.

OECD, "Government at a Glance 2015," Technical Report 2015.

Olken, Benjamin A., "Monitoring Corruption: Evidence from a Field Experiment in Indonesia," *Journal of Political Economy*, 2007, *115*, 200–249.

Pomeranz, Dina, "No Taxation Without Information: Deterrence and Self-Enforcement in the Value Added Tax," *American Economic Review*, 2015, *105* (8), 2539–2569.

Pop-Eleches, Cristian and Miguel Urquiola, "Going to a Better School: Effects and Behavioral Responses," *American Economic Review*, 2013, *103* (4), 1289–1325.

Tran, Anh, "Which Regulations Reduce Corruption? Evidence from the Internal Records of a Bribe-Paying Firm," Working Paper, Indiana University 2010.

Yang, Dean, "Can enforcement backfire? Crime displacement in the context of customs reform in the Philippines," *The Review of Economics and Statistics*, 2008, *90* (1), 1–14.

Table 1:
Summary Statistics

Purchase Modality	(1) Amount in Millions of USD	(2) Share of Total Value of Purchase	(3) Number of Orders	(4) Share of Total Number of Orders	(5) Average Number of Bidders/Quotes
			Panel A: Full Sample		
Auction	3,760	66.06%	2,595,662	51.15%	9.8
Direct contract	973	17.09%	706,016	13.91%	1.8
Framework agreement	938	16.48%	1,416,036	27.90%	
Small purchases	21	0.38%	356,901	7.03%	
			Panel B: Estimation Sample		
Auction	2,597	66.58%	1,827,455	52.76%	10.0
Direct contract	675	17.30%	482,816	13.94%	1.7
Framework agreement	613	15.72%	889,745	25.69%	
Small purchases	16	0.40%	263,575	7.61%	

Notes: The full sample consists of all public entities with medium level of risk (720 and 953 in 2011 and 2012 respectively). The estimation sample consists of public entities with medium risk whose normalized importance scores for the year in question was within the ± 10 range (610 and 392 in 2011 and 2012 respectively). Column (5) shows the average number of bidders in auctions and the average number of required quotes for direct contracting.

Table 2:
Balance Test for Pre-Treatment Variables

	(1)	(2)	(3)	(4)	(5)	(6)
	Comparison mean (±4)	Linear estimate (±4)	Linear estimate (±4)	Comparison mean (±10)	Quadratic estimate (±10)	Quadratic estimate (±10)
Auctions share, *t-1*	0.656	-0.007	-0.026	0.695	-0.035	−0.048*
		(0.035)	(0.026)		(0.035)	(0.026)
Direct contracting share, *t-1*	0.146	0.050	0.023	0.123	0.050	0.036
		(0.032)	(0.026)		(0.033)	(0.026)
Framework agreement share, *t-1*	0.183	-0.044	-0.004	0.168	-0.021	0.004
		(0.029)	(0.021)		(0.030)	(0.022)
Log (+1) of total amount purchased, *t-1*	13.331	0.317	-0.123	13.244	0.096	-0.100
		(0.322)	(0.128)		(0.311)	(0.130)
Auctions share, *t-2*	0.694	0.009	-0.032	0.731	-0.005	-0.056
		(0.041)	(0.036)		(0.039)	(0.036)
Direct contracting share, *t-2*	0.128	0.020	0.004	0.111	0.019	0.025
		(0.028)	(0.030)		(0.032)	(0.029)
Framework agreement share, *t-2*	0.155	-0.028	0.021	0.138	-0.014	0.024
		(0.029)	(0.026)		(0.028)	(0.024)
Log (+1) of total amount purchased, *t-2*	13.176	0.233	-0.001	13.079	-0.001	-0.096
		(0.349)	(0.178)		(0.339)	(0.140)
Audited, *t-1*	0.187	0.042	0.002	0.163	0.085	0.067
		(0.069)	(0.055)		(0.074)	(0.069)
Right-wing	0.671	-0.047	0.107*	0.695	-0.092	-0.003
		(0.099)	(0.056)		(0.111)	(0.058)
Independent	0.108	0.069	0.011	0.103	0.071	0.049
		(0.052)	(0.042)		(0.057)	(0.043)
F-statistic		0.70	1.20		0.84	1.50
[p-value]		[0.744]	[0.278]		[0.596]	[0.124]

Notes: This table tests whether there is a discontinuity for any of the covariates, by running an RDD with the covariates as an outcome, as in Equation (4). Columns (1) and (4) show RDD comparison means in the ±4 and ±10 range. Columns (2) and (3) show linear estimates in the ±4 range, without controls and with both stratum fixed effects and covariates, respectively. Columns (5) and (6) display the corresponding quadratic estimates. The covariates include: political affiliation (if appropriate), log (+1) of total amount purchased one and two years prior to the variable whose balance is being tested, a dummy for having been audited in the preceding year (audits data is not available for two years earlier) and auction and direct contract share measured one and two years prior. Standard errors are clustered at the stratum level. A stratum refers to a cell defined by year, responsible control department and type of entity. ***$p<0.01$, **$p<0.05$, *$p<0.1$.

Table 3:
Impact on Audit Probability

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Audit Probability							
$1\{\text{Relative importance} \geq \text{cutoff}\}$	0.296***	0.203**	0.181**	0.299***	0.220**	0.193**	0.158**	0.183**
	(0.074)	(0.078)	(0.069)	(0.088)	(0.087)	(0.084)	(0.067)	(0.076)
Bandwidth	±4	±4	±4	±10	±10	±10	±6.51	±6.51
Observations	482	482	477	1,002	1,002	992	716	716
R-squared	0.035	0.311	0.396	0.050	0.276	0.354	0.402	0.402
Comparison mean	0.136	0.136	0.136	0.071	0.071	0.071	0.118	0.118
Spline	Linear	Linear	Linear	Quadr.	Quadr.	Quadr.	Linear	Linear
Stratum fixed effects	No	Yes	Yes	No	Yes	Yes	Yes	Yes
Additional controls	No	No	Yes	No	No	Yes	Yes	Yes

Notes: First stage RDD estimates following the specification of Equation (3). Columns (7) and (8) employ the mean-squared-error-optimal bandwidth following Imbens and Kalyanaraman (2012). Column (8) in addition reports bias-corrected estimates and robust standard errors following Calonico et al. (2014). The control variables include political affiliation, log (+1) of total amount purchased one and two years prior, a dummy for having been audited in the preceding year (audits data is not available for two years earlier) and auction and direct contract share measured one and two year prior. Standard errors are clustered at the level of the strata. A stratum refers to a cell defined by year, responsible control department and type of entity. ***p<0.01, **p<0.05, *p<0.1

Table 4:
Share of Purchases through Auctions and Direct Contracting

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
				Panel A: Auctions				
1{Relative importance ≥ cutoff}	-0.065	-0.073^{*}	-0.069^{**}	-0.081^{**}	-0.126^{***}	-0.085^{***}	-0.079^{***}	-0.089^{**}
	(0.045)	(0.043)	(0.032)	(0.038)	(0.036)	(0.027)	(0.030)	(0.036)
Bandwidth	±4	±4	±4	±10	±10	±10	±5.19	±5.19
Observations	482	482	477	1,002	1,002	992	604	604
R-squared	0.030	0.350	0.614	0.016	0.257	0.578	0.573	0.573
Comparison mean	0.637	0.637	0.637	0.665	0.665	0.665	0.666	0.666
				Panel B: Direct Contracting				
1{Relative importance ≥ cutoff}	0.087^{***}	0.079^{**}	0.061^{**}	0.097^{***}	0.109^{***}	0.073^{***}	0.069^{***}	0.077^{***}
	(0.032)	(0.037)	(0.028)	(0.032)	(0.038)	(0.025)	(0.024)	(0.028)
Bandwidth	±4	±4	±4	±10	±10	±10	±5.05	±5.05
Observations	482	482	477	1,002	1,002	992	593	593
R-squared	0.043	0.221	0.535	0.017	0.114	0.508	0.498	0.498
Comparison mean	0.136	0.136	0.136	0.110	0.110	0.110	0.125	0.125
Spline	Linear	Linear	Linear	Quadr.	Quadr.	Quadr.	Linear	Linear
Stratum fixed effects	No	Yes	Yes	No	Yes	Yes	Yes	Yes
Additional controls	No	No	Yes	No	No	Yes	Yes	Yes

Notes: Reduced form RDD estimates following the specification of Equation (4). Columns (7) and (8) employ the mean-squared-error-optimal bandwidth following Imbens and Kalyanaraman (2012). Column (8) in addition reports bias-corrected estimates and robust standard errors following Calonico, Cattaneo and Titiunik (2014). The control variables include political affiliation, log (+1) of total amount purchased one and two years prior, a dummy for having been audited in the preceding year (audits data is not available for two years earlier) and auction and direct contract share measured one and two year prior. Standard errors are clustered at the stratum level. A stratum refers to a cell defined by year, responsible control department and type of entity. ***p<0.01, **p<0.05, *p<0.1

Table 5:
Share of Purchases through Direct Contracting by Justification

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Unique Supplier				Emergency			
1{Relative importance $\geq$ cutoff}	0.012	0.013*	0.015*	0.018**	0.042**	0.049***	0.044**	0.049**
	(0.008)	(0.008)	(0.008)	(0.009)	(0.018)	(0.018)	(0.018)	(0.021)
R-squared	0.468	0.395	0.413	0.413	0.297	0.204	0.269	0.269
Comparison mean	0.025	0.026	0.025	0.025	0.014	0.008	0.014	0.014
Observations	477	992	573	573	477	992	528	528
Bandwidth	±4	±10	±4.85	±4.85	±4	±10	±4.49	±4.49
	Trust in Suppliers				Disproportionate Cost			
1{Relative importance $\geq$ cutoff}	-0.004	0.001	0.004	0.004	-0.001	-0.002	-0.001	-0.001
	(0.006)	(0.006)	(0.003)	(0.004)	(0.002)	(0.002)	(0.002)	(0.002)
R-squared	0.513	0.442	0.427	0.427	0.299	0.271	0.323	0.323
Comparison mean	0.014	0.012	0.012	0.012	0.003	0.004	0.003	0.003
Observations	477	992	976	976	477	992	843	843
Bandwidth	±4	±10	±9.67	±9.67	±4	±10	±7.96	±7.96
	Cost Less than 720 USD				Other			
1{Relative importance $\geq$ cutoff}	0.001	-0.004	-0.004	-0.005	0.009	0.023	0.018	0.020
	(0.004)	(0.005)	(0.004)	(0.005)	(0.017)	(0.018)	(0.016)	(0.019)
R-squared	0.638	0.542	0.638	0.638	0.676	0.560	0.617	0.617
Comparison mean	0.017	0.015	0.017	0.017	0.062	0.045	0.043	0.043
Observations	477	992	470	470	477	992	722	722
Bandwidth	±4	±10	±3.92	±3.92	±4	±10	±6.58	±6.58
Spline	Linear	Quadr.	Linear	Linear	Linear	Quadr.	Linear	Linear
Stratum fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Additional controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Notes: Reduced form RDD estimates following the specification of Equation (4). Columns (3) and (7) employ the mean-squared-error-optimal bandwidth following Imbens and Kalyanaraman (2012). Columns (4) and (8) in addition reports bias-corrected estimates and robust standard errors following Calonico, Cattaneo and Titiunik (2014). The control variables include political affiliation, log (+1) of total amount purchased one and two years prior, and the outcome measured one and two years prior. Standard errors are clustered at the stratum level. A stratum refers to a cell defined by year, responsible control department and type of entity. ***p<0.01, **p<0.05, *p<0.1

Table 6:
Share of Purchases Made through Auctions and Direct Contracting
by Number of Suppliers

	(1)	(2)	(3)	(4)
	Auctions		Direct contracting	
	Bidders$\leq$3	Bidders$>$ 3	1 Quote	3 Quotes
1{Relative importance $\geq$ cutoff}	-0.040	-0.074^{***}	0.045^{*}	0.027
	(0.033)	(0.027)	(0.025)	(0.021)
Bandwidth	±7.87	±5.14	±6.34	±5.48
Observations	831	602	699	631
R-squared	0.431	0.308	0.497	0.332
Comparison mean	0.358	0.239	0.118	0.026
Spline	Linear	Linear	Linear	Linear
Stratum fixed effects	Yes	Yes	Yes	Yes
Additional controls	Yes	Yes	Yes	Yes

Notes: RDD estimates following the specification of Equation (4). Mean-squared-error-optimal bandwidth following Imbens and Kalyanaraman (2012) with bias-corrected estimates and robust standard errors following Calonico, Cattaneo and Titiunik (2014). The control variables include political affiliation, log (+1) of total amount purchased one and two years prior, a dummy for having been audited in the preceding year (audits data is not available for two years earlier) and outcome measured one and two year prior. Standard errors are clustered at the stratum level. A stratum refers to a cell defined by year, responsible control department and type of entity. ***p<0.01, **p<0.05, *p<0.1

