# Past Performance and Procurement Outcomes

Pasha Andreyanov, Francesco Decarolis,
Riccardo Pacini and Giancarlo Spagnolo*

July 31, 2023

**Abstract**

In procurement, balancing costs with contract execution performance is key. Our study evaluates the implementation of a vendor rating system and a transition from price-only to scoring auctions, where the price is weighed against past performance. Results indicate a significant performance jump – from 25% to 90% – across all audited parameters. These improvements were mirrored by buyer output quality but not the price. We attribute this anomaly to the interplay of cost and quality considerations and firms' heterogeneity across these two dimensions. A structural model is estimated to quantify the difference in the outcomes of the price-only and scoring auctions.

JEL: H57, D47, K12

Keywords: Public procurement, past performance, audit, vendor rating, public utility

*Andreyanov: HSE University, Moscow, *pandreyanov@gmail.com*; Decarolis: Bocconi University, *francesco.decarolis@unibocconi.it*; Pacini: Agenzia del Demanio, *riccardo.pacini01@agenziademanio.it*; Spagnolo: University of Rome Tor Vergata, EIEF & SITE, *spagnologianca@gmail.com*; We are grateful to Nick Bloom, Luis Cabral, Liran Einav, Rosa Ferrer, Stefano Gagliarducci, Ricard Gil, Luigi Guiso, Hugo Hopenhayn, Neale Mahoney, Rob Metcalfe, Enrico Moretti, Gustavo Piga, Andrea Prat, Piero Sabbatini, Stephane Saussier, Steve Tadelis, Paola Valbonesi and to the participants at the seminars at Barcelona GSE, Berkeley, FGV, Stanford University, Stockholm University, University of Bologna and the University of Paris I for their useful suggestions. We thank Acea for sharing its data. We also acknowledge financial support from both the European Research Council (ERC Starting Grant 679217, Decarolis) and the Wallander and Hedelius Foundation (Handelsbanken P2013-0162, Spagnolo).

# Contents



# 1 Introduction

To mitigate corruption, global legislators have prioritized the use of open auctions - ensuring equal treatment for bidders - even when there's significant variance in suppliers' past performances for analogous contracts. However, competitive auctions often fail in procurement scenarios characterized by incomplete contracts. Intensifying price competition might

compromise ex-post performance.[1] Thus, striking a balance between costs and contract execution performance is a recurring challenge in public procurement.

In the private sector, this issue can be addressed through the use of discretion. In the public sector, where auctions are often mandatory, a buyer could implement such a strategy using a *scoring auction* that weighs the supplier's bid against its past performance.[2] In essence, suppliers are evaluated based on attributes directly influencing or associated with their work quality in past contracts. We refer to this as a *vendor rating* system. A higher rating allows the firm to win contracts more often or at a more favorable price in the scoring auction. Such a system gives an advantage to firms with superior organizational frameworks, commendable work ethics, and established reputations and motivates underperforming suppliers to elevate their standards.

While giving weight to past performances seems inherently effective and is commonly employed across many markets, especially in the digital ones, the assessment of using a scoring auction with vendor rating is notably absent from the academic literature. Our study seeks to bridge this gap by empirically analyzing the adoption and the dismissal of such a system by a prominent utility provider in Italy, Acea. Catering to a substantial region in central Italy, including the capital Rome, Acea's objective was to elevate the performance standards in the construction tasks associated with maintaining and enhancing the electrical grid. Prior to implementing the system, contracts were allocated solely via the standard first-price, sealed bid auctions.

In October 2007, Acea's auditors started carrying out random worksite audits, meticulously examining the relevant parameters of both quality of the job done and adherence to the needed safety standards in execution. Two months later, Acea informed suppliers that these evaluations would form a *reputation index* (RI). This index would subsequently become a component of a linear scoring rule, attributing 75 percent of the score to pricing and the remaining 25 percent to the RI.[3] The exclusive price-centric auctions persisted for

[1] For instance, Spulber (1990) indicates that open competition in the construction sector can instigate adverse selection and ex-post opportunism among contractors. For a contemporary perspective, refer to Lopomo, Persico and Villa (2023) and additional references detailed in the literature review below.

[2] Scoring auctions, common in procurement, have garnered significant scholarly attention, as evident in works like Che (1993); Asker and Cantillon (2008, 2010); Andreyanov (2019); Laffont et al. (2020); Camboni et al. (2023).

[3] It's pertinent to note that Acea's choice of the term "reputation" for its vendor rating system refers

another 37 months, granting firms ample opportunity to optimize their RIs before transitioning to scoring auctions in January 2011.[4] However, in a twist of events, Acea shelved the system in May 2011 due to mounting pressures from suppliers, seemingly discontented with the approach, and ensuing legal contentions over the perceived discriminatory nature of the scoring auction. The auction process recommenced in May 2012, but under a revamped system that amalgamated traditional price-based auctions with rigorous compliance checks through audits.

A unique feature of this environment is that it allows us to observe the evolution of the reputation index between the announcement and the enforcement of the scoring auction, separately from the strategic adjustment of bids at the moment of the switch. Together with the availability of a control group, this provides us with a credible empirical strategy to identify the causal effects of the reform on a firm's decisions involving entry, bidding, and performance.

Our analysis combines several data sources. First, complete audit data on the parameters measuring quality and safety standards in suppliers' contract execution are observed for 10 years: from the introduction of the new audits in 2007, up until 2017. We refer to these as the *internal measures* of performance. Second, we observe the procurement auctions held between 2004 and 2017 by Acea and every other utility company in the country. Finally, from the public regulatory authority, we obtained *external measures* about the output quality for both the buyer and comparable utility companies.

We begin the analysis with the events between the first announcement of vendor rating and the switch to the new auction mechanism, to which we refer as the *grace period*. Through this period, compliance in the parameters audited monotonically increased from 25 percent to more than 80 percent just before the switch. We also find that essentially all active suppliers improved their compliance in similar ways, and they did so strategically, with compliance increasing relatively more for those parameters with higher weights in the computation of the reputation index, see Figures 2 and 3.

It is possible, of course, for the firms to improve compliance with the observed parameters,

strictly to observable, audited parameters. This definition diverges from the traditional economic context where reputation denotes a belief regarding a player's nature Bar-Isaac and Tadelis (2008).

[4]Before initiating a full series of scoring auctions, two such auctions took place on 5 August 2010.

but not performance itself. To counter this argument, we study external measures since they are not part of the vendor rating system. By comparing Acea to a control group of utility companies, we find a significant drop in the frequency, duration, and unpredictability of blackouts in the electricity distribution service related to Acea's contracts. Interestingly, the quality of the water distribution service provided by Acea, but not related to the electricity grid, did not improve or worsen.

The evolution of prices during the grace period is more complex, so we split it into two phases. In the first phase, while RI rapidly grows, we find that prices decrease, which can be explained by the temporary stimulus to win more auctions to get a chance to be audited. In the second phase, when RI was satiated, we found that prices rebounded to roughly the same level as before the grace period. This makes sense because the costs of acquiring RI were effectively sunk.

Proceeding to the events that took place once the scoring auction was implemented, to which we refer as the *scoring period*, one could expect that the firms, especially those that have raised their RI the most, would take advantage of the new auction mechanism by substantially increasing their bids and that the prices would finally increase. On the contrary, we observe that the prices have declined sharply. Figure 6 shows that the discounts (i.e., the percentage difference between the price offered and the reserve price set by the buyer) have jumped roughly from 10 to 25 percent. At the same time, the pool and number of participating firms and their reputation index remained stable in the neighborhood of the switch, see Figure 7.[5]

In a nutshell, the combination of the scoring auction with the vendor rating appears to have produced a large increase in performance at zero or even negative cost for the buyer. Moreover, after the scoring period and till the end of our records, the prices remained relatively low, see Figure 5, while RI kept improving. This is an empirical puzzle.

The intuitive impossibility of increasing performance at no cost has deep roots in auction theory. One could argue that, through the revenue equivalence principle, firms should enjoy the same expected utility across different auction formats; thus, higher auction profits should compensate for the investment costs. However, this argument relies on a seemingly harmless

[5]The reputation index is continuous in time because it is a rolling average of weighted audit results.

assumption that firms only vary in one dimension. We show that the revenue equivalence principle no longer holds if firms vary in two dimensions: cost-efficiency and *quality*. This opens a possibility of transfer of informational rents from the firms to the buyer, potentially explaining the puzzle.

There are two main factors at play here. Recall that the scoring auction puts the low-quality/low-cost firms at a disadvantage so that high-quality/high-cost firms can win more often. Thus, the first factor is the ranking of firms by the auction mechanism, contributing positively to the observed winner's quality and price, as common sense would predict. The second factor is more subtle. Note that, by the nature of the scoring auction, the high-cost-high-quality firms are pitted against the low-cost-low-quality ones. If quality and costs are positively correlated, firms may find themselves less differentiated in the score dimension than in the price dimension. Thus, bidding will be more aggressive in a scoring auction than in a price-only auction, contributing negatively to the price.

This reminds of two tales in economics: Bertrand supertraps, see Cabral and Villas-Boas (2005) and the Giffen Paradox, simultaneously. The two factors can be thought of as a direct effect on the price, that is, holding the firm's actions fixed, and the strategic effect. Alternatively, they can be thought of as the substitution effect since cost-reduction is substituted for quality and the income effect. To explain the puzzle, the two effects have to work against each other, and the second (strategic, income) effect has to outweigh the first (direct, substitution) effect.

We illustrate our explanation using a model of a scoring auction. The model differs from those of Che (1993) and Asker and Cantillon (2010) in two crucial elements: the costs of building up RI (to which we refer as quality) are sunk, and the types are bivariate. These are both key features of the environment that we analyze, and they reasonably capture a broad range of settings that are relevant in procurement. As a proof of concept, to illustrate the nuanced predictions of the model, we show an example where quality and discounts are negatively correlated, and both increase when switching from the price-only to the scoring auction, see Figure 9. This observed negative correlation underscores the presence of quality concerns or adverse selection within the setting. This feature is a crucial component in the theoretical discourse surrounding adverse selection in contract procurement, as explored by

Manelli and Vincent (1995) and Lopomo, Persico and Villa (2023). Furthermore, our data reveal traces of this correlation, though predominantly among the most successful bidders.

The estimates of the model reinforce our previous findings. Indeed, the model predicts that the switch to the scoring auction is associated with a decrease in the winning firm's bid shading by roughly 12 percent (€1.1 thousand per auction), on average, offsetting a significant portion of the increase in the winning firm's costs (€3.1 thousand per auction) due to selection. This means lower-than-expected prices for the buyer and, at the same time, lower-than-expected profits for the firms. [6]

To summarize, the vendor rating system, in conjunction with the scoring auction, appears to have been a very successful experiment, but mostly for the buyer. In other words, it was not a Pareto improvement, which likely contributed to the abortion of this experiment, despite its great merits. Indeed, we also include in our assessment of the reform a simple cost-effectiveness analysis revealing the remarkable improvements in safety and quality. In particular, we use external measures to estimate the consumption value of some of the observed increases in quality and safety. This is achieved by mapping quality into a measure of the buyer output (blackout duration) and safety to the probability of deadly accidents. We estimate that a lower bound for the social benefit from increased contract compliance ranges between €3.5 and €5.3 million per year regarding safety improvements. At the same time, the reform implies a social benefit of at least €6.6 million for the reduction in blackouts.

## 1.1 Literature

This study offers the first in-depth analysis of introducing such a vendor rating system in public procurement, i.e., scoring auctions with past performance.

At the most general level, this study is related to the law and economics literature on agency problems. Shavell (1984), and the following research line, modeled the theoretical question of whether ex-ante or ex-post interventions are more effective tools for dealing with a firm engaging in potentially risky behaviors and having private information about the

[6] The structural model shows a smaller effect than Figure 6 suggests. This is likely due to the large number of firms (14 on average) formally participating in every auction, which makes the estimates of bid shading very small. The effect would be larger if real competition happened among the strongest few firms.

extent of potential hazards.[7] Acea, with its dominant position as the largest buyer in the market, is akin to a regulator that decides to bolster the role of ex-post incentives to curb risky behaviors by its suppliers.[8]

Our paper also relates to the literature on industrial organization studies of auctions and competition. In particular, the need to use more complex auction systems relative to the standard price-only auctions is a key pillar of the literature on bidding for contracts. A vast theoretical literature has highlighted the limits of competitive auctions in this type of setting, starting at least from Spulber (1990) and including Manelli and Vincent (1995), Zheng (2001), Bajari and Tadelis (2001), D'Alpaos et al. (2013) and Burguet, Ganuza and Hauk (2012). This theoretical insight has found support in a handful of empirical studies, including Carril, Gonzalez-Lira and Walker (2020), Decarolis et al. (2020) and Bosio et al. (2020). These findings have contributed to the proposal of new auction formats to trade off price vs. quality, such as those in Andreyanov, Krasikov and Suzdaltsev (2023) when quality is contractible and in Lopomo, Persico and Villa (2023) when it is not.

Several more recent empirical studies have confirmed this result highlighting how price competition can backfire in various contract performance measures ranging from quality to cost overruns and time delays.[9] Compared to this literature, our emphasis is on the use of past performance, which is novel.[10] Our findings are also related to a recent wave of studies highlighting the importance of considering dynamic incentives to understand procurement auctions. In this respect, this study is close in spirit to those of Jofre-Bonet and Pesendorfer (2003), Marion (2017), and Chassang and Ortner (2016). Our theoretical model resonates with the seminal works by Che (1993), Asker and Cantillon (2008), and Asker and Cantillon

[7] A large body of subsequent studies has extended this original result and explored applications ranging from environmental protection to banking. See, among others, Kolstad, Ulen and Johnson (1990), Rose-Ackerman (1991), Saussier (2000), Hiriart, Martimort and Pouyet (2004) and Beuve and Saussier (2021).

[8] The success of this strategy likely hinges on the fact that the enforcement of ex-ante (i.e., contract) clauses through penalties is limited by the well know inefficacy of the Italian civil court system. See Djankov et al. (2008) for a cross-country study, Giacomelli and Menon (2016), and Coviello et al. (2018) for Italy.

[9] See Bajari, McMillan and Tadelis (2009), Decarolis (2014), Chong, Staropoli and Yvrande-Billon (2014), Liebman and Mahoney (2016), Lewis-Faupel et al. (2016), Kang and Miller (2021), and Carril, Gonzalez-Lira and Walker (2022).

[10] A few theoretical studies have argued in favor of the positive role that reputation mechanisms linking the award of future contracts to the quality of past performance may improve contract performance in repeated public procurement under imperfect contracting. See, among others, Calzolari and Spagnolo (2009), Board (2011) and Andrews and Barron (2016).

(2010) on the design of scoring auctions, but with the focus on the investment costs.

The last strand of the literature is the design and use of contract audit measures. Detailed audit data on public procurement are used by Olken (2007) on Indonesia and Colonnelli and Prem (2021) on Brazil, as well by Duflo et al. (2013, 2018) on environmental compliance. The mechanism that we study is based on third-party audits of past performance. Hence it is also closer to the recent literature on the design of feedback mechanisms in platforms (Tadelis, 2016) than to the classic literature on reputation as an incentive to work hard to affect beliefs (Klein and Leffler, 1981; Holmstrom, 1999). Still, our findings square well with the argument in List (2006) that reputation and quality verification are complements in that repeated interaction only increases the price/quality correlation when a quality rating system is present.