Table 7:

Additional Audits: Checks and Infractions

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Checks			Infractions			Follow-up
	Total	Creation	Execution	Total	Creation	Execution	
	Panel A: Without Control Variables						
Auction	31.31***	28.29***	3.45***	1.65***	1.52***	0.62***	-0.085
	(2.43)	(1.90)	(0.67)	(0.53)	(0.45)	(0.22)	(0.132)
Constant	20.94***	7.33***	11.58***	2.48***	1.24***	0.52***	0.515***
	(2.03)	(1.32)	(0.50)	(0.58)	(0.42)	(0.16)	(0.143)
Observations	105	105	105	105	105	105	105
R-squared	0.643	0.757	0.166	0.041	0.058	0.066	0.006
	Panel B: With Control Variable						
Auction	31.81***	28.54***	3.13***	2.67**	2.27**	0.46	0.110
	(2.21)	(1.69)	(0.95)	(1.15)	(0.82)	(0.40)	(0.133)
Amount of purchase	Yes	Yes	Yes	Yes	Yes	Yes	Yes
UCE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Month of purchase	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Product code	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Dummy for September audit	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	104	104	104	104	104	104	104
R-squared	0.921	0.933	0.691	0.726	0.752	0.462	0.578

Notes: OLS estimations. Each observation is an audited purchase. Column (1) shows the total number of checks conducted. Columns (2) and (3) show the number of checks in the creation and execution stages of the purchase, respectively. Column (4) shows the total number of infractions detected. Columns (5) and (6) show the number of infractions in the creation and execution stages. Column (7) shows the probability of a follow-up action for serious infractions. Panel B has one less observation since control variables were missing for that purchase. Standard errors are clustered at the entity level. ***$p<0.01$, **$p<0.05$, *$p<0.1$

Table 8:
Impact on the Audit Probability in the Subsequent Year

	(1)	(2)	(3)	(4)
		Probability of Audit in *t+1*		
$1\{\text{Relative importance} \geq \text{cutoff}\}$	0.041	0.078	0.089	0.079
	(0.114)	(0.105)	(0.079)	(0.098)
Bandwidth	±4	±10	±7.38	±7.38
Observations	482	1,002	796	796
R-squared	0.361	0.266	0.274	0.274
Comparison mean	0.162	0.161	0.176	0.176
Spline	Linear	Quadratic	Linear	Linear
Stratum fixed effects	Yes	Yes	Yes	Yes
Additional controls	Yes	Yes	Yes	Yes

Notes: RDD estimates following the specification of Equation (3). Columns (3) and (4) employ the mean-squared-error-optimal bandwidth following Imbens and Kalyanaraman (2012). Column (4) in addition reports bias-corrected estimates and robust standard errors following Calonico, Cattaneo and Titiunik (2014). The control variables include political affiliation, log (+1) of total amount purchased one and two years prior, and the outcome measured one year prior (audits data not available for two years before treatment). Standard errors are clustered at the stratum level. A stratum refers to a cell defined by year, responsible control department and type of entity. ***p<0.01, **p<0.05, *p<0.1

Table 9:
Auction Share by Pre-Treatment Share of Direct Contracting
(RDD IV Estimates)

	(1)	(2)	(3)
		Panel A: Nonweighted	
Audit $\times$ 1{DC share $>$ p(50)}	-0.479	-0.486	-0.641
	(1.734)	(12.535)	(66.950)
Audit $\times$ 1{DC share $<$ p(50)}	0.008	-0.314	-0.260
	(19.240)	(1.779)	(5.136)
Difference Estimate	-0.487	-0.173	-0.381
	(19.306)	(12.849)	(66.296)
Observations	477	992	604
		Panel B: Propensity Score Weighted	
Audit $\times$ 1{DC share $>$ p(50)}	-0.442	-0.443	-0.581
	(29.813)	(2.455)	(3.454)
Audit $\times$ 1{DC share $<$ p(50)}	0.155	-0.199	-0.185
	(1.768)	(2.726)	(3.054)
Difference Estimate	-0.597	-0.244	-0.396
	(30.231)	(3.668)	(5.010)
Observations	471	990	600
Bandwidth	$\pm$4	$\pm$10	$\pm$5.19
Spline	Linear	Quadr.	Linear
Stratum fixed effects	Yes	Yes	Yes
Additional controls	Yes	Yes	Yes

Notes: RDD IV estimations. Column (3) employs the mean-squared-error-optimal bandwidth following Imbens and Kalyanaraman (2012). Panel B was obtained using equations (5) and (6). The control variables include political affiliation, log (+1) of total amount purchased one and two years prior, a dummy for having been audited in the preceding year (audits data is not available for two years earlier) and auction and direct contract share measured one and two year prior. Standard errors are bootstrapped. A stratum refers to a cell defined by year, responsible control department and type of entity. ***p<0.01, **p<0.05, *p<0.1

Table 10:
Direct Contracting Share by Pre-Treatment Share of Direct Contracting
(RDD IV Estimates)

	(1)	(2)	(3)
		Panel A: Nonweighted	
Audit $\times$ 1{DC share > p(50)}	0.408	0.467	0.541
	(2.117)	(1.079)	(11.875)
Audit $\times$ 1{DC share < p(50)}	-0.010	0.033	0.034
	(7.875)	(5.714)	(1.348)
Difference Estimate	0.418	0.434	0.507
	(8.135)	(5.713)	(11.954)
Observations	477	992	593
		Panel B: Propensity Score Weighted	
Audit $\times$ 1{DC share > p(50)}	0.389	0.424	0.486
	(2.543)	(0.718)	(16.141)
Audit $\times$ 1{DC share < p(50)}	-0.251	-0.014	-0.078
	(2.439)	(4.891)	(0.773)
Difference Estimate	0.640	0.438	0.564
	(3.217)	(4.935)	(16.196)
Observations	471	990	591
Bandwidth	$\pm$4	$\pm$10	$\pm$5.05
Spline	Linear	Quadr.	Linear
Stratum fixed effects	Yes	Yes	Yes
Additional controls	Yes	Yes	Yes

Notes: RDD IV estimations. Column (3) employs the mean-squared-error-optimal bandwidth following Imbens and Kalyanaraman (2012). Panel B was obtained using equations (5) and (6). The control variables include political affiliation, log (+1) of total amount purchased one and two years prior, a dummy for having been audited in the preceding year (audits data is not available for two years earlier) and auction and direct contract share measured one and two year prior. Standard errors are bootstrapped. A stratum refers to a cell defined by year, responsible control department and type of entity. ***p<0.01, **p<0.05, *p<0.1

Figure 1:
Timeline

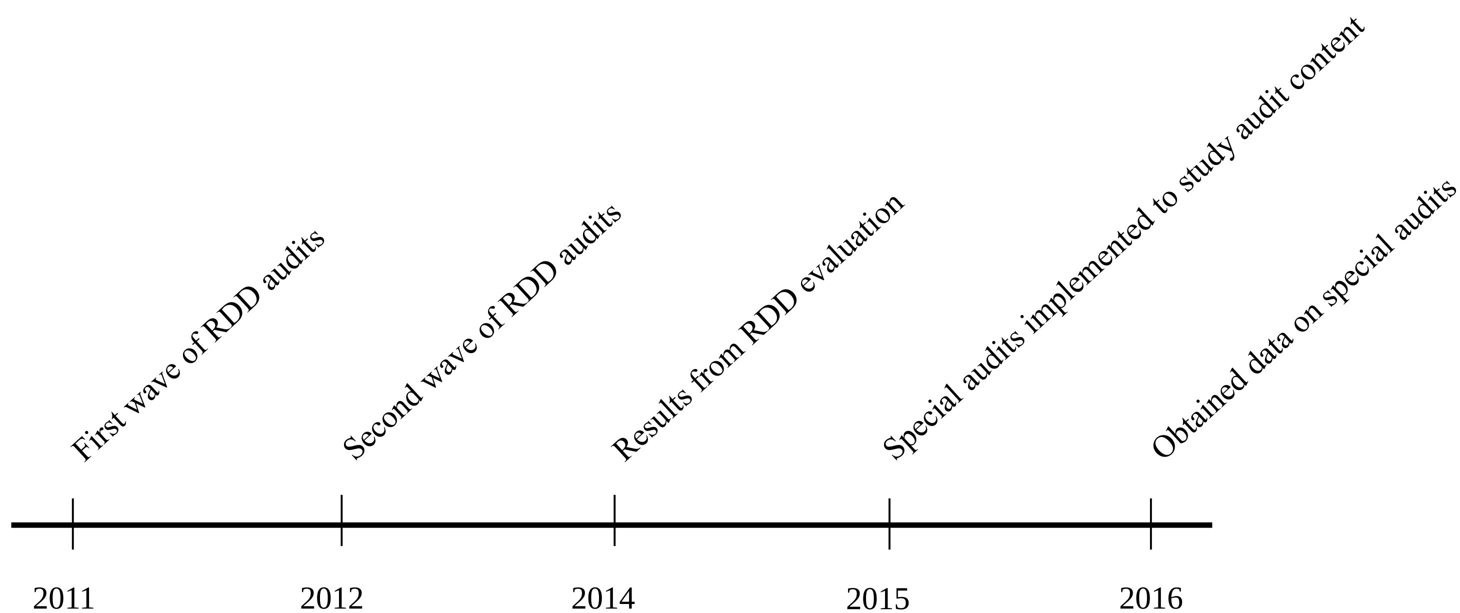
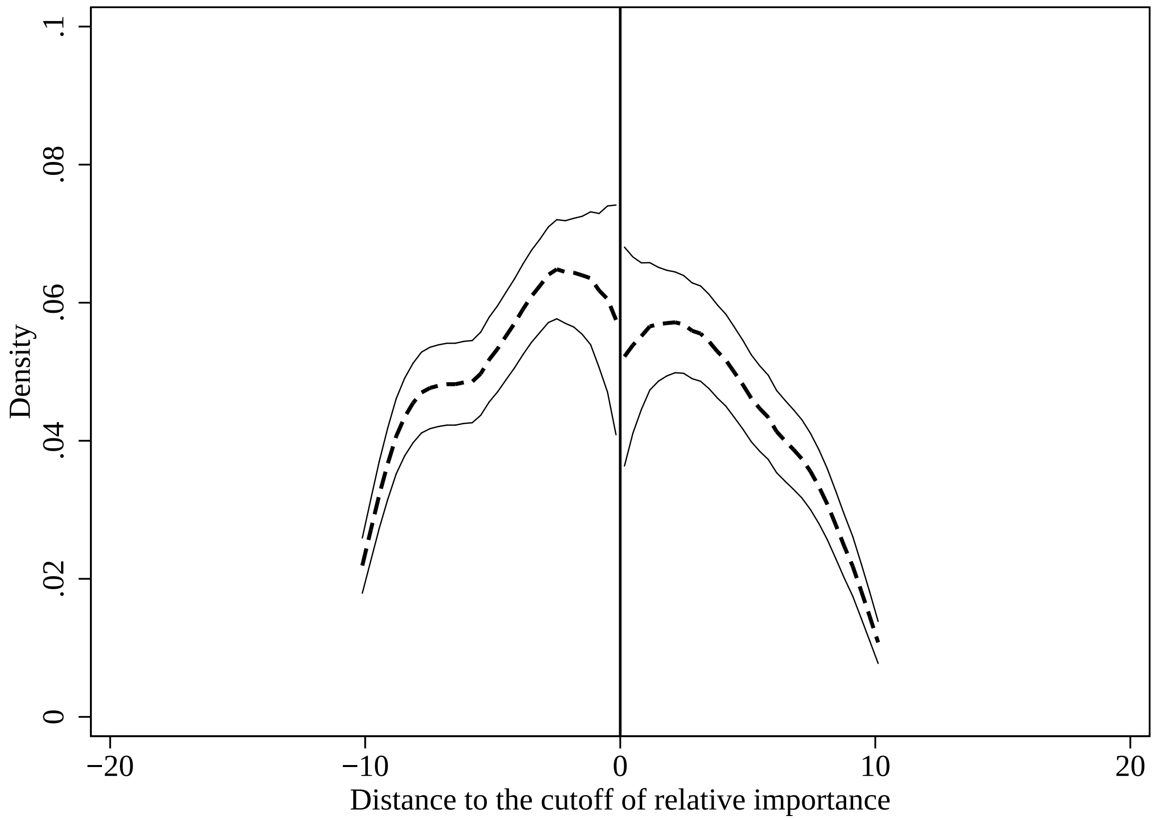