# 2 Institutional Details

The context of the analysis is that of a multi-utility company, Acea s.p.a., offering electricity and water services to about 1.6 million customers: private households and business establishments in Rome. The firm is vertically integrated, owning and operating most of its generation, transmission, and distribution systems. From this point of view, it is very similar to some of the largest US power operators, such as the Los Angeles Department of Water and Power (LADWP), ComEd (Chicago), BGE (Baltimore), and PECO (Philadelphia). As shown in Table 1, all of these firms spend significant resources every year on works to preserve their power grid's operational efficiency.

In 2015, Acea spent about US $200 million on procuring the kind of works which are the focus of this study. The jobs typically entail maintaining, upgrading, and replacing transformers, poles, underground cables, underground vaults, station transformers, and distribution and receiving stations. These are all works exposing workers to safety hazards linked to electricity-induced accidents. In 2007, after these risks materialized in some deadly accidents, Acea decided to take action to improve contract execution by revising its audit system. Until then, the auditors (i.e., a team of Acea engineers) inspecting the work sites used to prepare a written memo describing the state of the work site. On October 16, 2007,

Acea's engineers conducted their first audit with the new auditing system, which streamlined and digitized the process: using tablet computers, the inspection required evaluating a fixed list of 136 parameters by scoring them as pass, fail, or uninspectable.

The reform exclusively involved the electricity sector, leaving out the water sector and the contracts related mostly to maintaining the public illumination (IP) and electricity distribution (DIST) systems.[11] The list of 136 parameters was identified as exhaustive of the quality and safety standards that needed to be audited: they ranged from the types of materials and machinery used to the adherence to the worksite safety specifications and legal status of all workers (the full list is reported in Appendix F). The logic followed by Acea was to cover with these 136 parameters all of the relevant features of contract performance.

A few institutional details are important for our analysis and worth emphasizing:

1. In each audit, auditors evaluate all of those parameters which it is feasible to evaluate due to the type of job and the condition of the worksite: on average, 34 parameters are evaluated per auditor visit. Each evaluated parameter $p$ is given a value of 0 or 1. We will denote the corresponding value as $v_{ap}$.

2. Each parameter is assigned a weight $w_p$ between 1 and 10 when it is being evaluated and 0 otherwise, in which case it is absent from the data. These weights are constant across audits.

3. For any given firm, its reputation index (RI) at the moment $t$ is a rolling average of its weighted evaluations

$$\text{RI} = \frac{1}{m} \sum_{a=1}^{m} \frac{\sum_{p=1}^{136} v_{ap} w_p}{\sum_{p=1}^{136} w_p}. \tag{1}$$

   The audits $a = 1, \dots m$ considered for the calculation of RI are those in the 12 months before $t$.

4. To limit the risk of corruption and biased evaluations, randomization was implemented at two levels. First, which work sites to inspect every week was randomly drawn from those where suppliers were actively working. The same worksite could thus be audited

[11] For Acea, this was motivated by the intent to have a benchmark against which to evaluate the new system's effectiveness. While this reform experiment does not satisfy all the characteristics of an ideal field experiment (List and Reiley, 2008), it is nevertheless a very useful natural experiment.

multiple times or never. Second, the composition of the 3-member auditor teams was randomly drawn from the pool of Acea auditors (about 12 engineers).

5. The RI is part of the following scoring formula, which determines the winner of the scoring auction[12] as the bidder with the highest score:

$$\text{Score} = (1-\omega)\cdot \text{Discount} + \omega \cdot \text{RI} \tag{2}$$

$$\text{Discount} = 1 - \frac{\text{Price offered}}{\text{Reserve price}}, \quad \omega = 0.25. \tag{3}$$

   A price-only auction is a scoring auction with weight $\omega = 0$.

6. For new entrants and firms with very sparse audit data, Acea decided to calculate the RI only if at least seven audit visits had been done in the previous 12 months.[13] Otherwise, the supplier would be assigned a RI equal to the average RI of the bidders in the auction.

7. Penalties were formally always part of the contracts, both before and after the reform of the audit system. However, they were rarely enforced to avoid taking disputes to the civil court of law.[14]

A peculiar element of the reform's timing is that Acea initially concealed its motivation for switching to digitized audits. It was only three months after the new audit system had started that Acea announced to its suppliers in a public meeting held on December 20, 2007 ($t1$ in Figure 1) the intention to switch its contract procurement system from price-only auctions to scoring auctions. This allowed Acea to evaluate the distribution of RI in the price-only auctions. In five consecutive meetings with the suppliers (marked $t1, t2, \ldots, t5$), Acea explained this new system and showed simulations of how a firm would benefit from

[12] This scoring auction is strategically equivalent to the one with Score = Discount$+\alpha \cdot RI$, where $\alpha = \frac{\omega}{1-\omega}$.

[13] This requirement concerns the number of audits and not the number of contracts, as a supplier can be audited multiple times for the same contract. This rule limited a "cold start" problem as a barrier to entry, as discussed in Butler et al. (2020).

[14] This explanation for the limited enforcement of penalties was given to us by the Acea personnel but seems coherent with the data from the National Statistical Institute (ISTAT) indicating that, in the period between 2002 and 2012, the average length of this type of proceedings in front of the Rome court is on average 3 years and 3 months. Two levels of appeals could delay even further the final judgment.

higher RI. It also privately informed each firm of its current RI and the distribution of RI across all firms.[15]

Importantly, the audits continued for over three years before the first scoring auction took place. This gave firms more than enough time to increase their compliance and for it to be reflected in their RI (which is a rolling average over the past 12 months). The delayed adoption of the scoring auction was motivated by the need to collect sufficient audit data and an internal dispute on whether it was compliant with Italian law and EU Procurement Directives.

In total, 36 scoring auctions were held in roughly 3 months, when the disagreements among Acea's managers about the legal status of the system, compounded by complaints from the suppliers led Acea to halt the auctions altogether.

After a long break, the system was eventually resumed, entailing a return to price-only auctions, but with some ad hoc provisions to allow Acea inspectors to suspend the contract execution if the audits revealed major problems, which is a feasible substitute for the incentives the scoring auction would otherwise provide. We will refer to this period as the *hybrid* period.

Our paper mostly focuses on the system's original design, which combines scoring auctions with past performance evaluations. Still, we will also use data from the hybrid period to evaluate the long-term effects of the reform.

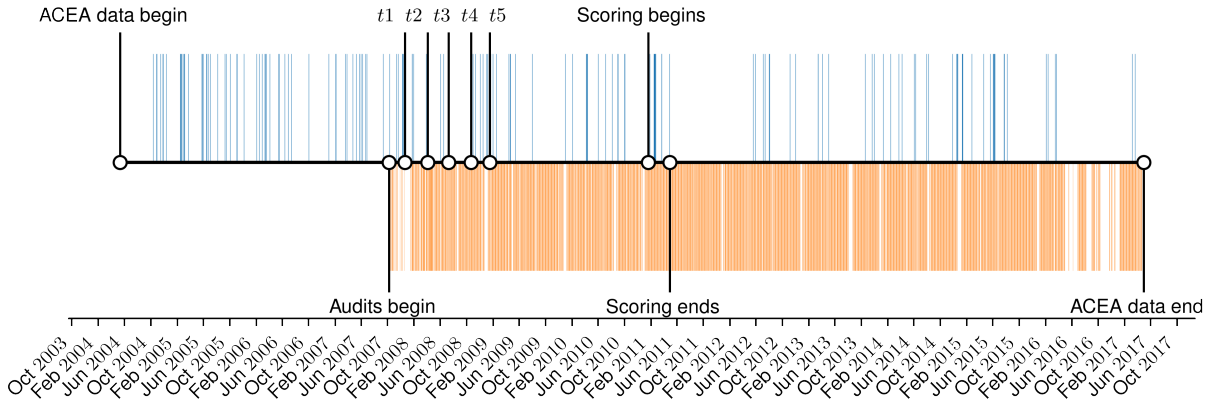
# 3 Descriptive and reduced-form evidence

The analysis is based on four main sets of data: Acea audit data, Acea auction data, Telemat auction data, and, finally, Arera and Istat external measures data.

The first two datasets come from Acea - the designer of the vendor rating system. It contains detailed audit data, recorded through the new auditing system, and detailed auction data, including RI and bids of all participants. The third dataset comes from Telemat, a large provider of public tender data. It contains auction data covering bidding and other auction-

[15] To our best knowledge, no simulations were made by Acea to evaluate the future equilibrium distribution of RI once the scoring auctions would take place.

Figure 1: Timeline




*Note: Upward blue ticks represent auctions; downward orange ticks represent audits. Acea's five announcements of the future switch to equation (2) are marked with t1,...,t5, with t1 being the first announcement date, and t2,...,t5 the dates of follow up meetings where Acea provided an additional explanation to its suppliers regarding this new system.*

related information. The Telemat auction data is less detailed than Acea's, as companies in Italy are only legally bound to report the winning bid, but not the losing bid or the total number of bids. The last two datasets come from Arera and Istata - public authorities supervising the power and water sectors and contain external performance measures.

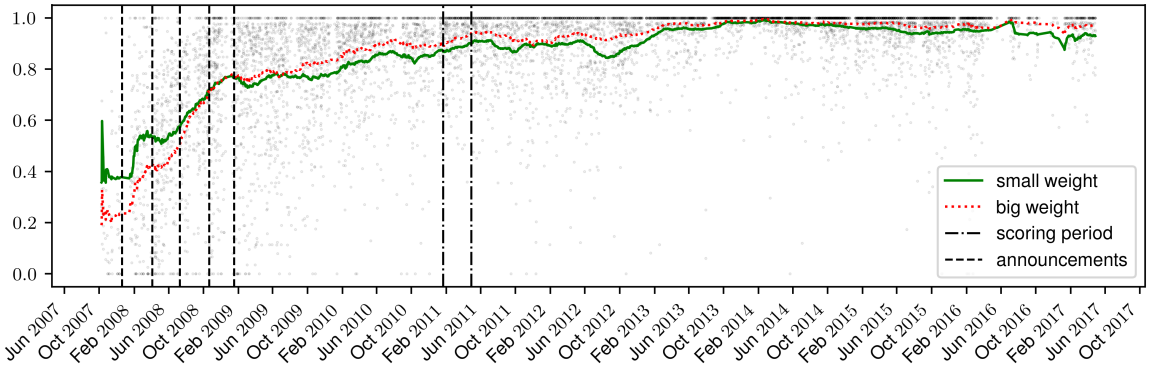
The timeline is split into four main periods, see Figure 1:

- price-only - before December 2007 (t1),
- grace - between December 2007 (t1) and January 2011 (scoring begins),
- scoring - between January 2011 (scoring begins) and May 2011 (scoring ends),
- hybrid - after May 2011 (scoring ends).

## 3.1 Acea's audits.

The records of Acea's audits under the vendor rating system span between 2007 and 2017. There are 365,896 values assigned during 8,973 audits involving 634 contracts and 84 different contractors. Recall that since the subset of worksites inspected in each given week is randomly drawn at the beginning of that week, a contract might receive no inspections or

Figure 2: Internal performance measures by parameter importance



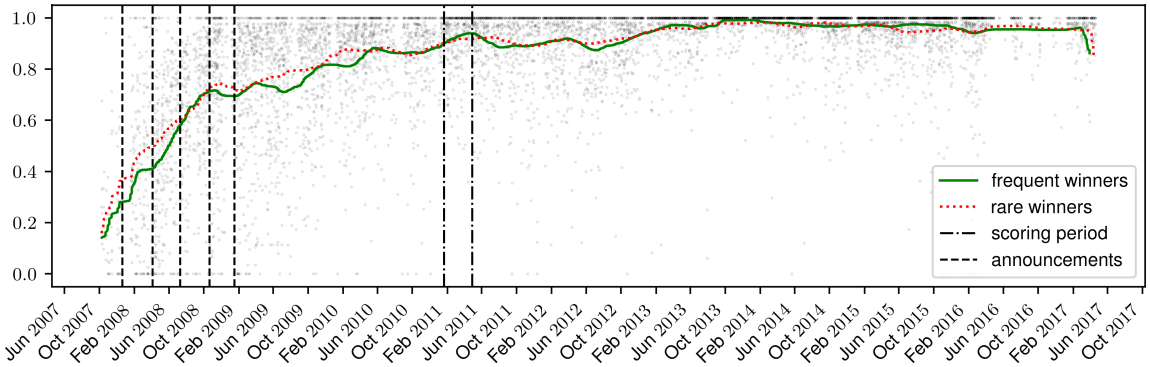

*Note: Weighted audit results, constrained to parameters with small (<=8) and large (>8) weights. The plot represents a moving average over 90 days.*

multiple inspections during its life. Although the shorter-lasting contracts might be rarely observed in the data, the level of detail of this dataset offers a rare opportunity to evaluate how contractual performance evolved over ten years.

Table 2 offers some initial descriptive evidence by reporting summary statistics and aggregating parameters at the level of the 12 categories. The table shows that there is substantial heterogeneity in the frequency with which different parameters are scored: very few contracts entail features that allow inspectors to check parameters in the "Customer relationship mgmt" category while, at the opposite end of the spectrum, parameters in the "Works site regularity" and "Works site safety" categories are systematically assessed. Table 2 also reports the share of compliant parameters separately for each of the 12 categories and for four time periods. Across nearly all categories, we can see a sharp increase throughout the grace period. The increase is more moderate in later periods. We perform several exercises to gain insights into the evolution of internal performance measures.

In the first exercise, we split the parameters into two categories: those with small and large weights $w_p$. We then reconstruct the analog of RI, but separately for each group, with a rolling window of 90 days. As seen on Figure 2, in the beginning, compliance with more important parameters (large weights) was smaller than with less important ones (small weights), as the firms' perception of performance was probably different from that of Acea. However, as the announcements happen, we see firms realigning their behavior, putting relatively more effort into the parameters with large weights.

Figure 3: Internal performance measures by firm type



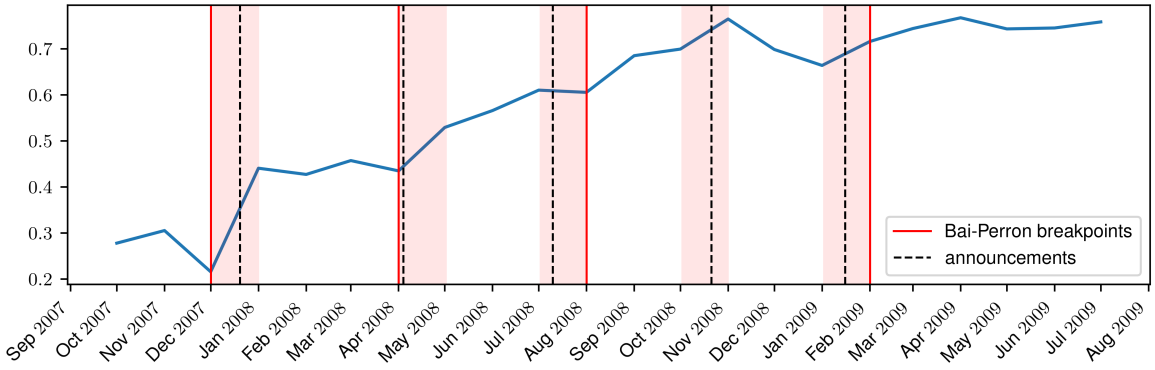

*Note: Weighted audit results by frequent (top 10) winners in first-price auctions and rare (everybody else) winners. The plot represents a moving average over 90 days.*

In a different exercise, we split firms into frequent winners in price-only auctions (top 10 winners before 20 December 2007) vs. everybody else, and we compute their average RI with a rolling window of 90 days. As seen on Figure 3, all suppliers improved their performance, albeit with different timing. The firms that used to win more frequently in the price-only auctions took longer to comply. We interpret this as quality concerns (adverse selection) in the market, in the sense that firms that are relatively better at cost reduction are also relatively worse at complying.

As discussed earlier, most performance improvements observed over the long run occurred during the grace period. To formally show the connection between performance changes and announcement timing, Table 3 reports the results of Bai-Perron tests for the presence of structural breaks. The variable of interest is the monthly weighted average compliance across all parameters. We do not specify the dates of the breaks but let the test determine them, either without specifying how many breaks there are (column 1) or specifying that there are 5 breaks at unknown dates (column 2). Surprisingly, all breaks appear at either the beginning or the end of the month during which an announcement was made, see Figure 4.

In Table 4, we complement the time series evidence with estimates of linear regressions of the average monthly compliance by contract and supplier on dummy variables for the four break dates detected by the Bai-Perron test (see column (1), panel (b) of Table 3) and a control for whether the contract is for public illumination. Column (1) confirms the significance of all four break dates. When we gradually augment the set of regressors to

Figure 4: Internal performance measures' response to announcements




include fixed effects for suppliers, contracts, and months, we find that the dummy for $t1$ preserves its statistical significance and large magnitude, thus confirming its relevance. In the Appendix E, we also present additional results exploring the composition of the set of parameters audited and firms inspected.

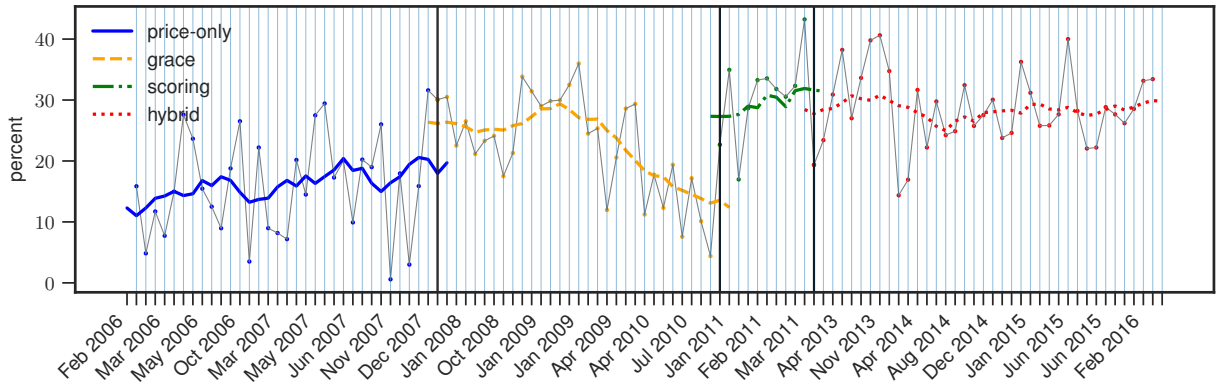
## 3.2 Acea's auctions.

The records of Acea's auctions span between the years 2004 and 2017. We are primarily interested in the evolution of Acea's winner's discounts, see Figure 5. Unlike internal performance measures, which increase steadily and monotonically over time, the discounts change non-monotonically. However, the cumulative effect appears positive if we compare the trajectory endpoints between the first announcement and the last scoring auction.

If we focus on the grace period, the discounts grow, then rebound. We believe the initial increase is due to the temporary stimulus to win more auctions. Firms need to win more auctions to increase their chances of being audited and increase one's reputation index. Indeed, during this time, we observe the largest increase in RI. However, closer to the end of the grace period, the discounts fall back to levels similar to or even below those observed before the first announcement (t1).

An important detail is an apparent discontinuity around the switch from grace to the scoring period. It is observed for both average and winning discounts but is higher for the latter. Unfortunately, the auctions are located sparsely over time, preventing us from formally testing continuity. Still, we can attempt to estimate two regression lines in the

Figure 5: Evolution of discounts over time.




*Note: average (per auction day) discount of Acea's winning firm and its rolling window average (over five auction days). The timeline is split into four periods: price-only (before t1), grace (after t1), scoring, and hybrid.*

small neighborhood of the switch, see Figure 6. Indeed, we observe a sharp discontinuity. If we average over three auction days before and after the switch, the discount jumps up by approximately 14 percentage points (from 11 to 25).

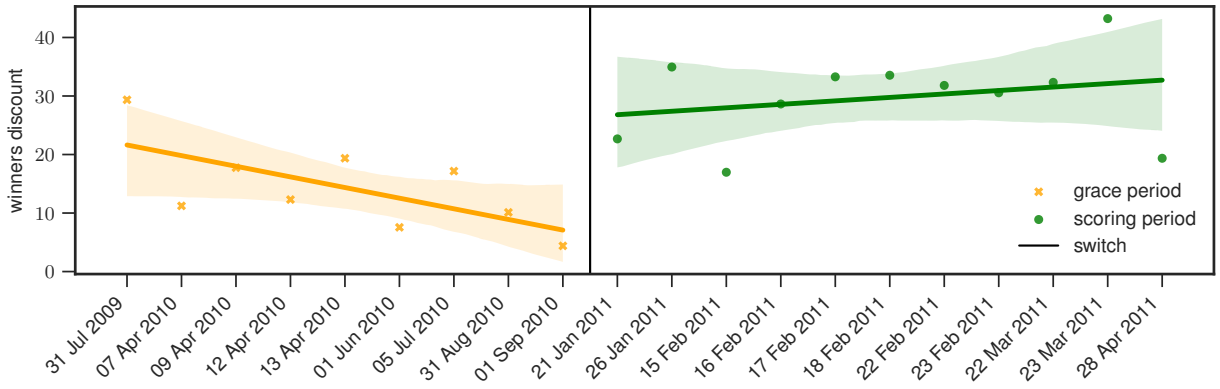
The jump indicates that the switch in the auction mechanism is likely behind the rise in discounts. However, several exogenous factors could also contribute to the discount: (i) the reserve price, since the discount is the percentage distance between the winning bid to the reserve price, (ii) the number of bidders, since the winning bid is the lowest order statistic of bids.

It is unlikely that the reserve price explains the oscillation of the discount. First, public buyers are not in full control of it: it is obtained by multiplying input quantities (estimated by the procurer's engineers) by their prices and summing up these products. Crucially, input prices are not the current market prices but the list prices set yearly by the region where Acea operates and used exclusively by contracting authorities to calculate reserve prices, thus excluding deliberate manipulation.[16] These prices are much higher than the typical production costs, so the reserve price is always non-binding.

On the other hand, the number of bidders contributes to the winning discount in two different ways. First, as it is an order statistic, the winning discount is increased by the

[16]However, there is a trend in Acea's contracting to concentrate its demands into fewer, larger lots, which we observe in the scoring and hybrid periods.

Figure 6: Evolution of discounts around the switch.




*Note: average (per auction day) discount of Acea's winning firm around the switch.*

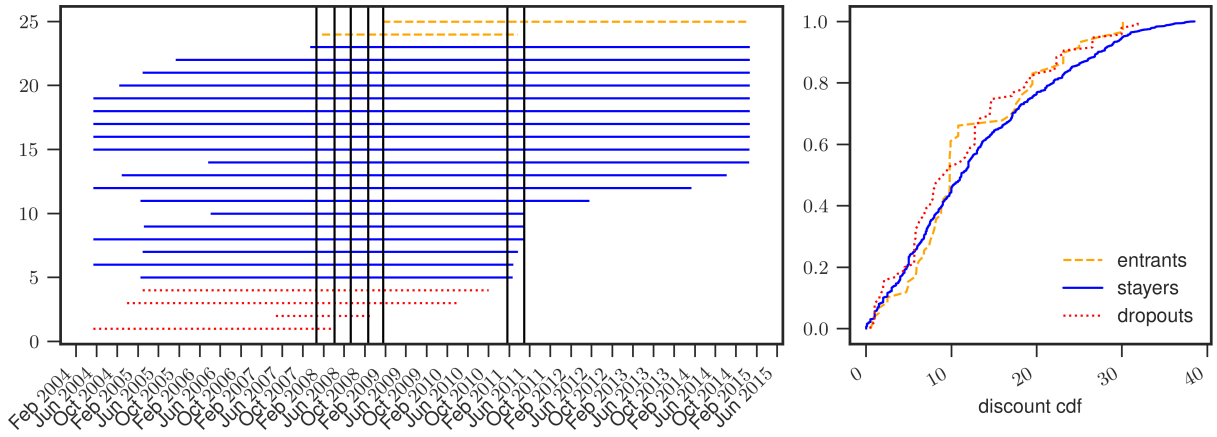
number of bidders. Second, if the participants observe or anticipate an increase in bidders, they bid more aggressively, increasing the discount even further. However, in our data, there is no noticeable jump in the number of bidders, in the neighborhood of the switch. Still, we repeat the exercise with residualized discounts, see Appendix E.

Another possible explanation for the change in discounts is the change in composition (as apposed to the size) of the pool of participants in auctions. First, the equilibrium analysis of the price-only and scoring auctions predicts full participation, see Section 4, when the reserve price is not binding. Second, because Acea is a public buyer, and because of the EE and Italian Regulation, the entry barriers into these auctions are small, so endogenous entry is negligible. Nevertheless, we turn to the data to investigate this question.

Among the 25 firms that had entered the auctions before the hybrid period and exited no earlier than the grace period, we define three non-overlapping groups. 2 *entrants* - those who only entered in the grace period and participated in the scoring period at least once. 19 *stayers* - those who participated before $t1$ and participated in the scoring period at least once. 4 *dropouts* - those who exited in the grace period. We show their descriptive statistics is Table 5.

We plot the duration of their activity in Acea's auctions and the CDF of their discounts before the scoring period, see Figure 7. Note first that the distribution of discounts across the three groups is very similar if we consider price-only auctions before the scoring period. The average discount of the dropout is only a few percentage points lower (11.37 as opposed

Figure 7: Entrants, stayers, and dropouts.




*Note: left figure - duration of firms activity in Acea's auctions, separated into dropouts, entrants, and stayers; right figure - the CDF of their discounts before the scoring period.*

to 13.07) than that of the stayers. Moreover, the dropouts have participated in only a small fraction of auctions. Thus the upwards jump in the discount can not be explained by an exodus of firms of a particular kind. If we consider entrants and stayers, their average discounts are also similar before scoring (12.08 and 13.07), but they both increase (up to 19.65 and 19.37, respectively) in the scoring period. Thus, the jump can not be explained by the entry of new firms either.

Before moving to the next section, we ask one last question. Is the discount related positively or negatively to the reputation index? The answer depends on the segment of bidders considered, see Table 6. Considering a full set of bidders, the regression coefficient of offered discount on the reputation index is positive and equal to 0.44. However, if we consider the rank of the bidder in the scoring auction, it turns out that, among the strongest, in terms of score, participants (rank 1), the correlation is negative: -0.32.[17] Since the strongest participants determine the outcome of the auction, we interpret this as strong evidence of adverse selection in the environment.

[17]For comparison, the slope of the scoring rule would correspond to -0.33.

## 3.3 Telemat auction data

To assess how Acea's prices relate to those in the rest of the market, we form a control group composed of similar auctions by other companies.[18] The data include the object of the contract, the reserve price, the winning discount and date, the identity of both the procurer and the winning contractor, and various other information on the call for tenders, such as the award procedure and criterion. [19]

Table 7 reports summary statistics for Acea and the control group, dividing them into four panels, one per each studied period. One can see that the discounts in the control group increase steadily across periods, unlike Acea's auctions.

Regarding the other variables reported in Table 7, there are no major differences between the top two panels, neither for Acea nor for the Control group. This is the case, for instance, for contract duration or the share of public illumination contracts.[20] The number of bidders is only slightly higher in the scoring period than in the grace period for Acea, and it seems unlikely that it explains such a dramatic change in discounts. To summarise, none of the exogenous factors are strong enough to explain the increased discounts on their own. We will therefore attempt to explain it with all these factors combined, in reduced form.

We estimate the effect of the announcements t1-t5, as well as the switch to the scoring auction, on the winner's discount, using a difference-in-differences (DID) strategy. Consider the equation below:

$$D^{w}_{ift} = a_f + b_t + \beta_1 d^{Acea}_{grace,t1-t5} + \beta_2 d^{Acea}_{grace,>t5} + \beta_3 d^{Acea}_{scoring} + \beta_4 d^{Acea}_{hybrid} + \gamma X_{ift} + \epsilon_{ift}, \qquad (4)$$

where $D^{w}_{ift}$ is the winning discount (over the reserve price) and the index $i$ indicates the auction, $f$ is the entity awarding the contract, and $t$ is the year. There are four treatment

[18] These jobs belong to a well-defined contract category identified by the Italian regulation as "$OG10$," which makes it feasible to select comparable projects across different buyers. Furthermore, using textual search methods, we separated $OG10$ contracts into those involving public illumination and those involving electrical substations. Finally, to ensure contract comparability, we trimmed a few particularly large or small contacts (i.e., those with a reserve price below €10,000 or above €2.5 million).

[19] For a subset of auctions, we integrate the data with the information on losing bids and the subsequent life of the contracts using data from the authority supervising public contracts (ANAC).

[20] It is important to stress that the main effort to ensure the comparability of the auctions was at the data collection stage, where we selected only auctions that, in terms of their object, were a close fit to the public illumination and electricity distribution contracts auctioned off by Acea.

dummy variables, for the contracts awarded by Acea in each of the 4 periods: grace before t5, grace after t5, scoring, and hybrid. The base group is therefore the price-only period before the announcements. Fixed effects are the entity awarding the contract ($a_f$) and year ($b_t$) and other covariates ($X_{ift}$) involving contract characteristics.

Table 8 presents these baseline estimates in four specifications differing in the set of covariates and control group observations. In particular, we consider limiting the sample to either buyers located in central Italy (which might be more similar to Acea in terms of input prices, the pool of suppliers, and environmental conditions) or only outside this area (which might serve to limit contamination concerns). Across all specifications and samples, we find fairly consistent estimates: in all periods, except for the second half (>5) of the grace period, the discounts are between 4 and 8 percent higher than in the price-only period.