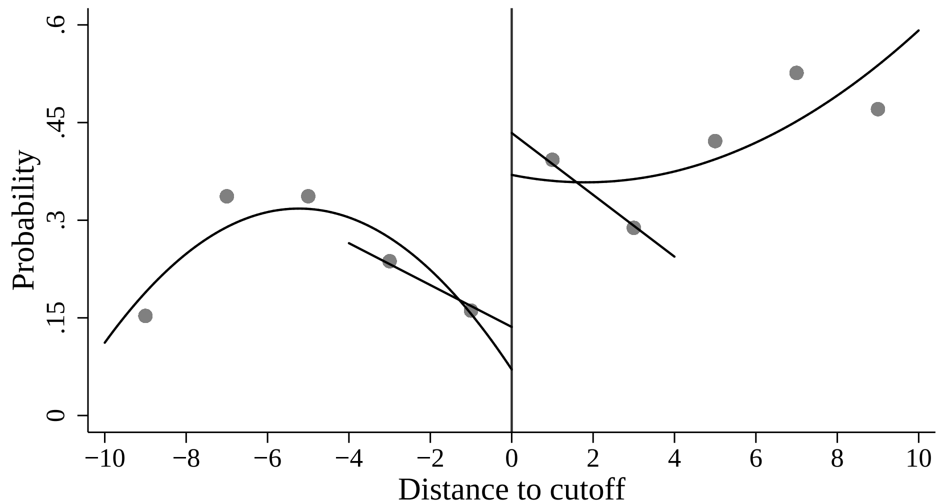
Figure 2:
McCrary Density Test



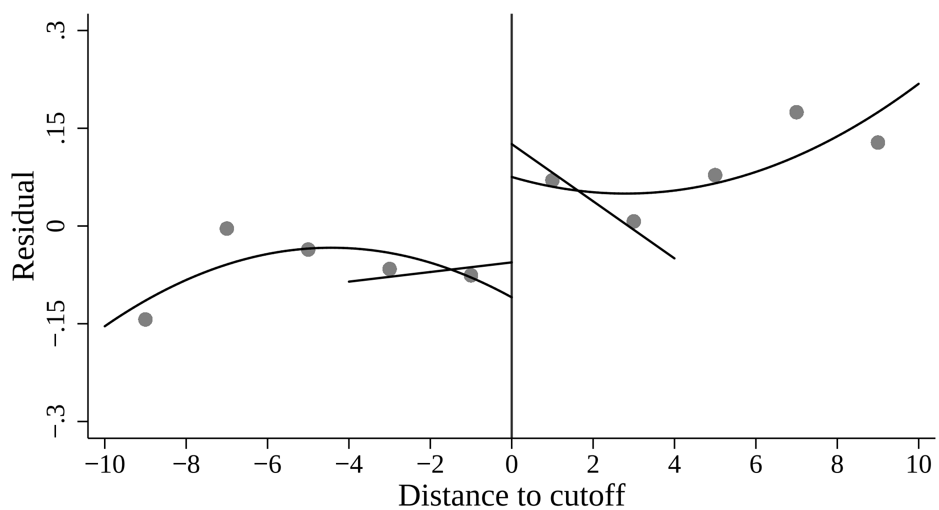
Notes: Dashed line indicates density estimate, and solid lines show the 95% confidence interval. The analysis includes the pooled sample of entities with medium level of risk in the ±10 range of the relative importance score. Relative importance scores are normalized by stratum-level cutoff. A stratum refers to a cell defined by year, responsible control department and type of entity.

Figure 3:
Audit Probability (First Stage)

(a) Audit Probability

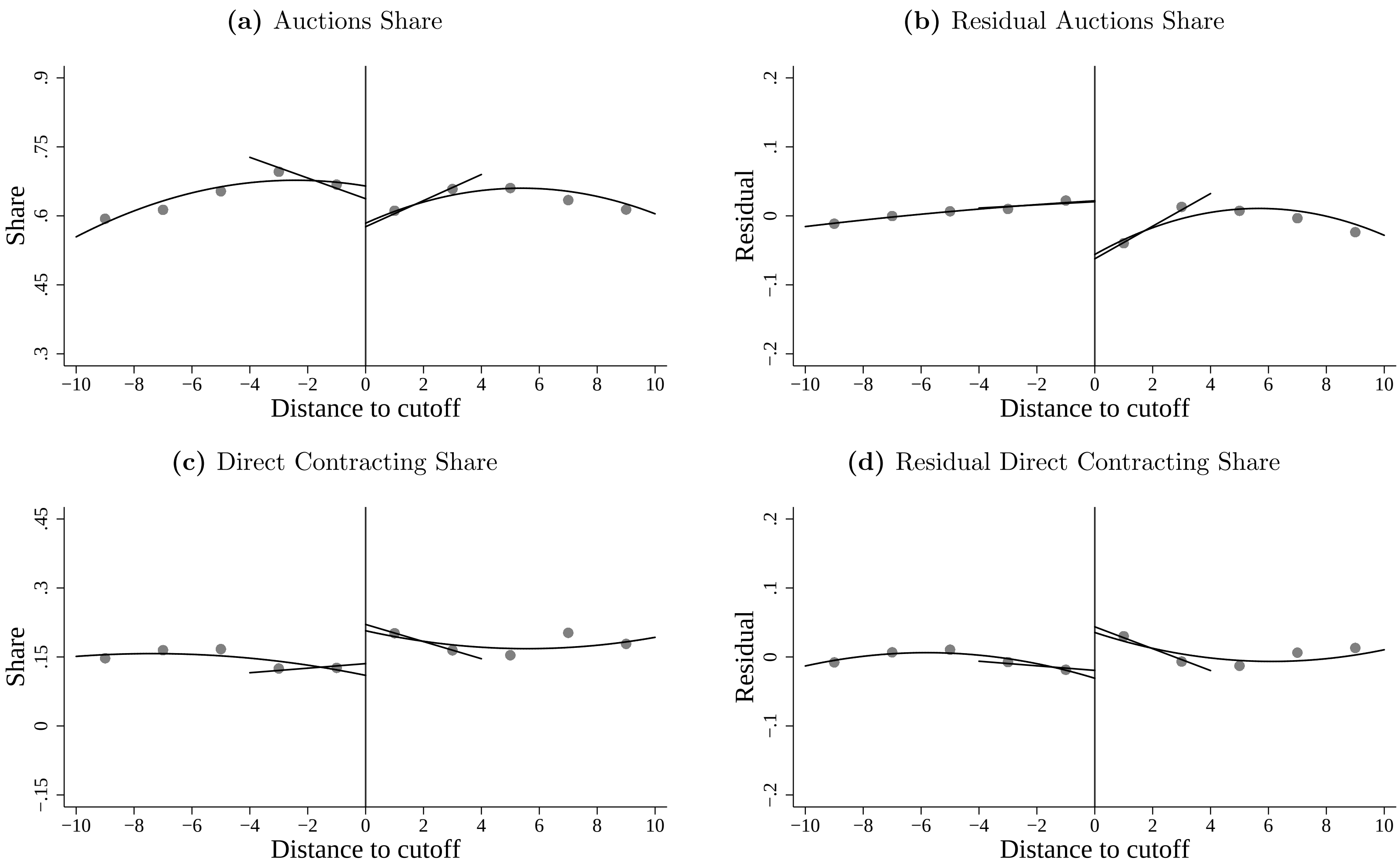


(b) Residual Audit Probability



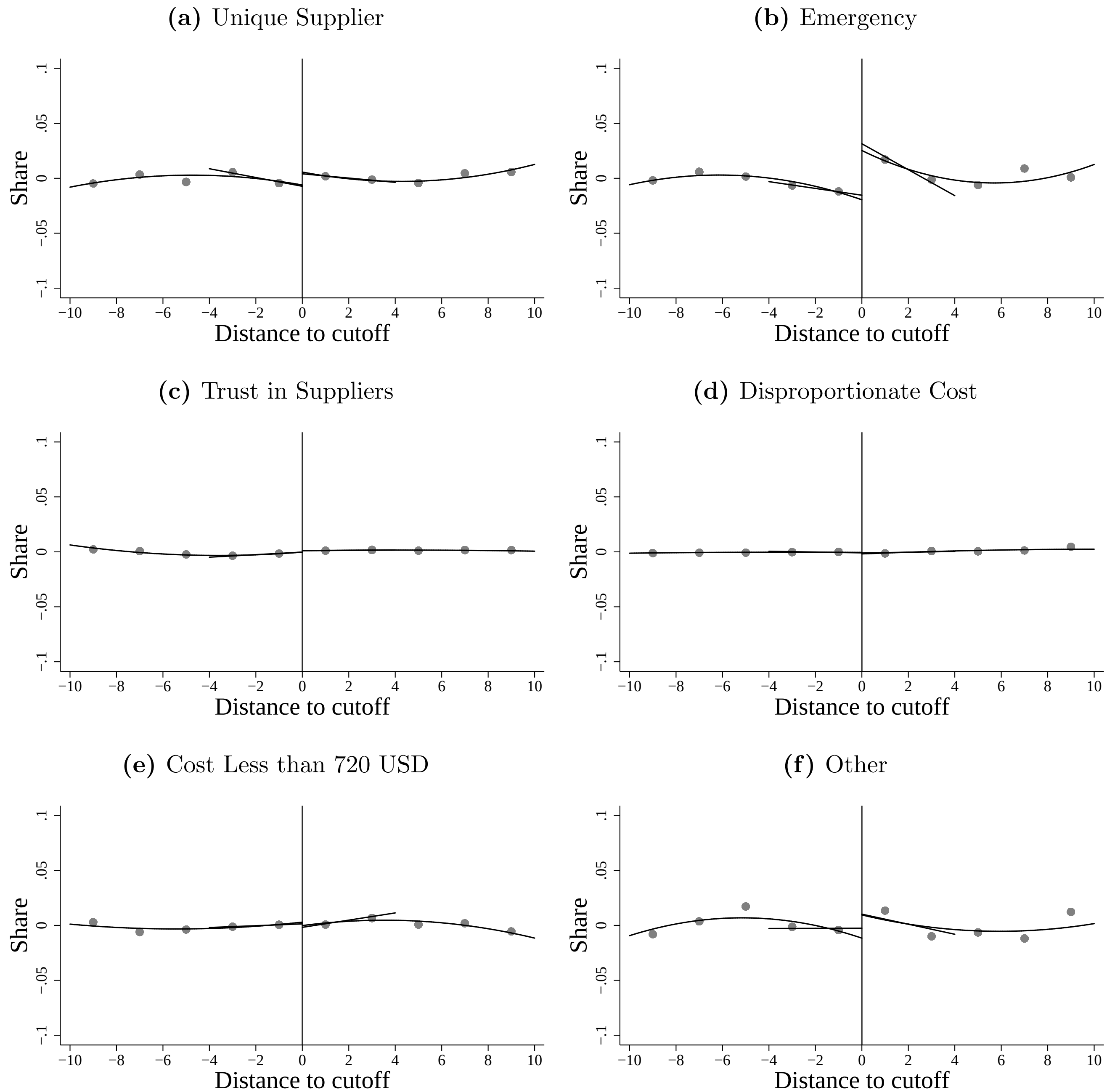
Notes: The figures show the audit probability for entities with medium level of risk in the ±10 range of the relative importance score. In panel *(a)*, the dots represent audit probabilities averaged within 2-point-wide intervals of the relative importance score. In panel *(b)* the dots represent averaged residuals. The residuals are obtained from a regression of the dummy for having been audited in the year corresponding to the score on stratum fixed effects and control variables. Control variables include political affiliation, log (+1) of total amount purchased one and two years prior, a dummy for having been audited in the preceding year (audits data is not available for two years earlier) and auction and direct contract share measured one and two years prior. Relative importance scores are normalized by stratum-level cutoff. A stratum refers to a cell defined by year, responsible control department and type of entity. Solid lines show linear and quadratic fits.

Figure 4:
Share of Purchases through Auctions and Direct Contracting



Notes: The figures show the amounts purchased through auctions and direct contracting as shares of total purchases, for entities with medium level of risk in the ± 10 range of the relative importance score. In panels *(a)* and *(c)*, the dots represent modality shares averaged within 2-point-wide intervals of the relative importance score. In panels *(b)* and *(d)*, the dots represent averaged residuals. The residuals are obtained from a regression of the outcome (amount purchased through a given modality over total amount purchased) in the year corresponding to the score on stratum fixed effects and control variables. Control variables include political affiliation, log (+1) of total amount purchased one and two years prior, a dummy for having been audited in the preceding year (audits data is not available for two years earlier) and auction and direct contract share measured one and two years prior. The relative importance score for each entity is normalized by the stratum-level cutoff. A stratum refers to a cell defined by year, responsible control department and type of entity. Solid lines show linear and quadratic fits.

Figure 5:
Share of Purchases by Justifications for Direct Contracting



Notes: The figures show the amounts purchased through each justification of direct contracting as shares of total purchases, for entities with medium level of risk in the ±10 range of the relative importance score. In panels *(a)* and *(c)*, the dots represent justification shares averaged within 2-point-wide intervals of the relative importance score. In panels *(b)* and *(d)*, the dots represent averaged residuals. The residuals are obtained from a regression of the outcome (amount purchased through a given justification over total amount purchased) in the year corresponding to the score on stratum fixed effects and control variables. Control variables include political affiliation, log of total amount purchased +1 one and two years prior, and the outcome measured one and two years prior. The relative importance score for each entity is normalized by the stratum-level cutoff. A stratum refers to a cell defined by year, responsible control department and type of entity. Solid lines show linear and quadratic fits.

A Appendix Tables and Figures

Table A1:
Audit Probability (First Stage) by Year

	(1)	(2)	(3)	(4)
		Panel A: 2011		
1{Relative importance $\geq$ cutoff}	0.101	0.104	0.099	0.098
	(0.067)	(0.079)	(0.075)	(0.087)
Bandwidth	$\pm$4	$\pm$10	$\pm$4.99	$\pm$4.99
Observations	321	610	392	392
R-squared	0.310	0.307	0.307	0.307
Comparison mean	0.075	0.040	0.092	0.092
		Panel B: 2012		
1{Relative importance $\geq$ cutoff}	0.395**	0.476***	0.334***	0.359**
	(0.164)	(0.154)	(0.128)	(0.147)
Bandwidth	$\pm$4	$\pm$10	$\pm$5.72	$\pm$5.72
Observations	161	392	222	222
R-squared	0.439	0.352	0.383	0.383
Comparison mean	0.278	0.214	0.260	0.260
Spline	Linear	Quadr.	Linear	Linear
Stratum fixed effects	Yes	Yes	Yes	Yes
Additional controls	Yes	Yes	Yes	Yes

Notes: RDD estimates following the specification of Equation (3). Standard errors are clustered at the stratum level. A stratum refers to a cell defined by each combination of responsible control department and type of entity. The control variables include political affiliation, log of total amount purchased one and two years prior, and a dummy for having been audited in the preceding year (audits data is not available for two years before treatment). For 2011, the analysis focuses on the cutoff between low and medium levels of relative importance, while for 2012, it focuses on the cutoff between medium and high importance. ***p<0.01, **p<0.05, *p<0.1.

Table A2:
Share of Purchases through Framework Agreements and as Small Purchases

	(1)	(2)	(3)	(4)
	Panel A: Framework Agreement			
1{Relative importance $\geq$ cutoff}	0.004	0.012	0.009	0.009
	(0.026)	(0.025)	(0.019)	(0.024)
Bandwidth	± 4	± 10	± 7.32	± 7.32
Observations	477	992	784	784
R-squared	0.656	0.615	0.604	0.604
Comparison mean	0.212	0.210	0.190	0.190
	Panel B: Small Purchases			
1{Relative importance $\geq$ cutoff}	-0.002	-0.004	-0.002	-0.003
	(0.004)	(0.004)	(0.002)	(0.003)
Bandwidth	± 4	± 10	± 8.86	± 8.86
Observations	477	992	915	915
R-squared	0.624	0.663	0.657	0.657
Comparison mean	0.016	0.015	0.014	0.014
Spline	Linear	Quadratic	Linear	Linear
Stratum fixed effects	Yes	Yes	Yes	Yes
Additional controls	Yes	Yes	Yes	Yes

Notes: Reduced form RDD estimates following the specification of Equation (4). Standard errors are clustered at the stratum level. A stratum refers to a cell defined by year, responsible control department and type of entity. The control variables include political affiliation, log of total amount purchased +1 one and two years prior, and the outcome measured one and two years prior. Columns (3) and (4) the employ mean-squared-error-optimal bandwidth following Imbens and Kalyanaraman (2012). Column (4) in addition reports bias-corrected estimates and robust standard errors following Calonico, Cattaneo and Titiunik (2014). ***p<0.01, **p<0.05, *p<0.1.

Table A3:
Share of Purchases through Auctions and Direct Contracting in 2011

	(1)	(2)	(3)	(4)
	Panel A: Auctions			
$1\{\text{Relative importance} \geq \text{cutoff}\}$	-0.086^{***}	-0.079^{**}	-0.068^{**}	-0.072^{*}
	(0.030)	(0.030)	(0.035)	(0.042)
Bandwidth	± 4	± 10	± 5.33	± 5.33
Observations	317	602	404	404
R-squared	0.630	0.598	0.567	0.567
Comparison mean	0.630	0.639	0.668	0.668
	Panel B: Direct Contracting			
$1\{\text{Relative importance} \geq \text{cutoff}\}$	0.082^{*}	0.061	0.054^{**}	0.059^{**}
	(0.044)	(0.039)	(0.024)	(0.029)
Bandwidth	± 4	± 10	± 5.80	± 5.80
Observations	317	602	432	432
R-squared	0.389	0.331	0.485	0.485
Comparison mean	0.139	0.131	0.123	0.123
Spline	Linear	Quadratic	Linear	Linear
Stratum fixed effects	Yes	Yes	Yes	Yes
Additional controls	Yes	Yes	Yes	Yes

Notes: Reduced form RDD estimates following the specification of Equation (4). Standard errors are clustered at the stratum level. A stratum refers to a cell defined by year, responsible control department and type of entity. The control variables include political affiliation, log of total amount purchased +1 one and two years prior, and the outcome measured one and two years prior. Columns (3) and (4) employ the mean-squared-error-optimal bandwidth following Imbens and Kalyanaraman (2012). Column (4) in addition reports bias-corrected estimates and robust standard errors following Calonico, Cattaneo and Titiunik (2014). ***p<0.01, **p<0.05, *p<0.1.

Table A4:
Share of Purchases through Auctions and Direct Contracting in 2012

	(1)	(2)	(3)	(4)
	Panel A: Auctions			
$1\{\text{Relative importance} \geq \text{cutoff}\}$	-0.016	-0.086	-0.135^{*}	-0.161^{*}
	(0.057)	(0.066)	(0.073)	(0.087)
Bandwidth	±4	±10	±4.11	±4.11
Observations	160	390	162	162
R-squared	0.635	0.582	0.625	0.625
Comparison mean	0.658	0.709	0.668	0.668
	Panel B: Direct Contracting			
$1\{\text{Relative importance} \geq \text{cutoff}\}$	0.062	0.113	0.109^{*}	0.127^{*}
	(0.050)	(0.071)	(0.058)	(0.069)
Bandwidth	±4	±10	±4.83	±4.83
Observations	160	390	195	195
R-squared	0.391	0.334	0.551	0.551
Comparison mean	0.132	0.092	0.137	0.137
Spline	Linear	Quadratic	Linear	Linear
Stratum fixed effects	Yes	Yes	Yes	Yes
Additional controls	Yes	Yes	Yes	Yes

Notes: Reduced form RDD estimates following the specification of Equation (4). Standard errors are clustered at the stratum level. A stratum refers to a cell defined by year, responsible control department and type of entity. The control variables include political affiliation, log of total amount purchased +1 one and two years prior, and the outcome measured one and two years prior. Columns (3) and (4) employ the mean-squared-error-optimal bandwidth following Imbens and Kalyanaraman (2012). Column (4) in addition reports bias-corrected estimates and robust standard errors following Calonico, Cattaneo and Titiunik (2014). ***p<0.01, **p<0.05, *p<0.1.

Table A5:
Share of Purchases through Auctions and Direct Contracts
Interacting the Running Variable with Stratum Dummies

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
				Panel A: Auctions				
1{Relative importance $\geq$ cutoff}	-0.065	-0.092^{**}	-0.089^{***}	-0.081^{**}	-0.110^{***}	-0.083^{***}	-0.086^{***}	-0.095^{***}
	(0.045)	(0.045)	(0.033)	(0.038)	(0.037)	(0.029)	(0.028)	(0.034)
Bandwidth	± 4	± 4	± 4	± 10	± 10	± 10	± 5.16	± 5.16
Observations	482	482	477	1,002	1,002	992	603	603
R-squared	0.030	0.456	0.675	0.016	0.329	0.628	0.630	0.630
Comparison mean	0.637	0.637	0.637	0.665	0.665	0.665	0.668	0.668
				Panel B: Direct Contracting				
1{Relative importance $\geq$ cutoff}	0.087^{***}	0.081^{**}	0.056^{**}	0.097^{***}	0.093^{**}	0.062^{**}	0.064^{***}	0.072^{***}
	(0.032)	(0.039)	(0.027)	(0.032)	(0.041)	(0.027)	(0.023)	(0.027)
Bandwidth	± 4	± 4	± 4	± 10	± 10	± 10	± 5.30	± 5.30
Observations	482	482	477	1,002	1,002	992	615	615
R-squared	0.043	0.367	0.604	0.017	0.183	0.576	0.575	0.575
Comparison mean	0.136	0.136	0.136	0.110	0.110	0.110	0.117	0.117
Spline	Linear	Linear	Linear	Quadr.	Quadr.	Quadr.	Linear	Linear
Stratum fixed effects	No	Yes	Yes	No	Yes	Yes	Yes	Yes
Additional controls	No	No	Yes	No	No	Yes	Yes	Yes

Notes: Reduced form RDD estimates following the specification of Equation (4) and additionally interacting each stratum dummy with the distance to the cutoff. Columns (7) and (8) employ the mean-squared-error-optimal bandwidth following Imbens and Kalyanaraman (2012). Column (8) in addition reports bias-corrected estimates and robust standard errors following Calonico, Cattaneo and Titiunik (2014). The control variables include political affiliation, log (+1) of total amount purchased one and two years prior, a dummy for having been audited in the preceding year (audits data is not available for two years earlier) and auction and direct contract share measured one and two year prior. Standard errors are clustered at the stratum level. A stratum refers to a cell defined by year, responsible control department and type of entity. ***p<0.01, **p<0.05, *p<0.1

Table A6:
Share of Purchases through Auctions and Direct Contracting in Subsequent Year

	(1)	(2)	(3)	(4)
		Panel A: Auctions in *t+1*		
1{Relative importance ≥ cutoff}	-0.039	-0.048	-0.041	-0.052
	(0.037)	(0.035)	(0.031)	(0.036)
Bandwidth	±4	±10	±5.54	±5.54
Observations	476	990	634	634
R-squared	0.543	0.493	0.478	0.478
Comparison mean	0.605	0.620	0.627	0.627
		Panel B: Direct Contracting in *t+1*		
1{Relative importance ≥ cutoff}	0.018	0.031	0.025	0.029
	(0.030)	(0.028)	(0.022)	(0.027)
Bandwidth	±4	±10	±6.27	±6.27
Observations	476	990	697	697
R-squared	0.512	0.428	0.442	0.442
Comparison mean	0.168	0.159	0.143	0.143
Spline	Linear	Quadratic	Linear	Linear
Stratum fixed effects	Yes	Yes	Yes	Yes
Additional controls	Yes	Yes	Yes	Yes

Notes: Reduced form RDD estimates following the specification of Equation (4). Standard errors are clustered at the stratum level. A stratum refers to a cell defined by year, responsible control department and type of entity. The control variables include political affiliation, log of total amount purchased +1 one and two years prior, and the outcome measured one and two years prior. Columns (3) and (4) employ the mean-squared-error-optimal bandwidth following Imbens and Kalyanaraman (2012). Column (4) in addition reports bias-corrected estimates and robust standard errors following Calonico, Cattaneo and Titiunik (2014). ***p<0.01, **p<0.05, *p<0.1.

Table A7:
Specification Bias in Subgroup Analysis with Interaction Terms in RDD: Auctions

	(1)	(2)	(3)
	Panel A: Estimation with Interaction Term		
1{Relative importance $\geq$ cutoff}	-0.048	-0.080^{***}	-0.038
	(0.038)	(0.030)	(0.031)
1{RI $\geq$ cutoff} $\times$ 1{DC > p(50)}	-0.030	-0.008	-0.032
	(0.040)	(0.026)	(0.035)
Sum of estimates	-0.079	-0.088^{**}	-0.070
	(0.055)	(0.039)	(0.047)
Bandwidth	±4	±10	±5.43
Observations	477	992	630
R-squared	0.607	0.576	0.560
	Panel B: Separate Estimation		
1{RI $\geq$ cutoff} $\times$ 1{DC share > p(50)}	-0.129^{***}	-0.114^{**}	-0.103^{**}
	(0.042)	(0.045)	(0.043)
1{RI $\geq$ cutoff} $\times$ 1{DC share < p(50)}	-0.004	-0.045	-0.043
	(0.061)	(0.052)	(0.042)
Difference Estimate	-0.125	-0.069	-0.060
	(0.085)	(0.077)	(0.060)
Bandwidth	±4	±10	±5.43
Observations	477	992	630
R-squared	0.666	0.624	0.608
Spline	Linear	Quadratic	Linear
Stratum fixed effects	Yes	Yes	Yes
Additional controls	Yes	Yes	Yes

Notes: RDD estimates following the specification of Equation (4). Standard errors are clustered at the stratum level. A stratum refers to a cell defined by year, responsible control department and type of entity. The control variables include political affiliation, log of total amount purchased +1 one and two years prior, and the outcome measured one and two years prior. Column (3) employs the mean-squared-error-optimal bandwidth following Imbens and Kalyanaraman (2012). ***p<0.01, **p<0.05, *p<0.1.