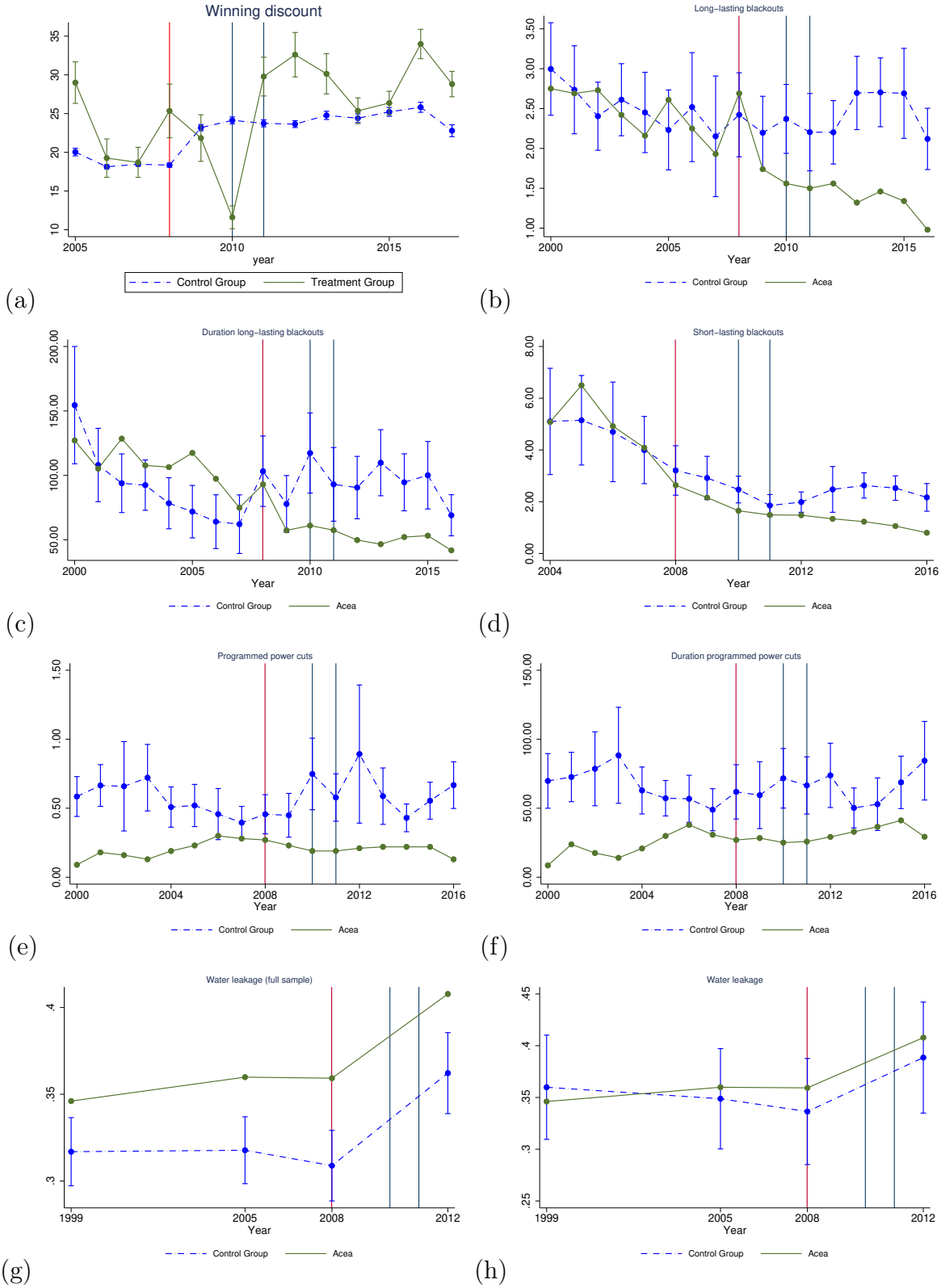
Overall, the Telemat data reinforces our finding that the increase in performance was accompanied by an increase in discounts in the scoring period, and it also lasted through the hybrid period.

## 3.4 External measures

Our final dataset is related to external performance measures. In Italy, electricity and water are both partially-regulated sectors. For electricity, although only power transmission is still under a regulatory regime, the regulator (ARERA) collects detailed information on the whole sector. From ARERA we were thus able to obtain various firm-level performance measures. These yearly data range from the year 2000 to 2016 and cover all low-voltage power distributors, including Acea. Herein, the main indicators of firm performance are the number and duration of blackouts and programmed power cuts. The top six rows of Table 9 report summary statistics for these external performance measures, none of which is part of the RI parameters.

In Figure 8, the different panels from (b) to (f) plot the evolution of all of our external measures: (b) number of long-lasting blackouts (i.e., lasting more than 3 hours), (c) duration of long-lasting blackouts (in minutes), (d) number of short-lasting blackouts (i.e., lasting less than 3 hours), (e) number of programmed power cuts and (f) duration of programmed power cuts. The observed pattern is similar across most measures: after $t1$, Acea's performance

Figure 8: Evolution of Discounts and External Performance Measures




*Note: Acea (in green) and other utilities (in blue, dashed). The red line indicates the date of Acea's first announcement. The blue, vertical lines indicate the scoring rule period. The dashed graph corresponds to the control group.*

gradually improves in absolute and relative terms. For instance, this is the case for the number of long-lasting blackouts experienced by Acea's customers: as the plot shows, this number declines in absolute terms and relative to those of the clients of the other utility company in the control group.

As for the programmed power cuts, they typically imply improved service quality as they are associated with work on the grid, and they substitute unplanned blackouts. The graphical evidence supports the idea that improvements in the performance of Acea's suppliers should lead to improvements in the external measures, although possibly with a time lag. The reasons why improvements in electric grid performance occur more slowly than those in internal performance are mostly due to technological constraints: even if suppliers use higher quality joints and materials (some of the quality parameters, see Table 2), only when a large enough portion of the grid is affected will blackouts fall.

The last two plots in Figure 8 cover the water sector: (g) water leakage (h) water leakage (pop. over 1 million). Recall that this sector never switched to the past performance auditing system introduced for the electricity sector. This sector thus serves the role of a placebo. External performance measures for the water sector have been obtained from the Italian Statistical Institute (Istat) environmental census. This census is performed in collaboration with the water distributors and includes information on water inflow and outflow in the distribution channel for each Italian county from 1999 to 2012. A performance measure is thus the extent of water leakage, calculated as the percentage incidence of leakage over water inflow. Although the data is released at the county level, it is easy to aggregate counties in such a way as to pin down the water leakage level experienced by Acea. In fact, by law, each county can have no more than one water distributor, so we aggregated up the water leakage data for all the counties served by Acea.[21] The bottom rows in Table 9 report summary statistics for the water sector, while the bottom panels of Figure 8 plot the dynamic of the water leakage indicator, separately for ACEA and other firms indicating that there is no evidence of lower leakages for Acea.

We now move towards evaluating, in reduced form, the impact of Acea's announcement

[21] This aggregation is performed by weighting the leakage in each of the counties served by a provider by its share of water customers relative to the total population of water customers served by the provider. County data are aggregated to mirror the "catchment areas" over which there is, by law, only one water provider.

at $t1$ on the six external performance measures introduced earlier. This is relevant as an additional check that multitasking effects are not muting the benefits of the reform implied by the internal performance measures and as an assessment of the reform on highly socially valuable measures. The estimation strategy is a DID based on the following equation:

$$O_{ft} = a_f + b_t + \beta_4 D^{Acea}_{t \geq t_1} + \gamma X_{ft} + \epsilon_{ft}, \tag{5}$$

where $O_{ft}$ is one of the performance outcomes we observe at the level of buyer, $f$, and year, $t$. On the right-hand side of the equation, $a_f$ and $b_t$ are fixed effects for buyers and years, $X_{ft}$ is a matrix of controls that includes the number of customers and, finally, $D^{Acea}$ is a dummy for Acea's auctions held after 2007. The coefficient of interest is $\beta_4$, which thus captures the difference in external performance between Acea and other buyers after Acea announced the change in the adopted award criterion in December 2007.

Table 10 reports the estimates. The first five columns cover different measures of the quality of electricity distribution, while the latter two cover water leakages for both the full sample of firms and for the subsample of larger firms. These estimates confirm the graphical evidence provided in Figure 8: for the five outcomes measuring quality in the low-tension electricity distribution sector, the effect of the treatment is to reduce the number and length of long-lasting blackouts, reduce the number of short-lasting blackouts and, on the contrary, increase programmed power cuts. The latter is most likely a signal of greater maintenance efforts. For the water sector, where no RI was introduced, Acea did not improve its performance (in terms of leakage) relative to other comparable firms. Regardless of whether we consider all distributors or only the largest players, the finding of no effect remains.

Overall, the external measures confirm the long-lasting performance improvements.

# 4 Stylized model

In this section, we propose a stylized model of a scoring auction, where non-price characteristics of the firm (i.e., its quality) are related to its past performance. The firm can invest in

its quality by performing better. However, since performance is measured in past contracts, the associated costs are effectively sunk from the current viewpoint. This differs from the classical models of scoring auctions in Che (1993) and Asker and Cantillon (2010), where costs or raising quality are not sunk.

Let the auction have $2 \leqslant n \leqslant N$ ex-ante identical firms, with exogenous probabilities $p_n$, competing for a single procurement contract. The reserve price is normalized to 1 and is non-binding[22]. Let $0 \leqslant \theta_i \leqslant 1$ be firm $i$'s efficiency parameter, such that its costs of executing the auctioned contract, i.e., its *production costs*, are equal to $1 - \theta_i$. Firm $i$ has initial quality $\underline{q}_i \geqslant 0$, which she can increase to $q_i \geqslant \underline{q}_i$, prior to the auction, at an additional *investment cost* $C_i(q_i)$, where

$$C_i(q) = \frac{(q - \underline{q}_i)^2}{2\beta}, \tag{6}$$

for some $\beta > 0$. The $(\theta_i, \underline{q}_i)$ pair captures the firm's private type and is i.i.d.

We will consider a *quasi-linear scoring rule* $s_i = \alpha q_i + d_i$, where $s_i$ is the firm's score and $d_i$ is its discount (i.e., the difference between the reserve price and its bid).[23] As in all scoring auctions, the firm with the highest score wins the contract.

We assume that there is no exchange of information between the firms after they invest in quality and before they choose their discounts. Thus the choice of $(q_i, d_i)$ or, equivalently, $(q_i, s_i)$ can be modeled as simultaneous. Following Asker and Cantillon (2010), we will create an auxiliary variable called the *pseudo-type* $\rho_i = \alpha q_i + \theta_i$, such that firm $i's$ profit margin $\theta_i - d_i$, upon winning the auction, is equal to $\rho_i - s_i$.

We are interested in a symmetric equilibrium with strictly monotone strategies $\sigma : \rho \to s$. Denoting the equilibrium distribution of score as $F_s(.)$, then each firm maximizes

$$U_i(q, s) = (\rho_i - s)G(s) - C_i(q), \quad G(s) = \sum_{n=2}^{N} p_n F_s^{n-1}(s)$$

[22] For simplicity, we allow for negative discounts.

[23] Quality $q$ in our model is measured by the reputational index RI in the data, and the $\alpha$ weight (often referred to as the *the dollar value of quality*) is related to the $\omega$ weight in the Experiment according to the simple formula $\alpha = \frac{\omega}{1-\omega}$.

subject to the $\rho_i = \alpha q + \theta_i$ constraint. This implies two sets of first-order conditions:

$$(\rho - s)\frac{\partial G}{\partial s}(s) - G(s) = 0, \tag{7}$$

$$\alpha G(s) - \frac{\partial C_i}{\partial q}(q) = 0. \tag{8}$$

Equation (7) is the standard optimality condition for auctions. It also shows that the score depends only on the pseudo-type $\rho = \alpha q + \theta_i$ since there is no binding reserve price. The pseudo-type is, however, endogenous. The equilibrium strategy can be written as

$$\sigma(\rho) = \int_{\underline{\rho}}^{\rho} z dH(z)/H(\rho), \quad H(\rho) = \sum_{i=2}^{N} p_n F_{\rho}^{n-1}(\rho) \tag{9}$$

where $F_\rho(.)$ is the equilibrium distribution of pseudo-type and $\underline{\rho}$ is the lowest participating pseudo-type. Equation (8) is the condition for the optimal choice of quality, which is both necessary and sufficient since $U_i$ is strictly concave in $q$, for any chosen score $s$. [24]

Our first observation is that there must be full participation (i.e., no endogenous entry). Indeed, if, in a symmetric equilibrium, a positive mass of types does not enter, then the auction has no participants with a positive probability. Moreover, the reserve price is, by construction, not binding. Thus, any potential entrant has incentives to enter.

Denote two functions $\eta_\alpha(x)$ and $\xi_\alpha(x)$ as below

$$\eta_\alpha(x) := \sum_{n=2}^{N} p_n \xi_\alpha^{n-1}(x), \quad \xi_\alpha(x) := Prob(\theta_i + \alpha \underline{q}_i \leqslant x), \tag{10}$$

and assume that they are continuously differentiable and strictly monotone. Since, in equilibrium, $G(s) = H(\rho)$, it follows from (6), (8) and (10) that

$$\rho_i - \alpha^2 \beta H(\rho_i) = \theta_i + \alpha \underline{q}_i, \quad H(\rho) = \eta_\alpha(\rho - \alpha^2 \beta H(\rho))$$

which uniquely identifies $H(\rho)$ and the equilibrium pseudo-type $\rho_i$ for each $(\theta_i, \underline{q}_i)$. Moreover, the pseudo-type $\rho_i$ is necessarily monotone in $\theta_i$, for all $q_i$, since $\theta_i = \eta_\alpha^{-1}(H(\rho)) - \alpha q_i$.

[24] Contrary to the analysis of Che (1993), where the choice of quality was independent of the score, here the marginal cost of quality equals $\alpha$ (the dollar value of quality) times $G(s_i)$.

With the pseudo-type at hand, we can compute equilibrium quality and score via equations (8) and (9), which completes the construction of the equilibrium. It remains to show that the second-order conditions are satisfied, see Appendix B for the details.

**Proposition 1.** *A unique, symmetric equilibrium, with full participation and strictly monotonic strategy $\sigma$ exists if $\xi_\alpha(.)$ is continuously differentiable and strictly increasing in the support of $(\theta_i, \underline{q}_i)$.*

We split the remaining analysis into two separate cases.

## 4.1 Univariate types.

For now, let the firms vary only in their efficiency parameter $\theta$, that is, only in their production costs. While seemingly innocuous, this assumption has major implications.

Recall that the score is monotone in $\rho$ and, therefore $\theta$. Furthermore, since all firms have the same parameter $\underline{q}$, the firm with the highest $\theta$ wins, independently of $\alpha$. This leads to several important conclusions. First, the winner does not depend on $\alpha$ thus winner's quality will increase by (8). Second, the firm's interim expected profits are fixed by the revenue equivalence principle, and she is indifferent between the scoring and the price-only auctions.

From the buyer's perspective, however, the scoring auction is more expensive. Indeed, the firm's auction profits are equal to its total profits net of the investment costs, thus its expected discounts are lower in the scoring auction than in the price-only auction.

**Proposition 2.** *In the symmetric equilibrium with univariate types of a scoring auction the expected quality is higher and the expected discounts are lower than in the symmetric equilibrium of the corresponding price-only, first-price auction.*

To illustrate the idea, we compare the outcomes between the scoring and the price-only auction for 2 firms, $\underline{q}_i = 0$ for all $i$, $\theta_i$ distributed uniformly on $[0, 1]$, and $\beta = 1$ in the box below.

| design | total profits | auction profits | discount | quality |
|---|---|---|---|---|
| price-only | $\theta^2/2$ | $\theta^2/2$ | $\theta/2$ | 0 |
| scoring | $\theta^2/2$ | $(1+\alpha^2)\theta^2/2$ | $(1-\alpha^2)\theta/2$ | $\alpha\theta$ |

Indeed, in a scoring auction, firms produce higher quality, and the associated investment costs are compensated, in expectation, by higher auction profits, adding up to the exact same total profit as in the price-only auction.

Thus, univariate types can not explain our empirical puzzle.

## 4.2 Bivariate types.

Let the firms vary in both $\theta_i$ and $\underline{q}_i$. Relative to univariate types, this introduces two new forces into the picture: the selection of winning firms and the evolution of bid shading.

Consider switching from the price-only auction to the scoring auction. Since the former is cost-efficient, the winning firm's $\theta_i$ can only decrease, and strictly so for generic distributions. Thus, if there was no strategic shading of bids, an increase in $\alpha$ would guarantee a decrease in discounts, just like with univariate types.

However, the strategies (9), and thus the informational rents of the firms can go either way. If with the introduction of the scoring auction, the pseudo-type distribution becomes more concentrated, the shading is likely to decrease, due to a more aggressive equilibrium strategy $\sigma$. Thus, while a switch to the scoring auction, through selection, necessarily leads to a decrease in the cot-efficiency of the winning firm, it may lead to an increase in discounts, if the shading of the winning firm decreases significantly.