Table A8:
Specification Bias in Subgroup Analysis with Interaction Terms in RDD:
Direct Contracting

	(1)	(2)	(3)
	Panel A: Estimation with Interaction Term		
1{Relative importance $\geq$ cutoff}	0.036	0.066***	0.030
	(0.025)	(0.025)	(0.019)
1{RI $\geq$ cutoff} $\times$ 1{DC > p(50)}	0.066***	0.012	0.040**
	(0.025)	(0.016)	(0.020)
Sum of estimates	0.102***	0.078***	0.070**
	(0.036)	(0.029)	(0.027)
Bandwidth	±4	±10	±5.29
Observations	477	992	615
R-squared	0.446	0.507	0.510
	Panel B: Separate Estimation		
1{RI $\geq$ cutoff} $\times$ 1{DC share > p(50)}	0.113**	0.109**	0.092**
	(0.045)	(0.042)	(0.040)
1{RI $\geq$ cutoff} $\times$ 1{DC share < p(50)}	0.002	0.003	0.027
	(0.026)	(0.025)	(0.029)
Difference Estimate	0.111**	0.105*	0.065
	(0.054)	(0.053)	(0.049)
Bandwidth	±4	±10	±5.29
Observations	477	992	615
R-squared	0.597	0.580	0.563
Spline	Linear	Quadratic	Linear
Stratum fixed effects	Yes	Yes	Yes
Additional controls	Yes	Yes	Yes

Notes: RDD estimates following the specification of Equation (4). Standard errors are clustered at the stratum level. A stratum refers to a cell defined by year, responsible control department and type of entity. The control variables include political affiliation, log of total amount purchased +1 one and two years prior, and the outcome measured one and two years prior. Column (3) employs the mean-squared-error-optimal bandwidth following Imbens and Kalyanaraman (2012). ***p<0.01, **p<0.05, *p<0.1.

Table A9:
Balance Improvement: Sample in ±4 Range around the Cutoff

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Original Balance				Balance after Propensity Score-Weighting			
	Low share (n=243)	High share (n=237)			Low share (n=234)	High share (n=237)		
	Mean	Mean	St.mean difference	P-value	Mean	Mean	St.mean difference	P-value
Dummy(independent)	0.173	0.135	0.105	0.252	0.150	0.154	-0.011	0.904
Dummy(left-wing)	0.222	0.232	-0.023	0.797	0.221	0.221	-0.001	0.992
Dummy(right-wing)	0.605	0.633	-0.058	0.529	0.629	0.624	0.009	0.921
Log (+1) of total amount in t-1	13.272	13.914	-0.385	0.000	13.686	13.622	0.039	0.674
Log (+1) of total amount in t-2	12.955	13.853	-0.463	0.000	13.581	13.507	0.047	0.617
Dummy(audited in t-1)	0.148	0.295	-0.355	0.000	0.231	0.226	0.012	0.895
Dummy(municipal entity)	0.564	0.519	0.090	0.326	0.528	0.527	0.003	0.975
Dummy(national entity)	0.128	0.148	-0.058	0.523	0.126	0.136	-0.028	0.755
Dummy(regional entity)	0.309	0.333	-0.053	0.563	0.346	0.337	0.018	0.850
Dummy(2012)	0.362	0.304	0.124	0.176	0.335	0.338	-0.006	0.948
Dummy(2011)	0.638	0.696	-0.124	0.176	0.665	0.662	0.006	0.948
Dummy(Central Region)	0.576	0.624	-0.099	0.281	0.582	0.585	-0.006	0.950
Dummy(North)	0.156	0.110	0.137	0.133	0.134	0.139	-0.016	0.862
Dummy(South)	0.267	0.266	0.004	0.967	0.284	0.276	0.019	0.840
Abs(Standardized mean diff.)			0.148				0.016	
F-statistic for joint significance				4.429				0.047
P-value for joint significance				0.000				1.000

Notes: Columns (1) and (5) show the mean of each variable for entities with below-median share of purchases made through direct contracting. Columns (2) and (6) show the means for entities with above-median shares of direct contracts. Columns (3) and (7) show the standardized mean differences. Columns (4) and (8) show the p-values of t-tests for statistical significance of the difference in means between the two groups. Propensity score-weighted statistics are based on equations (5) and (6).

Table A10:

Balance Improvement: Sample in ±10 Range around the Cutoff

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
		Original Balance			Balance after Propensity Score-Weighting			
	Low share (n=478)	High share (n=520)			Low share (n=471)	High share (n=519)		
	Mean	Mean	St.mean difference	P-value	Mean	Mean	St.mean difference	P-value
Dummy(independent)	0.180	0.137	0.119	0.060	0.152	0.154	-0.006	0.922
Dummy(left-wing)	0.230	0.200	0.073	0.247	0.213	0.215	-0.007	0.915
Dummy(right-wing)	0.590	0.663	-0.152	0.016	0.635	0.630	0.010	0.870
Log (+1) of total amount in t-1	13.464	13.933	-0.273	0.000	13.748	13.736	0.007	0.917
Log (+1) of total amount in t-2	13.155	13.819	-0.335	0.000	13.612	13.593	0.011	0.861
Dummy(audited in t-1)	0.186	0.308	-0.281	0.000	0.249	0.251	-0.003	0.966
Dummy(municipal entity)	0.579	0.481	0.198	0.002	0.519	0.521	-0.004	0.947
Dummy(national entity)	0.142	0.204	-0.162	0.010	0.178	0.177	0.002	0.976
Dummy(regional entity)	0.278	0.315	-0.081	0.200	0.304	0.302	0.003	0.963
Dummy(2012)	0.410	0.375	0.072	0.258	0.398	0.396	0.005	0.940
Dummy(2011)	0.590	0.625	-0.072	0.258	0.602	0.604	-0.005	0.940
Dummy(Central Region)	0.571	0.665	-0.194	0.002	0.618	0.617	0.001	0.984
Dummy(North)	0.138	0.096	0.131	0.039	0.119	0.120	-0.003	0.962
Dummy(South)	0.291	0.238	0.119	0.061	0.263	0.263	0.001	0.990
Abs(Standardized mean diff.)			0.162				0.005	
F-statistic for joint significance				6.437				0.010
P-value for joint significance				0.000				1.000

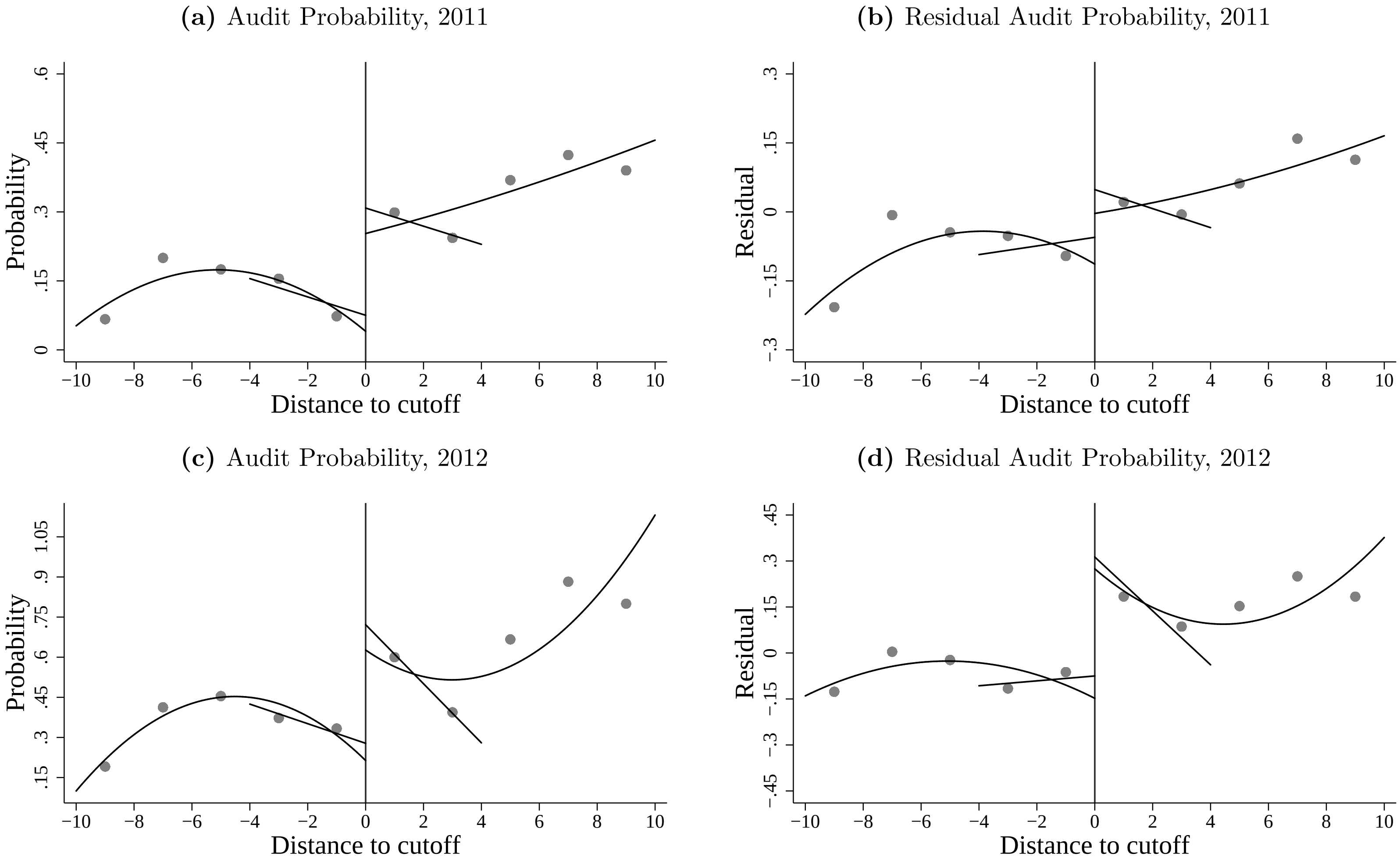
Notes: Columns (1) and (5) show the mean of each variable for entities with below-median share of purchases made through direct contracting. Columns (2) and (6) show the means for entities with above-median shares of direct contracts. Columns (3) and (7) show the standardized mean differences. Columns (4) and (8) show the p-values of t-tests for statistical significance of the difference in means between the two groups. Propensity score-weighted statistics are based on equations (5) and (6).

Table A11:
Share of Purchases through Auctions and Direct Contracting
Municipalities Only

	(1)	(2)	(3)	(4)
	Panel A: Auctions			
$1\{\text{Relative importance} \geq \text{cutoff}\}$	-0.091^{**}	-0.068^{***}	-0.092^{***}	-0.102^{***}
	(0.034)	(0.024)	(0.032)	(0.038)
Bandwidth	± 4	± 10	± 3.73	± 3.73
Observations	257	522	243	243
R-squared	0.422	0.548	0.440	0.440
Comparison mean	0.758	0.776	0.764	0.764
	Panel B: Direct Contracting			
$1\{\text{Relative importance} \geq \text{cutoff}\}$	0.082^{**}	0.060	0.075^{***}	0.083^{***}
	(0.031)	(0.036)	(0.025)	(0.029)
Bandwidth	± 4	± 10	± 4.24	± 4.24
Observations	257	522	277	277
R-squared	0.379	0.348	0.443	0.443
Comparison mean	0.097	0.087	0.098	0.098
Spline	Linear	Quadratic	Linear	Linear
Stratum fixed effects	Yes	Yes	Yes	Yes
Additional controls	Yes	Yes	Yes	Yes