At the same time, it is still impossible for quality to decrease in the scoring auction. Assume to the contrary, that in the price-only action, a firm with type $\theta'$ and quality $\underline{q}'$ wins, but in the scoring auction another firm with type $\theta''$ and quality $q'' < \underline{q}'$ wins. Since $\theta'' \leqslant \theta'$, we can derive that

$$\theta'' + \alpha q'' < \theta' + \alpha \underline{q}' \leqslant \theta' + \alpha q',$$

where $q' \geqslant \underline{q}'$ is the quality of the former winner, adjusted for the scoring auction. But this contradicts the fact that the firm with the highest pseudo-type wins.

**Proposition 3.** *In the symmetric equilibrium with bivariate types of a scoring auction, the expected quality is higher and the expected discounts are either higher or lower than in the symmetric equilibrium of the corresponding price-only, first-price auction.*

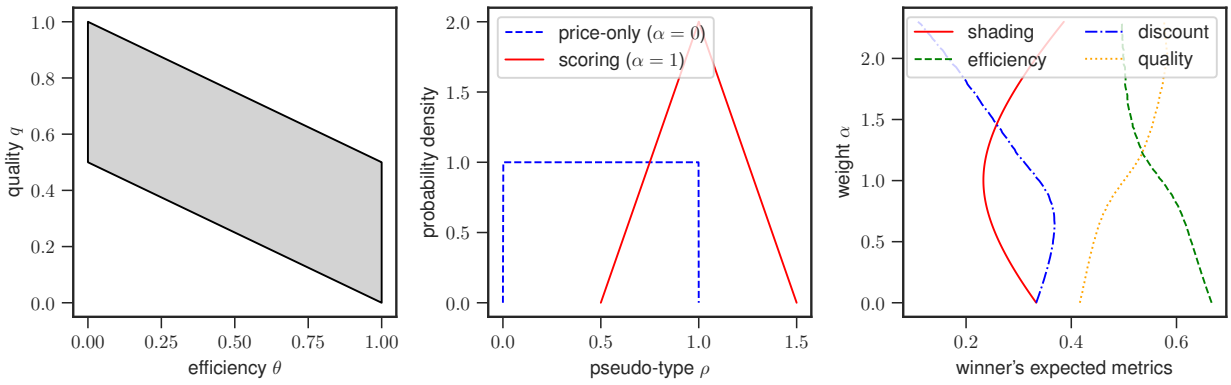



Figure 9: Bivariate types example.

To illustrate the idea, we pick a distribution such that $q$ and $\theta$ are slightly negatively correlated in the scoring auction. Namely, we take two firms and a uniform distribution of $(\theta, q)$ in the region defined by $0 \leqslant \theta \leqslant 1$ and $1/2 \leqslant q + \theta/2 \leqslant 1$, see Figure 9 (left).[25] For the price-only auction, that is, a scoring auction with weight $\alpha = 0$, the pseudo-type distribution is uniform, while with weight $\alpha = 1$, it is pyramid shaped, see Figure 9 (middle). The latter is more concentrated and is therefore associated with more aggressive bidding (i.e., less shading).

| design (weight) | $\partial H(\rho)/\partial\rho$ | $H(\rho)$ | expected winner's efficiency | expected winner's shading | expected winner's discount |
|---|---|---|---|---|---|
| price only ($\alpha = 0$) | $1,\ \rho \in (0,1)$ | $\rho$ | $2/3$ | $1/3$ | $1/3$ |
| scoring ($\alpha = 1$) | $\begin{cases} 4\rho - 2, & \rho \in (.5, 1) \\ 6 - 4\rho, & \rho \in (1, 1.5) \end{cases}$ | $\begin{cases} 2\rho^2 - 2\rho + 1/2 \\ -7/2 + 6\rho - 2\rho^2 \end{cases}$ | $37/60$ | $7/30$ | $23/60$ |

We present the results of the comparison between $\alpha = 0, 1$ in the box above, see Appendix B for the derivation. One can see that, while the winning firm's efficiency (type $\theta$)

[25] To rationalize the observed distribution of $(\theta, q)$, one has to verify that $\rho - \alpha^2\beta H(\rho)$ is increasing for all $\rho$ in the support. For $\alpha = 1$ it would suffice that $\beta < 1/2$. For simplicity, one can assume that quality is exogenous, which corresponds to $\beta \to 0$.

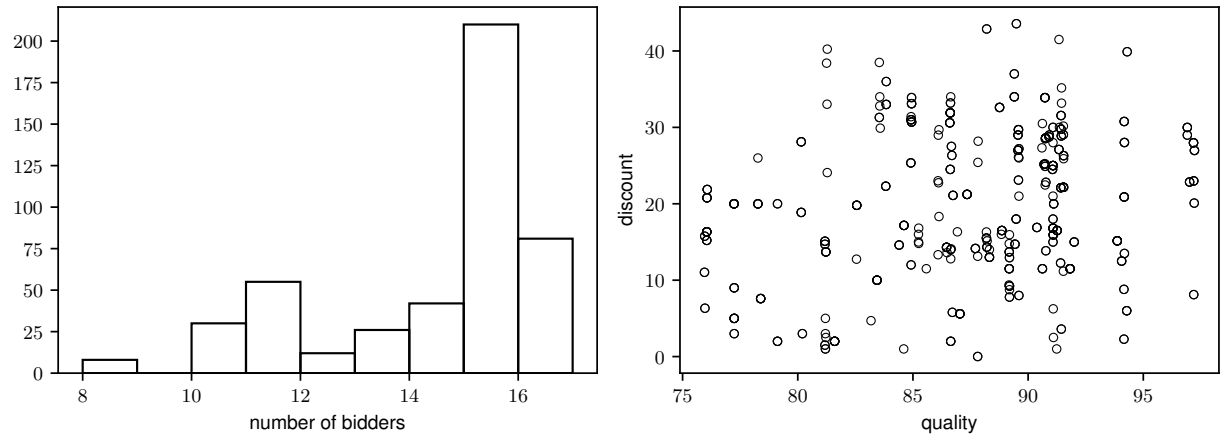



Figure 10: Distribution of data

has decreased through selection, the shading has decreased so much, that the final discounts increase in the scoring auction, albeit very slightly.

We also simulate numerically the evolution of the winner's characteristics as functions of $\alpha$, see Figure 9 (right), holding firm's quality fixed (which corresponds to $\beta$ approaching 0). Both expected quality and expected discount are increasing in the $\alpha$ weight, in the range between 0 and roughly .6.

Thus, the bivariate types model can explain our empirical puzzle.

# 5 Empirical model

In this section, we estimate a structural model of a scoring auction non-parametrically and simulate the outcomes of a counterfactual price-only auction. We are primarily interested in the behavior of discounts across the counterfactual simulations.

The dataset consists of 34 first-score sealed-bid auctions, held over 11 days between 2011-01-21 and 2011-04-28. The scoring rule is quasi-linear with weight $\alpha = 1/3$. We observe 479 quality-discount pairs, with quality (measured by the reputational index RI) distributed above 76 (out of 100), while discount distributed below 44 (out of 100). The number of bidders varies between 8 and 16, with an average of 13.64 and the mode at 15, see Figure 10.

To pick an appropriate structural model, we have to answer four key questions: (i)

whether the reserve price is binding, (ii) whether the number of bidders is known, (iii) which model of heterogeneity to use, and (iv) whether this is an IPV (independent private values) or APV (affiliated/correlated private values) environment.

The evidence from the Experiment suggests that the answer to the first two questions is negative. Indeed, the reserve price is intentionally set so that it is almost never binding. Moreover, since the format is sealed-bid, firms do not have hard information about who participates, so it makes sense to model the number of bidders as random.

Next, using a standard mapping[26] between the first-price and the first-score auctions when the reserve price is not binding, we can use the tests from Krasnokutskaya (2011) to pick a suitable model of auction-level heterogeneity. The additive model of heterogeneity is soundly not rejected, see Appendix C. The intuition behind it is that contracts have fixed production costs, common to all bidders in the auction. The variability in the scale of production costs is of lesser concern since discounts are already measured as a percentage of the reserve price.

Finally, we would like to test whether conditional on the observables, this is an IPV, rather than an APV model. We apply the analog of the sup-norm test, suggested by Haile, Hong and Shum (2003), and the IPV hypothesis is not rejected, see Appendix C.

## 5.1 Model Primitives and Identification.

For now, consider a single representative auction, as if there is no auction-level heterogeneity and quality is already chosen and observed by the buyer. As in the stylized model, denote quality as $q$, discount as $d$, pseudo-type as $\rho$, score as $s$, and the best possible discount the firm can offer for the contract as $\theta$.

The first primitive of the model is the probabilities $p_n$, which are trivially identified because the number of bidders is observed. The second primitive is the marginal distribution of $\theta$. Since we observe the joint distribution of $(s, q)$, through the optimality conditions (7), we observe the joint distribution of $(\rho, q)$ and, therefore, the marginal distribution of $\theta$ is identified. The joint distribution of $(\theta, \underline{q})$ is not identified, since $\beta$ - the parameter of investment costs - is unknown.

[26] See, for example, Che (1993); Asker and Cantillon (2010); Andreyanov (2019)

## 5.2 Estimation and simulation.

To account for auction-level heterogeneity, as well as a possible evolution of beliefs and strategies over time, we adopt a simple parametrization, where the location of the distribution of discounts (and thus the pseudo-types) is a linear function of the auction-day dummy variables.

Denote the auction-day fixed effects as $\gamma$. In other words, if $S, D, Q$ are the observed score, discount, and quality, then

$$S = s + \gamma, \quad D = d + \gamma, \quad Q = q, \quad S = \alpha Q + D.$$

Due to the linear scalability of the optimality conditions, a shift in the location of the distribution of pseudo-types does not affect the shape of the strategy. This motivates (additively) partialling-out the auction-level heterogeneity in a reduced form. We regress the observed score $S$ on auction-day dummies, see Table A.1, to obtain the estimates of fixed effects $\gamma$.

Similar to Guerre, Perrigne and Vuong (2000) and Li, Perrigne and Vuong (2000) we use a non-parametric approach to estimate the sample analog of equation (7), see Andreyanov and Franguridi (2021) for details. For any value of $\beta$, we can therefore obtain the pseudo-sample of estimated pairs $(\hat{\theta}_i, \hat{\gamma}_i)$, and simulate the outcomes in the counterfactual price-only auction, see Appendix C for details.

## 5.3 Counterfactuals

In this section, we present the counterfactual monetary outcomes for the price-only auction and compare them to the default (with $\alpha = 1/3$) scoring auction, see Table 11. Columns (1) and (3) contain outcomes, averaged over all bids. Columns (2) and (4) contain the average winner's outcomes. The standard deviation is computed via Bootstrap.

We can see that, relative to the scoring auction, the counterfactual price-only auction is associated with slightly lower (71.13% as opposed to 71.59%) production costs of the winning firm. This is not a surprise, since the price-only auction is the most efficient one. However, it is also associated with greater shading (1.83% as opposed to 1.604%), which partially offsets the former. Their combination leads to the price-only auction being slightly cheaper in terms

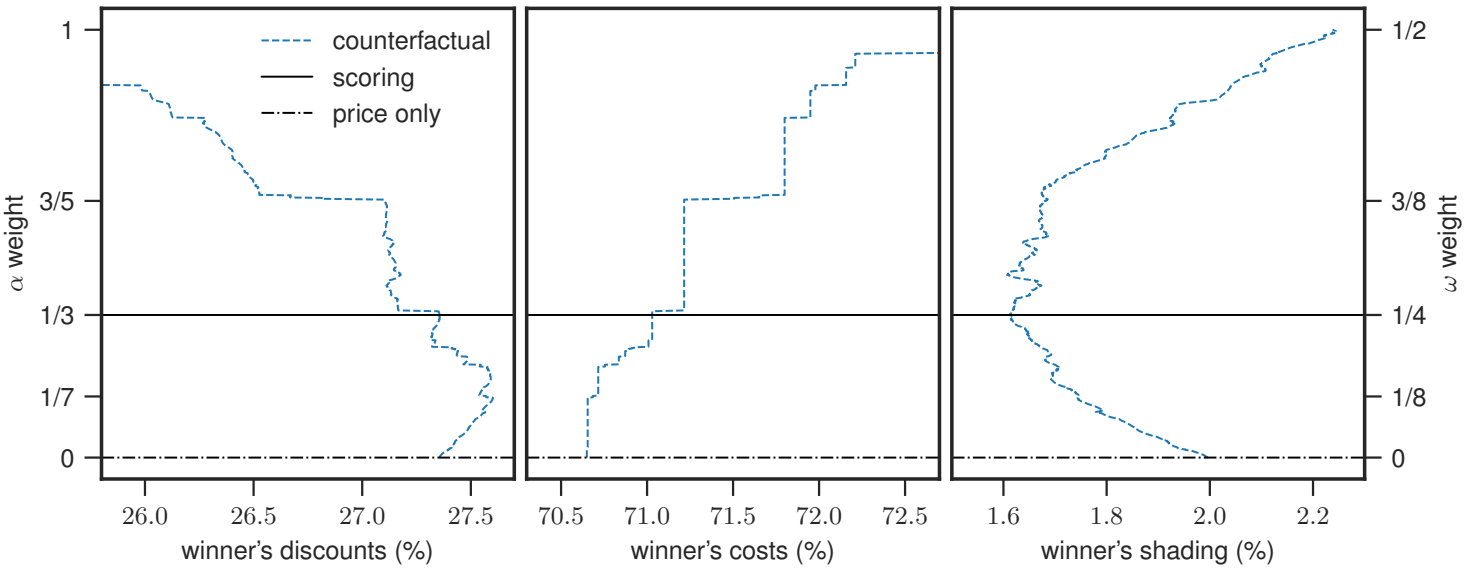



Figure 11: Counterfactuals

of discounts (27% as opposed to 26.81%), but this change is statistically insignificant. We stress that these results do not depend on the choice of $\beta$ or the exact shape of the cost function.

To put things into perspective, we also simulate counterfactual scoring auctions with other weights, with quality fixed at the level observed in the data. This can be interpreted as a temporary (or unexpected) change in the scoring weight away from $\alpha = 1/3$ or simply a limit when $\beta$ approaches 0. Thus, for a new scoring weight, $\alpha'$, we only have to calculate the new pseudo-type distribution $F_\rho(\cdot|\alpha')$ and re-evaluate bid shading. We present the results in Figure 11.

One can see that the relationship between the winner's expected production costs and the scoring weight is monotonic. This is not a surprise since higher weight means that high-quality firms have an advantage over low-cost firms. However, the relationship is not monotonic for the expected winner's bid shading. This leads to the observed discounts in the scoring auction being very close to those in the counterfactual price-only auction.

Our results indicate that the switch from the price-only auction to the scoring auction, through lower shading, has created an increase in quality without a visible increase in the price for the buyer (ACEA), which explains the empirical puzzle.

# 6 Discussion: consumer value of the reform

We conclude the analysis of the data with a back-of-the-envelope calculation comparing outcomes under Acea's reform and under the status quo absent any reform. This part of the analysis comes under two caveats. First, an exhaustive welfare analysis would require assigning a monetary value to the increased compliance on all parameters, but, since this would be hardly possible, we focus on a subset of specific outcomes. Second, the reform was characterized by several changes relative to the initial Acea plan of introducing a scoring rule system. Our analysis in this section compares the situation absent any reform to the one where the changes in price and performance are those implied by the most conservative estimates of the treatment at $t1$, as presented in section 3.

We start from the quality dimension. Here we focus on the quality of the service measured by one of the external measures of performance, the duration of long-lasting blackouts. Thus, we convert the estimate in column 2 of Table 10 into a measure of the number of blackout avoided per year: 43.272 hours on average per client. In the post-reform period, Acea has, on average, 1,597,066 customers, divided into 1,277,653 residential and 319,413 business customers. From the official statistics of the regulator (Arera),[27] we associate a cost of blackouts of 2.5 euro/hour for residential customers and 18.75 euro/hour for business customers. The result is that the reduction in blackouts implies a benefit of 6.623 million euros, 39 percent of which accrues to business customers and the rest to residential ones.