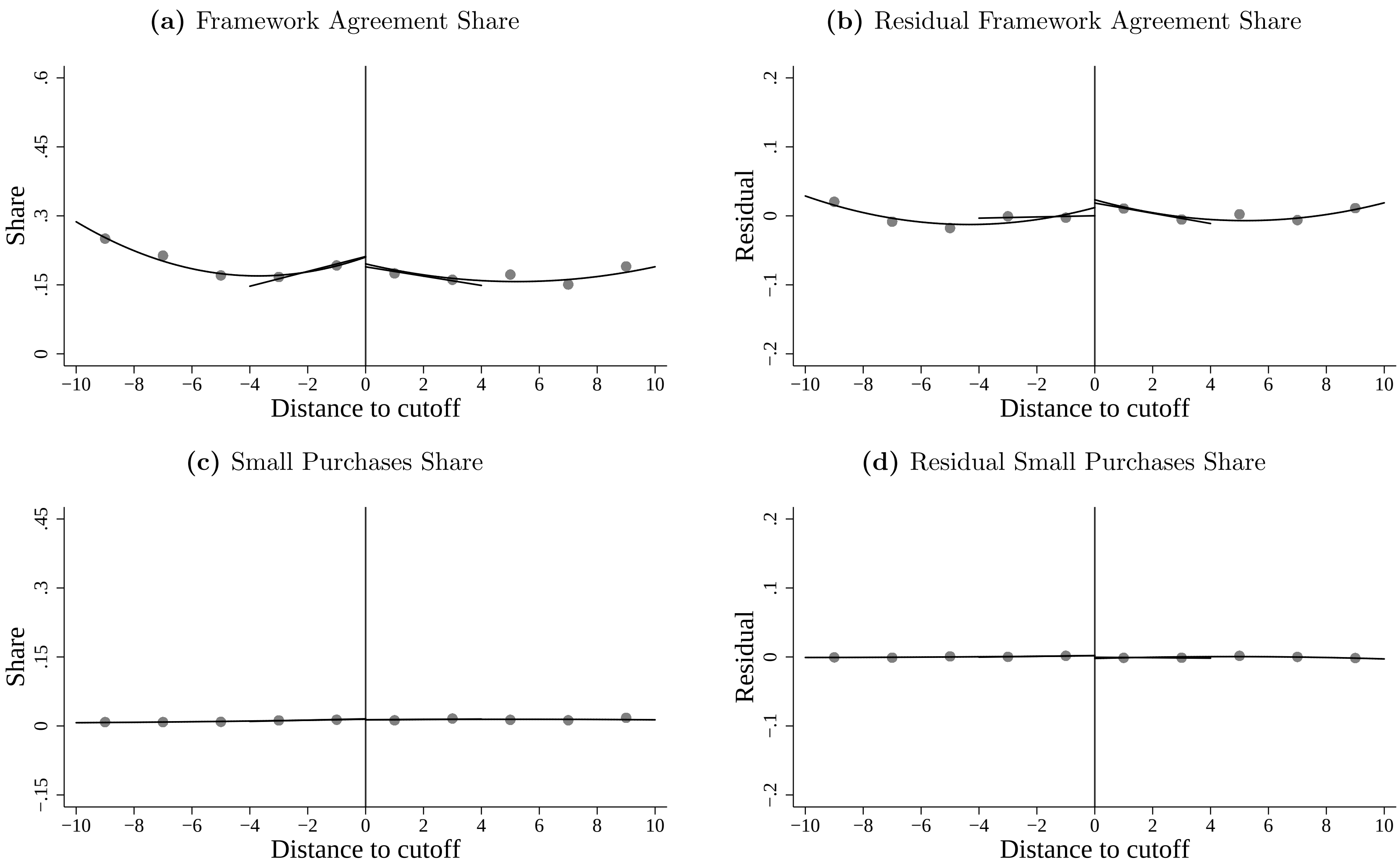
Notes: Reduced form RDD estimates following the specification of Equation (4). Standard errors are clustered at the stratum level. A stratum refers to a cell defined by year, responsible control department and type of entity. The control variables include political affiliation, log of total amount purchased +1 one and two years prior, and the outcome measured one and two years prior. Columns (3) and (4) employ the mean-squared-error-optimal bandwidth following Imbens and Kalyanaraman (2012). Column (4) in addition reports bias-corrected estimates and robust standard errors following Calonico, Cattaneo and Titiunik (2014). ***p<0.01, **p<0.05, *p<0.1.

Figure A1:
Audit Probability (First Stage) by Year



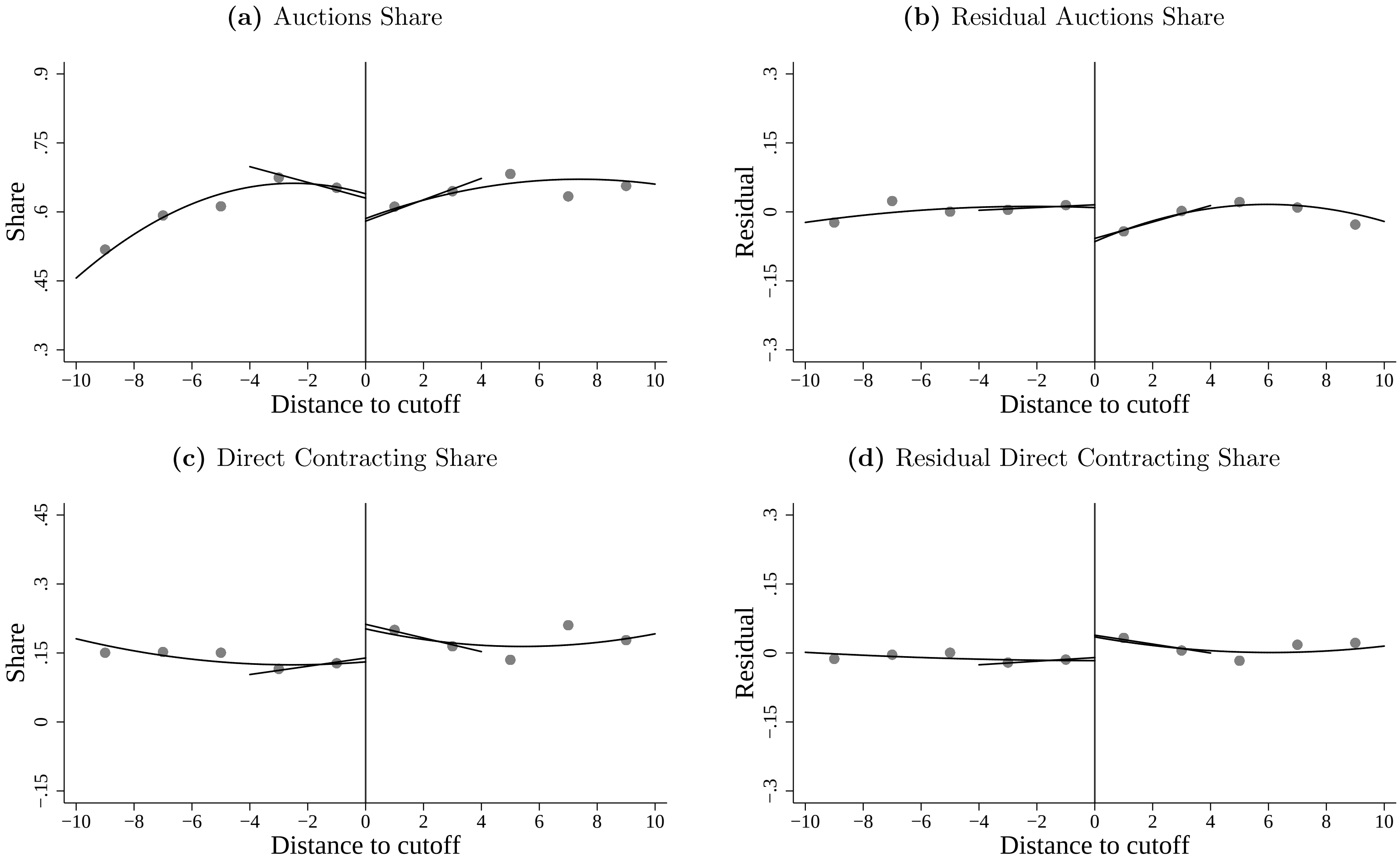
Notes: The figures show the audit probability for entities with medium level of risk in the ±10 range of the relative importance score. In panels *(a)* and *(c)*, the dots represent audit probabilities within 2-point-wide intervals of the relative importance score. In panels *(b)* and *(d)*, the dots represent averaged residuals. The residuals are obtained from a regression of the dummy for having been audited in the year corresponding to the score on stratum fixed effects and control variables. Control variables include political affiliation, log of total amount purchased +1 one and two years prior, and a dummy for having been audited in the preceding year (audits data are not available for two years earlier). Relative importance scores are normalized by stratum-level cutoff. A stratum refers to a cell defined by each combination of responsible control department and type of entity. Solid lines show linear and quadratic fits.

Figure A2:
Share of Purchases through Framework Agreements and as Small Purchases



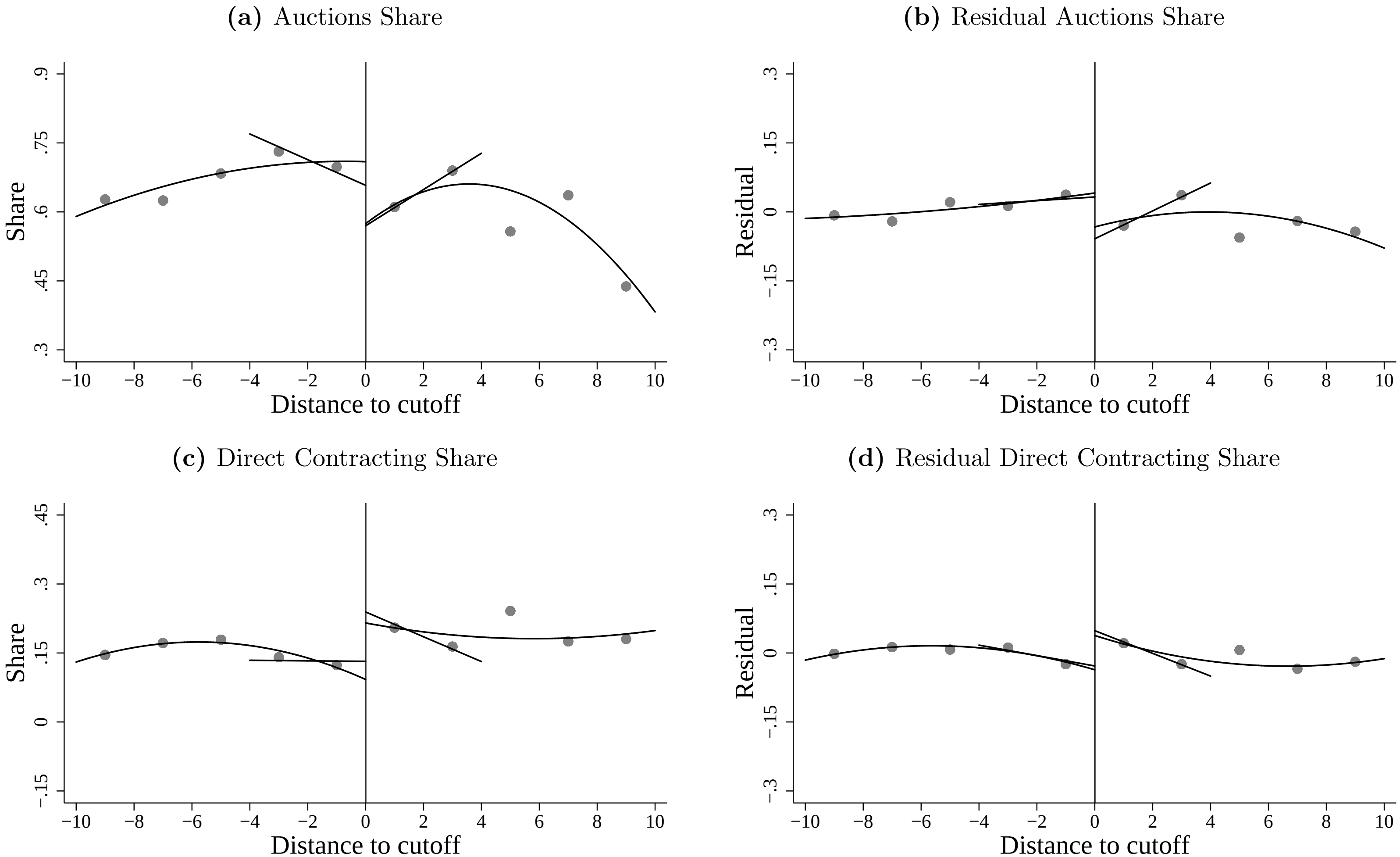
Notes: The figures show the amount purchased through framework agreement and the amount of small purchases as shares of total purchases, for entities with medium level of risk in the ±10 range of the relative importance score (reduced form). In panels *(a)* and *(c)*, the dots represent modality shares averaged within 2-point-wide intervals of the relative importance score. In panels *(b)* and *(d)*, the dots represent averaged residuals. The residuals are obtained from a regression of the outcome (amount purchased through a given modality over total amount purchased) in the year corresponding to the score on stratum fixed effects and control variables. Control variables include political affiliation, log of total amount purchased +1 one and two years prior, and the outcome measured one and two years prior. The relative importance score for each entity is normalized by the stratum-level cutoff. A stratum refers to a cell defined by year, responsible control department and type of entity. Solid lines show linear and quadratic fits.

Figure A3:
Share of Purchases through Auctions and Direct Contracting in 2011



Notes: The figures show the amounts purchased through auctions and direct contracting as shares of total purchases, for entities with medium level of risk in the ±10 range of the relative importance score. In panels *(a)* and *(c)*, the dots represent modality shares averaged within 2-point-wide intervals of the relative importance score. In panels *(b)* and *(d)*, the dots represent averaged residuals. The residuals are obtained from a regression of the outcome (amount purchased through a given modality over total amount purchased) in the year corresponding to the score on stratum fixed effects and control variables. Control variables include political affiliation, log of total amount purchased +1 one and two years prior, and the outcome measured one and two years prior. The relative importance score for each entity is normalized by the stratum-level cutoff. A stratum refers to a cell defined by each combination of responsible control department and type of entity. Solid lines show linear and quadratic fits.