Next, we look at the safety dimension. Here we focus on the change in the probability of fatal accidents as implied by improvements in a subset of internal measures that most likely cover safety parameters.[28] Construction and maintenance jobs for electricity generation and transmission are among those with the highest incidence of workplace accidents, including deadly accidents.[29] The occurrence of such accidents has costs for society and Acea, and the

[27] See Arera's decision n. 172/07 of 12/07/2007.

[28] This subset of parameters is identified with an * in Table A.4. The Acea engineers decided on the selection of this subgroup of parameters.

[29] Electricity is widely recognized as a serious workplace hazard, exposing employees to electric shocks, burns, fires, and explosions. A search among local newspapers revealed that 4 workers had died in the last 15 years while performing jobs procured by Acea. The U.S. Bureau of Labor Statistics recorded 5,587 fatal electrical injuries between 1992 and 2013, an average of 254 fatal electrical injuries each year. Death was due to electrocution or fires caused by electricity, see Campbell and Dini (2015).

public ownership of Acea only increased its management's concern about these safety risks.

To map the relationship between changes in safety parameter compliance and the occurrence of fatal accidents, we use the statistical model used by Acea's engineers, which are known as *Heinrich's pyramid* and are often used by practitioners in the context of industrial systems to link accidents of different intensity.[30] The pyramid entails the following ratios: 1 fatal accident to 10 major accidents, to 30 minor accidents, to 600 material damages, and – finally – to 200,000-300,000 small deviations from safe behaviors. If we assume that each case of non-compliance in the safety parameters audited by Acea corresponds to a small deviation in the pyramid, we can estimate a lower bound for the policy benefit of €3-5 million per year. This is calculated as follows: in a typical audit, 33.08 parameters are assessed, 85.3 percent of which are part of the subset of safety-related parameters. There are, on average, 43 contracts a year, with an average duration of 250 working days (see Table 7). If the same rate of compliance observed across audits applies to every working day, then a conservative estimate of a 55 percent improvement in parameter compliance implies a reduction in about 163,000 small deviations per year. Using the 200,000-300,000 figure from the pyramid, this maps into a reduction in the probability of a fatal accident of 0.54-0.82 per year. Finally, considering an average of 4 workers on the worksite per day and taking the lowest bound of the OECD (2012) estimates of the "value of a statistical life" of €1.62 million per life saved,[31] the estimated benefit ranges between 3.5 and 5.3 million euro/year.[32]

[30] See Heinrich (1931), Bird and Germain (1986) and Goodman (2012). See also its usage by modern safety apps: http://safesiteapp.com/blog/safety/the-safety-triangle-explained/.

[31] The number of workers present on the worksite was estimated for us by the same expert engineers who estimated the variable *quick* described earlier. The OECD (2012) values are converted to the 2007 nominal euro. We shall also remark that our approach is quite conservative because we have employed the lowest OECD estimate of the value of a statistical life for benefits. Using the upper bound of the OECD estimate (€5.3 million), the benefits would be in the range of €11.55 -17.33. Furthermore, our benefit calculation excludes all the additional savings from reductions in non-fatal accidents associated with better safety practices and all improvements in quality associated with increased compliance with the quality parameters.

[32] This range is not our interval estimate, but the result of using the two bounds of 200,000 and 300,000 small deviations.

# 7 Conclusions

This paper has studied the merits of using past performance audits to spur quality and safety in executing public works. The evaluation of the evidence from a reform undertaken by Acea, a large utility company, has shown strong improvements in contract performance after the announcement of its intention to use past performance scores to award future contracts.

Improvements involve all parameters and suppliers, are long-lasting (for at least 10 years after the initial reform), and reflect higher service quality by the utility. Regarding prices, we find mixed empirical evidence and argue that, essentially, all of the improvements in contract execution come at zero or negative cost for the buyer, which is an empirical puzzle.

To explain the puzzle, we employ a structural model of a scoring auction with past performance demonstrating a possibility of such an outcome through compression of bid shading. The necessary ingredients combine sunk costs of investments with quality concerns (i.e., adverse selection), both inherent to the environment. A structural estimation of the model also confirms this hypothesis.

Several aspects remain open and offer room for future research; for example, how to optimize the parameter weights, how to discipline the rating for new entrants, how to structure the weights in the award criteria, and how to choose the optimal "memory" of the indicator (i.e., how long should be the window of time over which the RI is calculated, and how heavily should older information be discounted). Even the ideal speed at which the switch to a rating system should occur is an interesting but little-studied problem.

The policy relevance of our findings is significant. There is an ongoing policy debate in Europe and the US on using contractors' past performance in public procurement. In the US, with the Federal Acquisitions Streamlining Act of 1994, federal agencies started to record past contractor performance evaluations and to share them through common platforms for use in future contractor selection.[33] Interestingly, the EU follows a very different system, essentially barring the use of past performance except for extremely severe misbehavior sanctioned by the judiciary (Gordon and Racca, 2014). Indeed, using mechanisms based on

[33]The reform was pushed by Steven Kelman when he served as Administrator of the Office of Federal Procurement Policy in the Office of Management and Budget from 1993 to 1997, playing a lead role in the Administration's "reinventing government" effort that led, among other things, to the Federal Acquisition Streamlining Act of 1994 and the Federal Acquisition Reform Act 1995, see Kelman (1990).

past performance has been one of the most contentious issues in the debate leading up to the 2004 and 2014 EU Procurement Directives.[34] To this debate, our results offer a clear empirical illustration of the potential benefits of a rating mechanism based on objective and targeted past performance measures.

We shall explicitly discuss the issues of external validity and scalability, to which we devote a specific section in Appendix D. Following List (2020), there are four necessary conditions for external validity (SANS conditions). In our case, (i) representativeness of the sample with respect to the full population and representativeness of the sample with respect to the relevant variables for the study are satisfied, as most clearly displayed by our usage of control groups in different portions of our analysis; (ii) attrition rates and reasons for attrition and noncompliance do not apply to the public buyer; (iii) naturalness of the setting, choices, tasks and time frame observed is also clearly satisfied since it is a standardized setting like that of public procurement where most rules are identically shared by all public buyers within Italy (and, to a lesser extent, the EU).

Finally, the fourth condition, scalability, is possibly both the most important and the hardest to meet. In our case, it amounts to arguing whether the results presented would hold at a national or international level. Critical concerns are the ability to develop a large-scale monitoring system and the corruption risk. For both problems, however, the case of the PPIRS (Past Performance Information Retrival System) in the US public procurement shows that large-scale monitoring of contractors is feasible (and more so every day with the deployment of the internet of things, IoT). Moreover, corruption risk can be contained via transparency of the public procurement auctions. A similar European system would thus be advisable to assess the effects of a wider use of past performance to assign public procurement works. It would also harmonize the rules for public procurement in different countries and ensure the comparability of the works within the European Union for evaluating firms' performance.

[34] Curiously enough, current EU regulation acknowledges the importance of past performance for some types of procurement. For example, the European Research Council (ERC) funds research (including this study) through peer review, and the track record of the principal investigator is one of the main selection criteria.

Table 1: Comparison with U.S. Multi-Utility Providers

| Y2015 | ACEA | LADWP | ComEd | BGE | PECO |
|---|---|---|---|---|---|
| Total Employees (000) | 5.0 | 9.4 | 6.8 | 3.3 | 2.6 |
| Power Customers (mln) | 1.6 | 1.4 | 3.8 | 1.3 | 1.6 |
| Power Grid (000/miles) | 19 | 14 | 90 | 26 | 14 |
| Total Turnover (bln/$) | 3.2 (2.1) | 4.4 (3.3) | 4.9 | 3.1 | 3.0 |
| Power Supply (TWh) | 11 | 26 | 86* | 29* | 36* |
| Works on Power Grid Works (mln/$) | 206 | 318 | 2,400 | 500 | 475 |

*Note: Acea and LADWP figures on employees and turnover include the water business too. BGE and PECO figures on employees and turnover include the gas business too. All values are for 2015. Values with a * symbol are estimates: the supply is estimated proportionally to the customers out of the total supply of all Exelon subsidiaries (195TWh). The values in parenthesis refer to power only for the total Turnover (bln $).*

Table 2: Summary Statistics for the Acea's Audits (Internal Performance Measures)

| Parameters | Share Compliant Parameters | | | | Number of |
|---|---|---|---|---|---|
| Category | price-only | grace | scoring | hybrid | |
| Air works | | .98 | 1 | 1 | 253 |
| Customer relationship mgnt | 1 | .93 | 1 | 1 | 154 |
| Documentation | .35 | .66 | .83 | .93 | 61,782 |
| Equipment and machinery | .7 | .93 | .96 | .95 | 51,032 |
| H.T. works site controls | | .8 | .96 | .96 | 2,858 |
| Personnel | .32 | .69 | .88 | .96 | 25,266 |
| Underground works | .4 | .69 | .85 | .81 | 18,245 |
| Works execution | .19 | .84 | .96 | .98 | 36,977 |
| Works on joints | 1 | .96 | .94 | .86 | 3,238 |
| Works on transformer station | 1 | 1 | 1 | 1 | 501 |
| Works site regularity | .11 | .62 | .83 | .94 | 72,475 |
| Works site safety | .31 | .76 | .91 | .96 | 93,076 |

*Note: The 136 parameters audited are partitioned into the 12 categories. For each of the four subperiods in which the sample is split, the share of compliant parameters indicates the share of scores equal to 1, over the sum of all scores that are either zero or 1.*

Table 3: Breakpoints in the Internal Performance Measure

| | F-stat breaks | 5 unknown breaks | |
|---|---|---|---|
| Number of breaks | 4 | 5 | |
| Dates of the breaks: | | | |
| Date 1 | 2007m12 | 2007m12 | t1 |
| Date 2 | 2008m5 | 2008m5 | t2 |
| Date 3 | 2008m8 | 2008m8 | t3+1 |
| Date 4 | 2009m2 | 2009m2 | t5+1 |
| Date 5 | - | 2009m8 | t5+7 |

*Note: The table reports the results of Bai-Perron tests. The variable is the monthly weighted average compliance, measured on all audited parameters. We indicate as $ty + x$ a breakpoint taking place $x$ months after Acea's announcement date $ty$, where $y = 1, ..., 5$. The test criterion used is that of sequential F-statistic determined breaks. Results are identical with the significant F-statistic largest breaks criterion.*

Table 4: Acea's Announcements and Supplier Compliance

| | (1) | (2) | (3) | (4) |
|---|---|---|---|---|
| t1 | 0.184*** | 0.184*** | 0.177*** | 0.607*** |
| | (0.029) | (0.028) | (0.031) | (0.128) |
| t2 | 0.088*** | 0.092*** | 0.068*** | 0.003 |
| | (0.020) | (0.019) | (0.019) | (0.073) |
| t3+1 | 0.139*** | 0.143*** | 0.139*** | -0.017 |
| | (0.015) | (0.015) | (0.014) | (0.061) |
| t5+1 | 0.185*** | 0.142*** | 0.073*** | 0.015 |
| | (0.010) | (0.010) | (0.012) | (0.039) |
| Safety parameters | 0.127*** | 0.104*** | 0.044*** | 0.032*** |
| | (0.012) | (0.012) | (0.011) | (0.011) |
| Public Illumination | -0.016*** | -0.030*** | -0.005 | 0.000 |
| | (0.006) | (0.007) | (0.028) | (0.028) |
| Firm FEs | | ✓ | ✓ | ✓ |
| Contract FEs | | | ✓ | ✓ |
| Month FEs | | | | ✓ |
| N | 4765 | 4765 | 4765 | 4765 |

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

*Note: The dependent variable is the average compliance (weighted with the RI parameter weights) for each firm-contract-month triplet. Regarding the regressors,* $t1$ *is a dummy variable equal to 1 from* $t1$ *onward and zero before then.* $t2, t3+1, t5+1$ *are constructed analogously. We indicate as* $ty+x$ *a breakpoint taking place* $x$ *months after the Acea's announcement date* $ty$*, where* $y = 1, ..., 5$*. All regressions also control for the Job type – the proportion of contracts classified as public illumination – calculated among those parameters audited in the firm-contract-month triplet. Standard errors are in parentheses.* * $(p < 0.10)$, ** $(p < 0.05)$, *** $(p < 0.01)$.

Table 5: Entrants, stayers and dropouts

| | entrant | stayer | dropout |
|---|---|---|---|
| number | 2 | 19 | 4 |
| avg. offered discount (before scoring) | 12.08 | 13.07 | 11.37 |
| avg. offered discount (scoring) | 19.65 | 19.41 | . |
| avg. reputation index (scoring) | 86.84 | 87.57 | . |
| auctions participated | 155 | 3524 | 96 |
| scoring auctions participated | 36 | 412 | 0 |
| auctions won | 13 | 326 | 6 |
| scoring auctions won | 2 | 28 | 0 |
| % auctions won vs. participated | 8.39 | 9.25 | 6.25 |
| % scoring auctions won vs. participated | 5.56 | 6.8 | . |

Table 6: Regression of offered discount on reputation index with and without bidders rank according to the score (the strongest bidder has rank 1, second strongest has rank 2, e.t.c.)

| | (1) | (2) | (3) | (4) | (5) | (6) |
|---|---|---|---|---|---|---|
| Reputation index | 0.442*** | 0.478*** | 0.478*** | -0.320** | -0.355*** | -0.361*** |
| | (0.078) | (0.069) | (0.070) | (0.128) | (0.069) | (0.067) |
| Rank | | | | -3.272*** | -3.575*** | -3.650*** |
| | | | | (1.134) | (0.607) | (0.587) |
| Reputation index × Rank | | | | 0.017 | 0.019*** | 0.019*** |
| | | | | (0.013) | (0.007) | (0.007) |
| Observations | 495 | 495 | 495 | 495 | 495 | 495 |
| $R^2$ | 0.06 | 0.31 | 0.33 | 0.58 | 0.89 | 0.90 |
| Auction FEs | | ✓ | ✓ | | ✓ | ✓ |
| Lot FEs | | | ✓ | | | ✓ |

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 7: Summary Statistics for the Auctions Data

| | Acea | | | Control | | |
|---|---|---|---|---|---|---|
| | Mean | SD | N | Mean | SD | N |
| Panel (a): price-only period | | | | | | |
| Winning discount | 21.93 | 10.41 | 196 | 18.52 | 9.600 | 11466 |
| Winning Bid | 546339.9 | 280667.9 | 6 | 1317302.6 | 4487640.2 | 11466 |
| Length (days) | 403.6 | 172.2 | 199 | 453.2 | 752.0 | 11732 |
| Num. Bidders | 11.09 | 4.133 | 45 | . | . | 0 |
| Public Illumination | 0.183 | 0.388 | 218 | 0.0764 | 0.266 | 13186 |
| Central Region | 1 | 0 | 218 | 0.189 | 0.392 | 13186 |
| Municipal Firm | 1 | 0 | 218 | 0.362 | 0.481 | 13186 |
| Panel (b): grace period | | | | | | |
| Winning discount | 18.68 | 10.48 | 142 | 21.07 | 11.97 | 12956 |
| Winning Bid | 557254.6 | 395848.6 | 42 | 1510379.5 | 4885153.9 | 12956 |
| Length (days) | 392.7 | 146.9 | 130 | 539.6 | 953.5 | 13220 |
| Num. Bidders | 11.04 | 4.429 | 125 | . | . | 0 |
| Public Illumination | 0.243 | 0.431 | 152 | 0.0776 | 0.268 | 15921 |
| Central Region | 1 | 0 | 152 | 0.194 | 0.396 | 15921 |
| Municipal Firm | 1 | 0 | 152 | 0.399 | 0.490 | 15921 |
| Panel (c): scoring period | | | | | | |
| Winning discount | 28.84 | 7.202 | 36 | 24.75 | 13.27 | 1273 |
| Winning Bid | 417061.7 | 203631.7 | 25 | 1631297.7 | 5758230.8 | 1273 |
| Length (days) | 416.3 | 101.4 | 36 | 590.4 | 989.7 | 1287 |
| Num. Bidders | 13.59 | 2.340 | 32 | . | . | 0 |
| Public Illumination | 0.611 | 0.494 | 36 | 0.0992 | 0.299 | 1623 |
| Central Region | 1 | 0 | 36 | 0.185 | 0.388 | 1623 |
| Municipal Firm | 1 | 0 | 36 | 0.384 | 0.486 | 1623 |
| Panel (d): hybrid period | | | | | | |
| Winning discount | 28.65 | 7.803 | 207 | 24.45 | 13.53 | 15380 |
| Winning Bid | 878906.9 | 3237609.7 | 175 | 1967913.5 | 6017188.4 | 15380 |
| Length (days) | 432.0 | 173.0 | 240 | 736.7 | 1318.8 | 15634 |
| Num. Bidders | 11.85 | 4.623 | 265 | . | . | 0 |
| Public Illumination | 0.293 | 0.456 | 283 | 0.0787 | 0.269 | 20890 |
| Central Region | 1 | 0 | 283 | 0.194 | 0.395 | 20890 |
| Municipal Firm | 1 | 0 | 283 | 0.471 | 0.499 | 20890 |

*Note: "Control" sample consists of auctions held by companies other than Acea. Length is the contractual duration of the contract in days (a contractual duration of 1 year corresponds to 250 working days), Num. Bids is the number of bids submitted, Public Illumination is a dummy equal to 1 if Acea classifies the contract type as public illumination and zero if it is classified as work on electrical substations, Central Region is a dummy equal to 1 if the CA is located in one of Italy's Center regions and zero otherwise and Municipal Firm is a dummy equal to 1 if the CA is a multi-utility company that is (at least partially) owned by the municipality in which it operates. The last two variables are not reported for Panel (c) as they are both always equal to 1 for Acea's auctions.*

Table 8: Baseline Price Estimates

| | (1) | (2) | (3) | (4) |
|---|---|---|---|---|
| Acea * grace (t1-t5) | 6.46** | 7.66*** | 6.96** | 8.62*** |
| | (2.94) | (2.79) | (3.11) | (2.30) |
| Acea * grace (>t5) | -5.62 | -4.70 | -5.69 | -3.40 |
| | (4.97) | (5.14) | (4.93) | (4.78) |
| Acea * scoring | 4.15 | 4.01 | 5.17 | 4.69* |
| | (3.15) | (3.09) | (3.68) | (2.61) |
| Acea * hybrid | 7.20* | 7.17* | 8.69** | 7.82** |
| | (3.91) | (4.05) | (3.64) | (3.66) |
| ABA | -7.18*** | -7.32** | -6.23** | -7.66** |
| | (2.54) | (3.06) | (2.95) | (3.06) |
| Log(1+N.bidders) | 1.74 | 1.33 | 2.08 | 1.31 |
| | (3.06) | (3.15) | (2.48) | (3.07) |
| Observations | 41482 | 13505 | 2608 | 11294 |
| $R^2$ | 0.54 | 0.63 | 0.58 | 0.64 |
| Buyer&Year FE | ✓ | ✓ | ✓ | ✓ |
| Size, region, category | | ✓ | ✓ | ✓ |
| Control sample | All | All | Center | North&South |

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

*Note: the dependent variable is the winning discount. The sample includes auctions by Acea (treatment group) and all other contracting authorities (control group). The first three columns report estimates for the model in equation (4), while the last three columns report estimates for the model in equation (**??**). For each model, the model specification gradually expands the set of contract characteristics included as controls: award criterion (columns 1 and 4), also fixed effects for four levels of the reserve price (columns 2 and 5), and also a dummy for whether the contract is for public illumination (columns 3 and 6). Standard error clusters by year and CA are reported in parentheses.* * $(p < 0.10)$, ** $(p < 0.05)$, *** $(p < 0.01)$.

Table 9: Summary Statistics for the Regulators' Reports (External Performance Measures)

| | Mean | St. Dev. | Median | Min | Max | N |
|---|---|---|---|---|---|---|
| Long-lasting blackouts (num/LV lines) | 2.43 | 2.50 | 1.76 | 0.00 | 24.00 | 1,433 |
| Long-lasting blackouts duration (min) | 94.03 | 134.01 | 49.40 | 0.00 | 960.00 | 1,419 |
| Short-lasting blackouts (num/LV lines) | 2.70 | 3.90 | 1.84 | 0.00 | 62.00 | 1,286 |
| Programmed power cuts (num/LV lines) | 0.60 | 1.24 | 0.30 | 0.00 | 29.47 | 1,431 |
| Duration Programmed power cuts (min/LV lines) | 65.63 | 113.96 | 31.18 | 0.00 | 988.53 | 1,428 |
| Low voltage users (thousands) | 365.45 | 815.30 | 6.42 | 0.00 | 4663.64 | 1,642 |
| Water Leakage (Water users ('000s) | 899.80 | 1061.58 | 491.33 | 118.63 | 4341.24 | 253 |

*Note: Long-lasting blackouts and Blackouts duration are, respectively, the average number and the average duration (in minutes) of long-lasting blackouts per user, Short-lasting blackouts is the average number of short-lasting blackouts per user, Programmed power cuts and Duration programmed power cuts are, respectively, the average number and average duration (in minutes) of programmed power cuts to the low voltage grid per user, Low voltage users is the total number of low voltage grid customers (in thousands), WaterLeakage is the percentage incidence of water leakage over water inflow (Water Leakage= (Inflow-Outflow)/Inflow), while Water users is the total number of customers (in thousands).*

Table 10: Estimates for the External Performance Measures: Electricity and Water Sectors

| | (1) Long-lasting blackouts | (2) Length long-lasting blackouts | (3) Short-lasting blackouts | (4) Programmed power cuts | (5) Length programmed power cuts | (6) Water leakage | (7) Water leakage |
|---|---|---|---|---|---|---|---|
| $\beta_4$ | -0.325* | -43.272** | -0.922** | 0.141 | 19.839* | -0.003 | 0.009 |
| | (0.163) | (13.350) | (0.296) | (0.074) | (9.154) | (0.010) | (0.015) |
| Observations | 386 | 386 | 298 | 386 | 386 | 253 | 59 |
| $R^2$ | 0.84 | 0.57 | 0.83 | 0.72 | 0.79 | 0.82 | 0.89 |
| Buyer&Year FE | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| Sample | All | All | All | All | All | All | Pop $\geq$ 1mln |

*Note: The table reports the difference-in-difference estimates for the available external performance measures. In the first five columns, the outcomes cover the electricity distribution sector, whereas the last two columns regard the water distribution sector. ACEA is the treated unit, and the treatment is the interaction term of indicators for ACEA and post-year 2007. The control units for the electricity sector include all the distributors with at least 200 thousand clients. For the water sectors, the control units include all the distributors (column 6) or only those in charge of geographical areas with at least 1 million customers (column 7). Robust standard errors in parentheses. Significance: *** p<0.01, ** p<0.05, * p<0.10.*

Table 11: Counterfactuals

| design (weight) | scoring ($\alpha = 1/3,\ \omega = 1/4$) | | price only ($\alpha = \omega = 0$) | |
|---|---|---|---|---|
| | all | winner's | all | winner's |
| | (1) | (2) | (3) | (4) |
| quality (%) | 87.78 (0.11) | 89.92 (0.4) | 87.78 (0.11) | 87.81 (0.38) |
| discount (%) | 19.48 (0.16) | 26.81 (0.3) | 19.52 (0.16) | 27.03 (0.23) |
| cost (%) | 79.43 (0.16) | 71.59 (0.31) | 79.43 (0.16) | 71.14 (0.26) |
| markup (%) | 1.43 (0.03) | 2.26 (0.16) | 1.39 (0.03) | 2.63 (0.15) |
| shading (%) | 1.09 (0.02) | 1.6 (0.11) | 1.05 (0.02) | 1.83 (0.1) |
| shading ($) | 5470.24 (127.75) | 8206.55 (508.14) | 5320.39 (126.4) | 9349.92 (506.16) |
| cost ($) | 390299.35 (3619.5) | 366266.7 (1512.1) | 390299.35 (3619.5) | 363169.38 (1190.45) |

# References


**Andrews, Isaiah, and Daniel Barron.** 2016. "The Allocation of Future Business: Dynamic Relational Contracts with Multiple Agents." *American Economic Review*, 106(9): 2742–59.

**Andreyanov, Pasha.** 2019. *Essays on Prior-Free Mechanism Design.* University of California, Los Angeles.

**Andreyanov, Pasha, and Grigory Franguridi.** 2021. "Nonparametric inference on counterfactuals in first-price auctions."

**Andreyanov, Pasha, Ilya Krasikov, and Alex Suzdaltsev.** 2023. "Scoring and Favoritism in Optimal Procurement Design." Working Paper.

**Asker, John, and Estelle Cantillon.** 2008. "Properties of scoring auctions." *The RAND Journal of Economics*, 39(1): 69–85.

**Asker, John, and Estelle Cantillon.** 2010. "Procurement when price and quality matter." *The Rand journal of economics*, 41(1): 1–34.

**Bajari, Patrick, and Steven Tadelis.** 2001. "Incentives versus Transaction Costs: A Theory of Procurement Contracts." *RAND Journal of Economics*, 32(3): 387–407.

**Bajari, Patrick, Robert S. McMillan, and Steven Tadelis.** 2009. "Auctions Versus Negotiations in Procurement: An Empirical Analysis." *Journal of Law, Economics and Organization*, 25(2): 372–399.

**Bar-Isaac, Heski, and Steven Tadelis.** 2008. "Seller Reputation." *Foundations and Trends(R) in Microeconomics*, 4(4): 273–351.

**Beuve, Jean, and Stephane Saussier.** 2021. "Renegotiations and Renewals of Public Contracts." *Review of Industrial Organization*, 59(3): 461–482.

**Bird, Frank E, and George L Germain.** 1986. *Practical loss control leadership.* Loganville, Ga. Institute Publications.

**Board, Simon.** 2011. "Relational Contracts and the Value of Loyalty." *American Economic Review*, 101(7): 3349–67.

**Bosio, Erica, Simeon Djankov, Edward L Glaeser, and Andrei Shleifer.** 2020. "Public Procurement in Law and Practice." National Bureau of Economic Research Working

Paper 27188.

**Burguet, Roberto, Juan Jose Ganuza, and Ester Hauk.** 2012. "Limited Liability and Mechanism Design in Procurement." *Games and Economic Behavior*, 76(1): 15–25.

**Butler, Jeffrey V., Enrica Carbone, Pierluigi Conzo, and Giancarlo Spagnolo.** 2020. "Past Performance and Entry in Procurement: an Experimental Investigation." *Journal of Economic Behavior and Organizations*, 173: 179–195.

**Cabral, Luis M. B., and Miguel Villas-Boas.** 2005. "Bertrand Supertraps." *Management Science*, 51(4): 599–613.

**Calzolari, Giacomo, and Giancarlo Spagnolo.** 2009. "Relational Contracts and Competitive Screening." CEPR Discussion Papers 7434.

**Camboni, R, L Corazzini, S Galavotti, and P Valbonesi.** 2023. "Bidding on Price and Quality: An Experiment on the Complexity of Scoring Rule Auctions." *The Review of Economics and Statistics*, forthcoming.

**Campbell, Richard B., and David A. Dini.** 2015. "Occupational Injuries from Electrical Shock and Arc Flash Events." Fire Protection Research Foundation Final Report 1.

**Carril, Rodrigo, Andres Gonzalez-Lira, and Michael Walker.** 2020. "Rules Versus Discretion in Public Procurement." *mimeo.*

**Carril, Rodrigo, Andres Gonzalez-Lira, and Michael Walker.** 2022. "Competition under Incomplete Contracts and the Design of Procurement Policies." *mimeo.*

**Chassang, Sylvain, and Juan Ortner.** 2016. "Collusion in Auctions with Constrained Bids: Theory and Evidence from Public Procurement." *Journal of Political Economy*, 127(5): 2269–2300.

**Che, Yeon-Koo.** 1993. "Design competition through multidimensional auctions." *The RAND Journal of Economics*, 668–680.

**Chong, Eshien, Carine Staropoli, and Anne Yvrande-Billon.** 2014. "Auction versus negotiation in public procurement: Looking for empirical evidence." *The manufacturing markets, legal, political and economic dynamics*, Cambridge University Pres.

**Colonnelli, Emanuele, and Mounu Prem.** 2021. "Corruption and Firms: Evidence from Randomized Audits in Brazil." *The Review of Economic Studies*, forthcoming.

**Coviello, Decio, Luigi Moretti, Giancarlo Spagnolo, and Paola Valbonesi.** 2018.

"Court Efficiency and Procurement Performance." *The Scandinavian Journal of Economics*, 120.

**D'Alpaos, C, M Moretto, P Valbonesi, and S Verglli.** 2013. "Time overruns as opportunistic behaviour in public procurement." *Journal of Economics*, 110(1): 25–42.

**Decarolis, Francesco.** 2014. "Awarding Price, Contract Performance and Bids Screening: Evidence from Procurement Auctions." *American Economic Journal: Applied Economics*, 6(1): 108–132.

**Decarolis, Francesco, Raymond Fisman, Paolo Pinotti, and Silvia Vannutelli.** 2020. "Rules, Discretion, and Corruption in Procurement: Evidence from Italian Government Contracting." *mimeo.*

**Djankov, Simeon, Oliver Hart, Caralee McLiesh, and Andrei Shleifer.** 2008. "Debt Enforcement around the World." *Journal of Political Economy*, 116(6): 1105–1149.

**Duflo, Esther, Michael Greenstone, Rohini Pande, and Nicholas Ryan.** 2013. "Truth-telling by Third-party Auditors and the Response of Polluting Firms: Experimental Evidence from India*." *The Quarterly Journal of Economics*, 128(4): 1499–1545.

**Duflo, Esther, Michael Greenstone, Rohini Pande, and Nicholas Ryan.** 2018. "The Value of Regulatory Discretion: Estimates From Environmental Inspections in India." *Econometrica*, 86(6): 2123–2160.

**Giacomelli, Silvia, and Carlo Menon.** 2016. "Does weak contract enforcement affect firm size? Evidence from the neighbour's court." *Journal of Economic Geography*, 17(6): 1251–1282.

**Goodman, William M.** 2012. "Measuring Changes in the Distribution of Incident-Outcome Severities: A Tool for Safety Management." *Case Studies in Business, Industry and Government Statistics*, 5(1): 32–43.

**Gordon, Daniel I., and Gabriella M. Racca.** 2014. "Integrity Challenges in the EU and U.S. Procurement Systems." *Integrity and Efficiency in Sustainable Public Contracts. Corruption, Conflict of Interest, Favoritism and Inclusion of Non-Economic Criteria in Public Contracts, Gabriella M. Racca & Christopher R. Yukins, eds., Bruylant.*

**Guerre, Emmanuel, Isabelle Perrigne, and Quang Vuong.** 2000. "Optimal nonparametric estimation of first-price auctions." *Econometrica*, 68(3): 525–574.