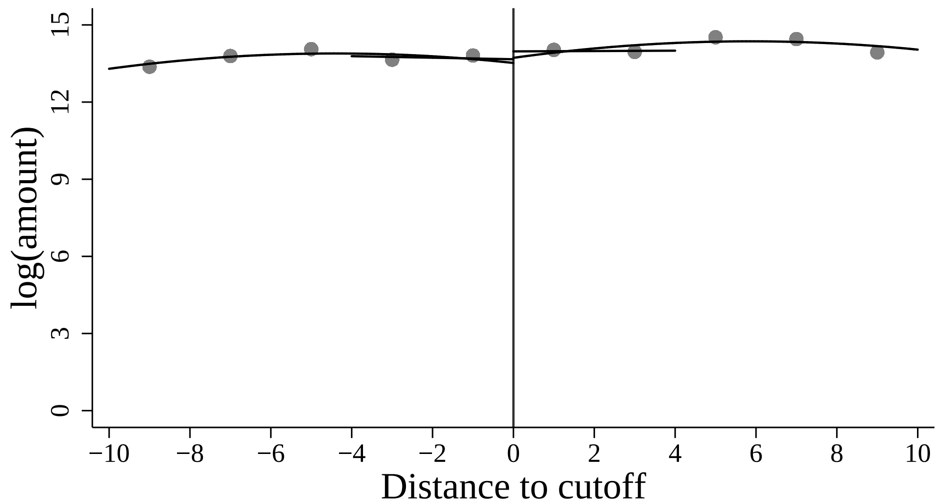
Figure A4:
Share of Purchases through Auctions and Direct Contracting in 2012



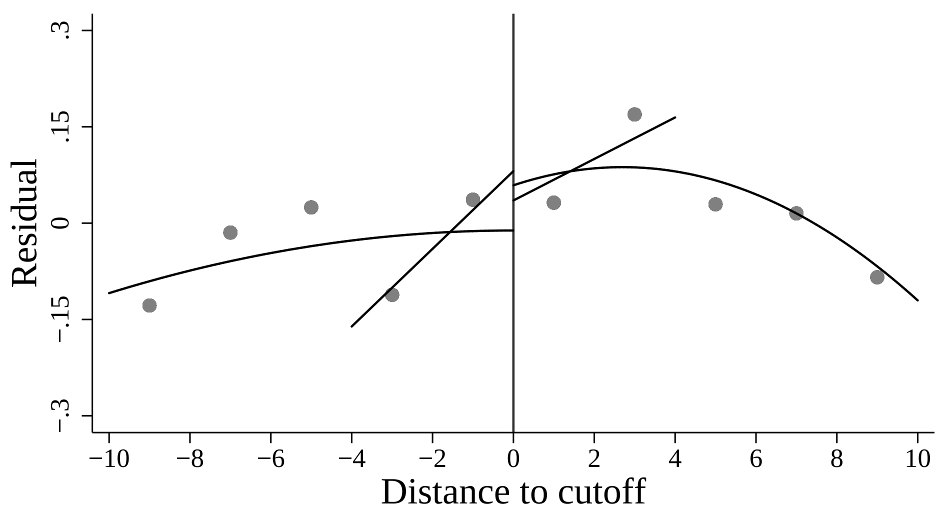
Notes: The figures show the amounts purchased through auctions and direct contracting as shares of total purchases, for entities with medium level of risk in the ±10 range of the relative importance score. In panels *(a)* and *(c)*, the dots represent modality shares averaged within 2-point-wide intervals of the relative importance score. In panels *(b)* and *(d)*, the dots represent averaged residuals. The residuals are obtained from a regression of the outcome (amount purchased through a given modality over total amount purchased) in the year corresponding to the score on stratum fixed effects and control variables. Control variables include political affiliation, log of total amount purchased +1 one and two years prior, and the outcome measured one and two years prior. The relative importance score for each entity is normalized by the stratum-level cutoff. A stratum refers to a cell defined by each combination of responsible control department and type of entity. Solid lines show linear and quadratic fits.

Figure A5:
Total amount purchased

(a) Total Amount



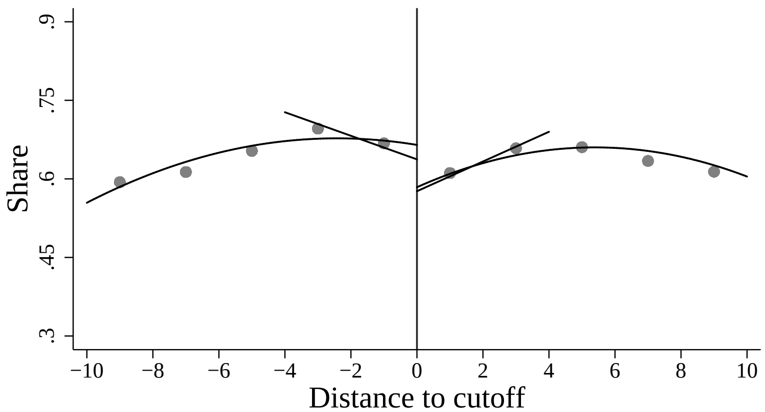
(b) Residual Total Amount



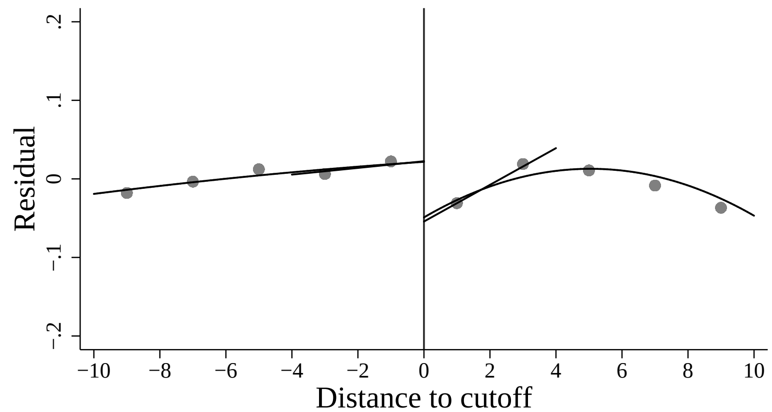
Notes: The figures show the total amount purchased in logarithm, for entities with medium level of risk in the ±10 range of the relative importance score. The relative importance score for each entity is normalized by the stratum-level cutoff. A stratum refers to a cell defined by year, responsible control department and type of entity. In panel *(a)*, the dots represent modality shares averaged within 2-point-wide intervals of the relative importance score. In panel *(b)*, the dots represent averaged residuals. The residuals are obtained from a regression of the Outcome (total amount purchased) in the year corresponding to the score on stratum fixed effects and control variables. Control variables include political affiliation, log of total amount purchased +1 one and two years prior and the outcome measured one and two years prior. Solid lines show linear and quadratic fits.

Figure A6:
Share of Purchases through Auctions and Direct Contracting
Interacting the Running Variable with Stratum Dummies

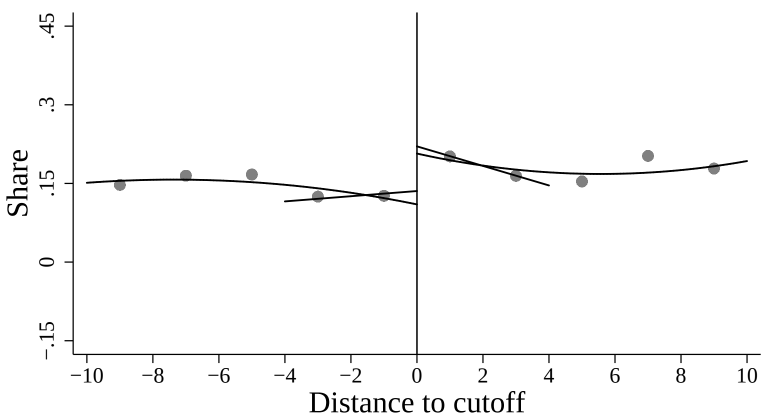
(a) Auctions Share



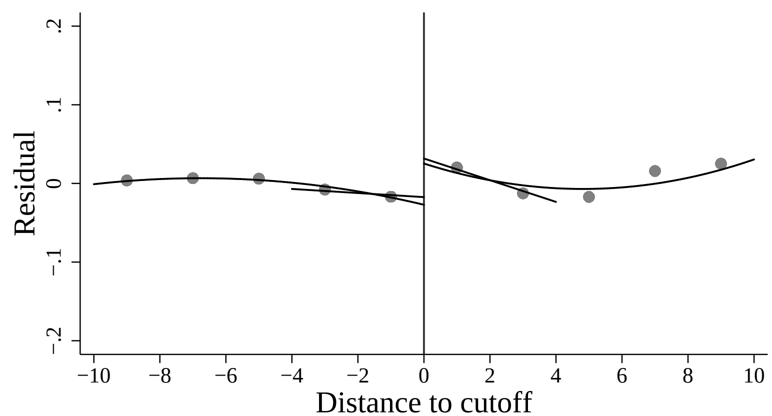
(b) Residual Auctions Share



(c) Direct Cntracting Share

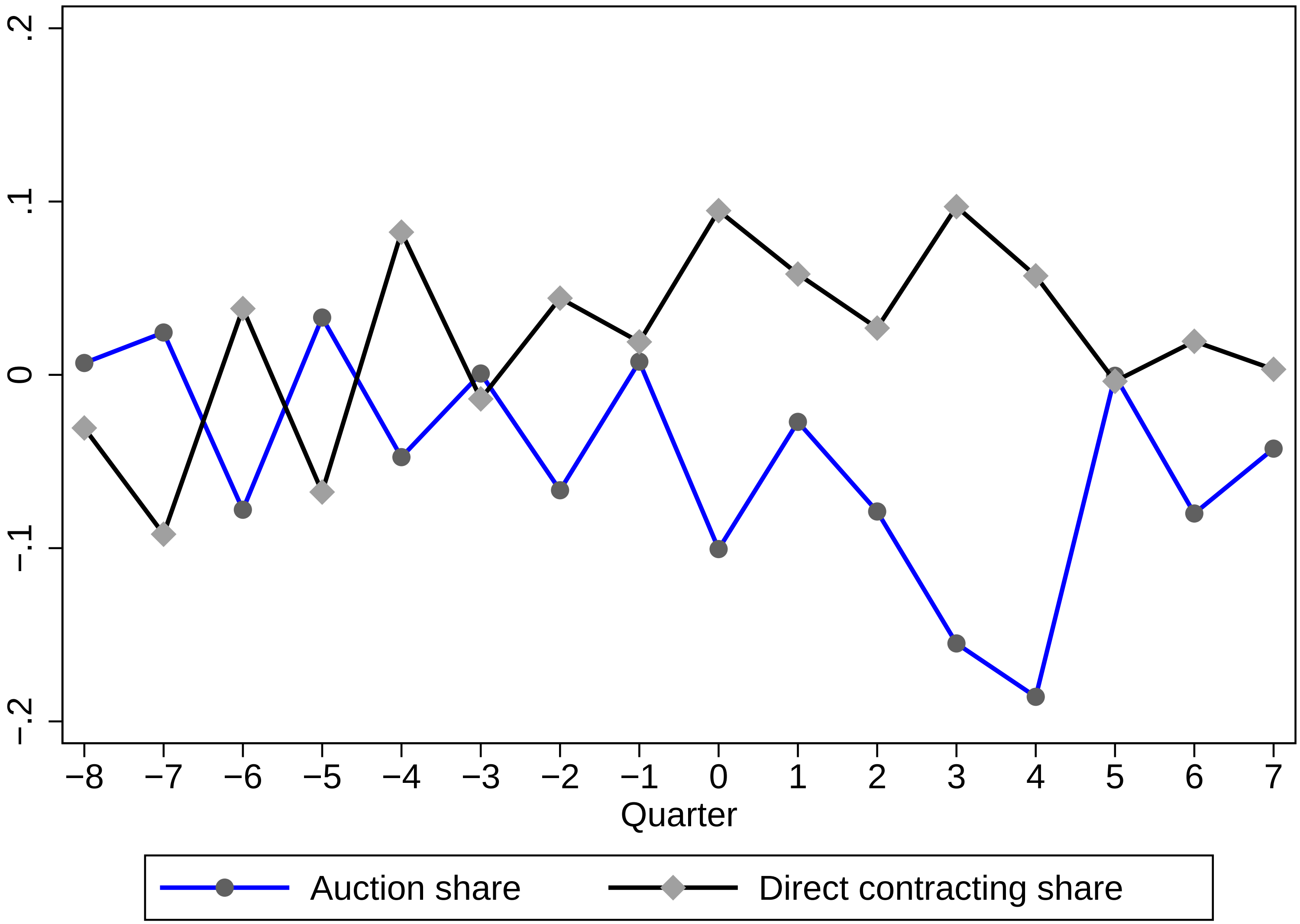


(d) Residual Direct Contracting Share



Notes: The figures show the amounts purchased through auctions and direct contracting as shares of total purchases, for entities with medium level of risk in the ±10 range of the relative importance score (reduced form). In panels *(a)* and *(c)*, the dots represent modality shares averaged within 2-point-wide intervals of the relative importance score. In panels *(b)* and *(d)*, the dots represent averaged residuals. The residuals are obtained from a regression of the outcome (amount purchased through a given modality over total amount purchased) in the year corresponding to the score on stratum fixed effects, and their interaction with running variable, and control variables. Control variables include political affiliation, log of total amount purchased one and two years prior, and the outcome measured one and two years prior. The relative importance score for each entity is normalized by the stratum-level cutoff. A stratum refers to a cell defined by each combination of responsible control department and type of entity. Solid lines show linear and quadratic fits.

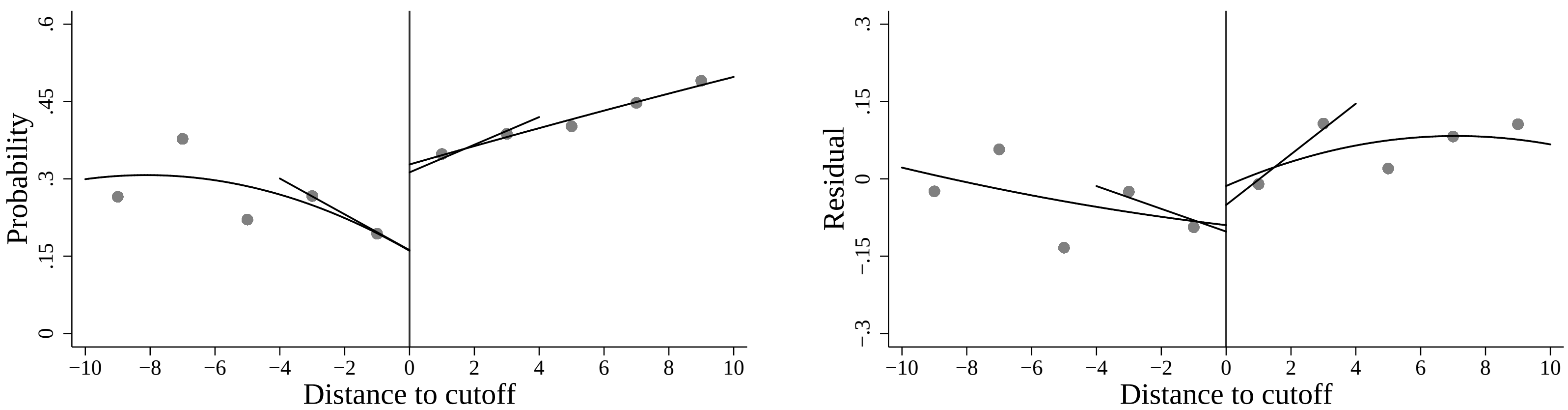
Figure A7: Treatment Effects on Purchases through Auctions and Direct Contracting over Time



Notes: The lines show the estimates for the impact on the share of purchases made through auctions and direct contracting by quarter. Coefficients plotted correspond to the bias-corrected estimates using the MSE-optimal bandwidth. Zero corresponds to the first year.

Figure A8:
Audit Probability (First Stage) in Subsequent Year

(a) Audit Probability in *t+1* **(b)** Residual Audit Probability in *t+1*



Notes: The figures show the audit probability in the year after treatment. In panel *(a)*, the dots represent audit probabilities averaged within 2-point-wide intervals of the relative importance score. In panel *(b)*, the dots represent averaged residuals. The residuals are obtained from a regression of the outcome on stratum fixed effects and control variables. Control variables include political affiliation, log of total amount purchased one and two years prior, and a dummy for having been audited in the year prior to treatment year (audits data is not available for two prior to treatment). Analysis for entities with medium level of risk in the ±10 range of the relative importance score. The relative importance score for each entity is normalized by the stratum-level cutoff. A stratum refers to a cell defined by year, responsible control department and type of entity. Solid lines show linear and quadratic fits.

B Specification Bias with a Single Binary Subgroup Interaction Term

This Appendix shows that the simple treatment-subgroup-interaction approach is not generally valid in the RDD setting, unless the relationship between the outcome and the running variable is the same across subgroups. And even when the model allows for separate slopes by subgroup (i.e. full interaction of the running variable polynomial with the subgroup indicator), the problem remains that other characteristics may vary systematically across subgroups.

When trying to estimate the differential impact on two subgroups, $G_i = 0$ and $G_i = 1$, there are now two estimands of interest, corresponding to the RD-gaps in Y at the cutoff in each of the two subgroups:

$$\lim_{X_i \downarrow 0} \mathrm{E}[Y_i|X_i, G_i = 0] - \lim_{X_i \uparrow 0} \mathrm{E}[Y_i|X_i, G_i = 0] = \alpha_{R0} - \alpha_{L0} \tag{C1}$$

and

$$\lim_{X_i \downarrow 0} \mathrm{E}[Y_i|X_i, G_i = 1] - \lim_{X_i \uparrow 0} \mathrm{E}[Y_i|X_i, G_i = 1] = \alpha_{R1} - \alpha_{L1} \tag{C2}$$

Now consider a linear spline specification, augmented only with the subgroup dummy G_i and an interaction term of the subgroup dummy with treatment assignment $Z_i \times G_i$, where $Z_i = \mathrm{I}[X_i \geq 0]$:

$$\begin{aligned} Y_i = \alpha_{L0} + (\alpha_{R0} - \alpha_{L0})Z_i + (\alpha_{L1} - \alpha_{L0})G_i + [(\alpha_{R1} - \alpha_{L1}) - (\alpha_{R0} - \alpha_{L0})]Z_i \times G_i \\ + \beta_1 X_i + \beta_2 X_i \times Z_i + U_i \end{aligned} \tag{C3}$$

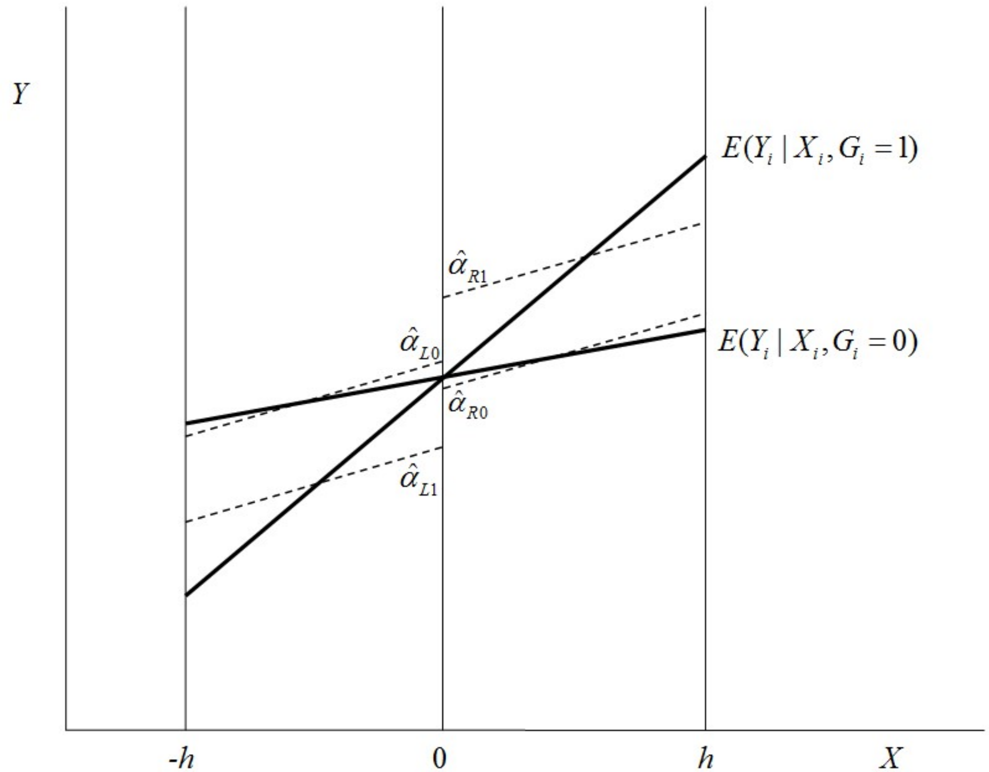
To see the correspondence between the regression specification (C3) and the parameters of interest (the αs), simply evaluate the predicted value of the estimated regression function at a given point. For example, when X, Z, and G are all zero, the predicted value is the estimated mean of Y in the $G_i = 0$ subgroup just before crossing the cutoff, i.e. $\hat{\alpha}_{L0}$. Just above the cutoff, the estimated mean of Y in the $G_i = 0$ subgroup is $\hat{\alpha}_{L0} + (\hat{\alpha}_{R0} - \hat{\alpha}_{L0}) = \hat{\alpha}_{R0}$.

Figure C1 illustrates the specification bias that is introduced when both true RD gaps are zero ($\alpha_{R0} = \alpha_{L0} = \alpha_{R1} = \alpha_{L1}$, i.e. the treatment has no effect in either subgroup) yet the relationship between Y and X is not the same in the two subgroups (i.e. different slope in each subgroup). The two solid lines show the linear approximations to the conditional expectation functions in the two subgroups, $\mathrm{E}[Y_i|X_i, G_i = 0]$ and $\mathrm{E}[Y_i|X_i, G_i = 1]$ within a neighborhood h around the cutoff. The dashed lines represent the slope estimates from equation (C3), which are allowed to differ to the left and to the right of the cutoff but are assumed constant across subgroups. Crucially, the slope estimates are necessarily between the true slope parameters because OLS tries to minimize deviations from the regression line across subgroups. Now as long as the slope estimates are biased, the intercept and discontinuity estimates at the cutoff are necessarily biased as well. In Figure C1, the discontinuity estimate for the $G_i = 1$ subgroup is upward biased and for the $G_i = 0$ subgroup the discontinuity estimate is downward biased. This specification bias is easily fixed if the model allows for separate slopes by subgroup (i.e. a linear spline fully interacted with G_i):

$$\begin{aligned} Y_i = \alpha_{L0} + (\alpha_{R0} - \alpha_{L0})Z_i + (\alpha_{L1} - \alpha_{L0})G_i + [(\alpha_{R1} - \alpha_{L1}) - (\alpha_{R0} - \alpha_{L0})]Z_i \times G_i \\ + \beta_1 X_i + \beta_2 X_i \times Z_i + \beta_3 X_i \times G_i + \beta_4 X_i \times Z_i \times G_i + U_i \end{aligned} \quad \text{(C4)}$$

The problem remains, however, that other characteristics may vary systematically across subgroups, which makes it difficult to interpret differential subgroup impacts.

Figure C1:
Specification bias



Notes: This figure illustrates the specification bias that arises when the true RD gaps in both subgroups are zero and the slope parameters of the running variable and the outcome are different across subgroups, yet the model imposes that the slope parameters are the same.

C Justifications for Direct Purchases

There are 21 justifications for direct purchases according to Chilean law number 19,886:

1. Acquisition originated in an unsuccessful private auction, where no bidders showed up (point 1)
2. Remnant of unfinished contract due to breach of contract from original supplier or other causes, when remnant does not exceed 70,000 USD (point 2)
3. Emergency or unforeseen urgency (point 3)
4. There is only one supplier of goods or services (point 4)
5. Contracting with foreign legal persons outside the national territory (point 5)
6. Confidential services (point 6)
7. Extension of utility contracts or ancillary services (point 7.a)
8. Procurement under representation expenses (point 7.b)
9. Hiring a specific supplier for the safety and integrity of authorities (point 7.c)
10. Hiring of consultancy services, considering the special qualities of the provider (point 7.d)
11. Contracting with holders of intellectual or industrial property (point 7.e)
12. Trust and security of providers, derived from their experience (point 7.f)
13. Replacing or complementing compatible accessories for models already acquired (point 7.g)
14. Public awareness of the tender could jeopardize the goal of the contract (point 7.h)
15. Procurement of goods to foreign suppliers for use outside the country (point 7.i)
16. Purchase is below 7,000 USD and cost of evaluating bids is disproportionate (point 7.j)
17. Goods or services for teaching or research projects (point 7.k)
18. Public auction without prior public tender offers, or inadmissible deals (point 7.l)
19. Specialized services below 70,000 USD as specified by article 107 of the present law [law number 19,886] (point 7.m)
20. Acquisition of less than or equal to 700 USD (point 8)
21. Exceptions for being required to operate in the online procurement system apply as specified in article 62 [of law number 19,886]