**Haile, Philip, Han Hong, and Matthew Shum.** 2003. "Nonparametric tests for common values at first-price sealed-bid auctions."

**Heinrich, Herbert William.** 1931. *Industrial accident prevention: a scientific approach.* McGraw-Hill.

**Hiriart, Yolande, David Martimort, and Jerome Pouyet.** 2004. "On the Optimal Use of Ex Ante Regulation and Ex Post Liability." *Economics Letters*, 84(2): 231–235.

**Holmstrom, Bengt.** 1999. "Managerial incentive problems: A dynamic perspective." *The Review of Economic Studies*, 66(1): 169–182.

**Jofre-Bonet, Mireia, and Martin Pesendorfer.** 2003. "Estimation of a Dynamic Auction Game." *Econometrica*, 71(5): 1443–1489.

**Jones, M Chris.** 1992. "Estimating densities, quantiles, quantile densities and density quantiles." *Annals of the Institute of Statistical Mathematics*, 44(4): 721–727.

**Kang, Karam, and Robert A Miller.** 2021. "Winning by Default: Why is There So Little Competition in Government Procurement?" *The Review of Economic Studies*, forthcoming.

**Kelman, Steven.** 1990. *Procurement and Public Management: The Fear of Discretion and the Quality of Government Performance.* AEI Studies.

**Klein, Benjamin, and Keith B Leffler.** 1981. "The Role of Market Forces in Assuring Contractual Performance." *Journal of Political Economy*, 89(4): 615–41.

**Kolstad, Charles D., Thomas S. Ulen, and Gary V. Johnson.** 1990. "Ex Post Liability for Harm vs. Ex Ante Safety Regulation: Substitutes or Complements?" *The American Economic Review*, 80(4): 888–901.

**Krasnokutskaya, Elena.** 2011. "Identification and estimation of auction models with unobserved heterogeneity." *The Review of Economic Studies*, 78(1): 293–327.

**Krishna, Vijay.** 2009. *Auction theory.* Academic press.

**Laffont, Jean-Jacques, Isabelle Perrigne, Michel Simioni, and Quang Vuong.** 2020. "Econometrics of scoring auctions." In *Essays in honor of Cheng Hsiao*. Vol. 41, 287–322. Emerald Publishing Limited.

**Lewis-Faupel, Sean, Yusuf Neggers, Benjamin A. Olken, and Rohini Pande.** 2016. "Can Electronic Procurement Improve Infrastructure Provision? Evidence from Public

Works in India and Indonesia." *American Economic Journal: Economic Policy*, 8(3): 258–83.

**Liebman, Jeffrey B., and Neale Mahoney.** 2016. "Do Expiring Budgets Lead to Wasteful Year-End Spending? Evidence from Federal Procurement." *American Economic Review*, 107(11): 3510–49.

**List, John.** 2006. "The Behavioralist Meets the Market: Measuring Social Preferences and Reputation Effects in Actual Transactions." *Journal of Political Economy*, 114(1): 1–37.

**List, John A.** 2020. "Non est Disputandum de Generalizability? A Glimpse into The External Validity Trial." National Bureau of Economic Research Working Paper 27535.

**List, John A., and David Reiley.** 2008. "Field Experiments in Economics: Palgrave Entry." In *The New Palgrave Dictionary of Economics.* , ed. Steven N. Durlauf and Lawrence E. Blume. Palgrave Macmillan.

**Li, Tong, Isabelle Perrigne, and Quang Vuong.** 2000. "Conditionally independent private information in OCS wildcat auctions." *Journal of Econometrics*, 98(1): 129–161.

**Lopomo, Giuseppe, Nicola Persico, and Alessandro Villa.** 2023. "Optimal Procurement With Quality Concerns." *American Economic Review*, forthcoming.

**Manelli, Alejandro M., and Daniel R. Vincent.** 1995. "Optimal Procurement Mechanisms." *Econometrica*, 63(3): 591–620.

**Marion, Justin.** 2017. "Affirmative Action Exemptions and Capacity Constrained Firms." *American Economic Journal: Economic Policy*, 9(3): 377–407.

**OECD.** 2012. *Mortality Risk Valuation in Environment, Health and Transport Policies.* OECD Publishing.

**Olken, Benjamin A.** 2007. "Monitoring Corruption: Evidence from a Field Experiment in Indonesia." *Journal of Political Economy*, 115(2): 200–249.

**Rose-Ackerman, Susan.** 1991. "Regulation and the Law of Torts." *The American Economic Review*, 81(2): 54–58.

**Saussier, Stephane.** 2000. "Transaction costs and contractual incompleteness: the case of Electricite de France." *Journal of Economic Behavior and Organization*, 42(2): 189–206.

**Shavell, Steven.** 1984. "A Model of the Optimal Use of Liability and Safety Regulation." *The RAND Journal of Economics*, 15(2): 271–280.

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

**Tadelis, Steven.** 2016. " The Economics of Reputation and Feedback Systems in E-Commerce Marketplaces." *IEEE Internet Computing*, 20(1): 12–19.

**Zheng, Charles Z.** 2001. "High Bids and Broke Winners." *Journal Economic Theory*, 100: 129–171.

# Online Appendix

## A Data

The data used in the paper come from three main sources plus several ancillary ones. The Audit data come directly from the firm implementing the reform, Acea (https://www.gruppo.acea.it/en). They were released to us for research and study purposes. The Auction data come from the database on public works of a private company, http://www.telemat.it/. This is a major information entrepreneur (IE), and its main activity is selling information about public contracts to construction firms. For the subset of auctions held by Acea, we also have the internal Acea's records regarding these auctions. The Regulatory Reports data come from the public authority, the yearly reports of the Italian Regulatory Authority for Energy, Networks, and Environment (ARERA, https://www.autorita.energia.it/it/inglese/). Additional data were obtained from the Observatory on Public Contracts of the Italian Anticorruption Authority http://www.anac.it, from which we take the data on time delays and cost overruns in contract execution. Furthermore, for the analysis of the consumer value of the reform, the value of statistical life figures come from the OECD (https://www.oecd.org/environment/mortalityriskvaluationinenvironmenthealthandtransportpolicies.htm), while those for the economic cost of 1 hour of a blackout, separately for business and residential customers come from Table 11 in the AREA's decision n. 172/07 of 12/07/2007.

## B Theory

### B.1 Second order conditions

Note that the firm's action is 2-dimensional. Instead of picking $s, q$ simultaneously, she can optimize over $q$ conditional on $s$. This is equivalent to plugging equation (8) into the utility:

$$U_i(q(s,\gamma_i),s) = (\theta_i + \alpha(1-\frac{1}{\beta})(\gamma_i + (\alpha G(s))^{\frac{1}{\beta}}) + \frac{\alpha}{\beta}\gamma_i - s)G(s),$$

which she can then maximize over $s$. The optimal score can be derived via the envelope conditions:

$$s(\theta_i, \gamma_i) = \theta_i + \alpha(1 - \frac{1}{\beta})(\gamma_i + (\alpha H(\rho(\theta_i, \gamma_i))^{\frac{1}{\beta}}) + \frac{\alpha}{\beta}\gamma_i - \int_0^{\rho(\theta_i,\gamma_i)} H(z)dz/H(\rho(\theta_i, \gamma_i)).$$

We can then invoke a standard mechanism design argument to show that the second-order conditions are satisfied. Indeed, if the agent reports a score associated with a different type $\theta'$ and chooses a quality that is optimal for that score, her utility will be equal to

$$(\theta_i - \theta')H(\rho(\theta', \gamma_i)) + \int_0^{\rho(\theta',\gamma_i)} H(z)dz$$

which has a unique critical point $\theta' = \theta_i$. Finally, the second derivative at the critical point is equal to $-2\frac{\partial H}{\partial \rho}\frac{\partial \rho}{\partial \theta}$, which is strictly negative, thus the second order conditions are satisfied.

## B.2 Univariate types example

Assuming monotonicity of score in type, $F_s(s(\theta)) = \theta$, we can compute the equilibrium quality $q = \alpha\theta$ and the equilibrium pseudo-type $\rho = (1 + \alpha^2)\theta$.

On the other hand, the total equilibrium profit of the firm is equal to $\int_0^\theta xdx = \theta^2/2$ by the envelope conditions. It does not depend on $\alpha$ or the sunk nature of costs because the firm with the highest type always wins, e.g., Krishna (2009). The auction profits are, therefore, equal to the total profits plus the investment costs: $(1 + \alpha^2)\theta^2/2$.

With the equilibrium auction profits at hand, we can compute the profit margins $(\rho - s) = (1 + \alpha^2)\theta/2$. The equilibrium score is, therefore, linear in type $s = (1 + \alpha^2)\theta/2$, and the discount is $d = (1 - \alpha^2)\theta/2$.

## B.3 Bivariate types example

For both weights $\alpha = 0, 1$, it is true that $\mathbb{E}\theta|\rho = \rho - \underline{\rho}$. The analytical expression for the expected winning firm's type is, therefore the same:

$$\int \mathbb{E}\theta|\rho dH^2(\rho) = \int 1 - H^2(\rho)d\rho.$$

The expression for the expected winning firm's bid shading, on the other hand, is

$$\int \frac{\int_{\underline{\rho}}^{\rho} H(x)}{H(\rho)} dH^2(\rho) = 2 \int H(\rho)(1 - H(\rho))d\rho.$$

The expected winning firm's discount is, therefore their difference.

# C Structural

## C.1 Specification tests

We test whether $s_1 - s_2$ is independent of $s_3 - s_4$, where $\{s_i\}_{i=1}^4$ are four scores, randomly picked in every auction, to validate an additive model of heterogeneity.[35] It is soundly not rejected, according to Pearson ($r = 0.004$ $p = 0.8$) and Spearman ($r = 0.01$ $p = 0.58$) correlation tests with 3000 randomly picked quadruples of scores, see Figure A.1.

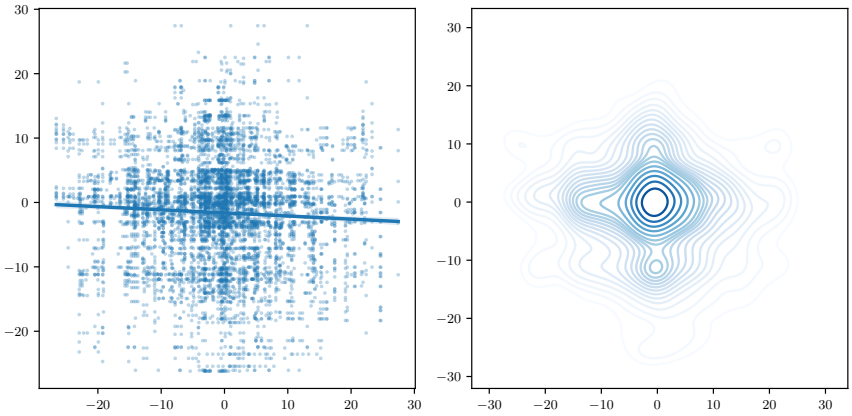



Figure A.1: Scatterplot and contourplot of score differences $s_1 - s_2$ and $s_3 - s_4$.

We apply the analog of the sup-norm test, suggested by Haile, Hong and Shum (2003), to compare the distributions of score residuals between auctions with different numbers of bidders. To test whether the distributions are identical, the statistic is formed

$$\delta = \sum_{10}^{16} \sup_v \{\hat{F}_{n+1}(v) - \hat{F}_n(v)\},$$

[35] Similarly, we could validate a multiplicative model by testing the independence of score ratios $s_1/s_2$ and $s_3/s_4$.

where $\hat{F}_n(v)$ is the empirical cdf of score residuals with $n$ bidders. The asymptotic distribution of the statistic is achieved via sub-sampling, and the IPV hypothesis is soundly not rejected ($\delta = 1.44$, $p = 0.52$), see Figure A.2.

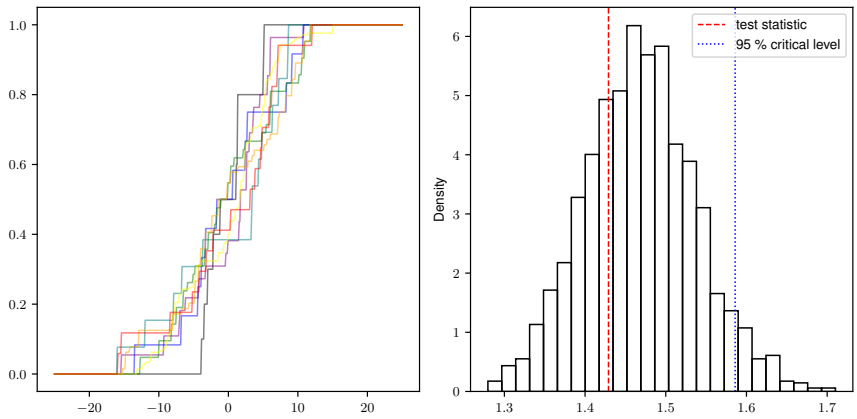



Figure A.2: Empirical CDF's of score residuals (left) with different numbers of bidders and the distribution of the $\delta$ statistic (right) obtained via sub-sampling.

## C.2 Quantile approach

We will rewrite the optimality conditions in the quantile form to make the optimality conditions more palatable. For this, we will need auxiliary functions that only depend on the probabilities $p_n$:

$$C(u) = \sum_{n=1}^{N} p_n u^{n-1}, \quad c(u) = \frac{C(u)}{C'(u)}.$$

Denote $Q_s(\cdot|\alpha), q_s(\cdot|\alpha)$ to be the equilibrium quantile function and density of the score, while $Q_\rho(\cdot|\alpha)$ the quantile function of the pseudo-type. Using the trivial identities $F_s(Q_s(u|\alpha)|\alpha) = u$ and $F_\rho(Q_\rho(u|\alpha)|\alpha) = u$, we can recast the first order conditions as

$$Q_\rho(u|\alpha) = Q_s(u|\alpha) + q_s(u|\alpha)c(u), \tag{11}$$

and the envelope conditions as

$$Q_s(u|\alpha) = Q_\rho(u|\alpha) - \frac{\int_0^u C(x)dQ_\rho(x|\alpha)}{C(u)}. \tag{12}$$

## C.3 Estimation

For convenience, we will divide the observed scores by 0.75, so that the default scoring rule has the form $s = \alpha q + d$, with $\alpha = 1/3$.

As is common in the literature, we will first residualize the observed scores to eliminate the auction-fixed effects. Note that this process does not change the ranking of firms within the auction, that is, the firm with the highest residual is also the winner in the data. Note also that while the location of the distribution of $\delta_i$ is not identified, it does not matter due to the linear scalability of the model.

The auction fixed effects account for roughly 21% of the variance of the score variable. Denote the residuals and fitted values from the regression as $\hat{s}_m$ and $\hat{\gamma}_m$, $m = 1, \dots M$. We further sort the observations w.r.t. residuals in ascending order and denote the new sample as $(q_{(m)}, \hat{s}_{(m)}, \hat{\gamma}_{(m)})$. We aim at using the identifying equation (11) in order to obtain the pseudo-sample $(q_{(m)}, \hat{\rho}_{(m)}, \hat{\gamma}_{(m)})$, where $\hat{\rho}_{(m)}$ are the estimates of pseudo-types.

Consider a sample analog of equation (11), evaluated at an evenly spaced grid:

$$\hat{Q}_\rho(u|\alpha) = \hat{Q}_s(u|\alpha) + \hat{q}_s(u|\alpha)\hat{c}(u), \quad u \in \{\frac{m}{M}\}_{m=1}^M. \tag{13}$$

Observe first that $\{\hat{Q}_s(\frac{m}{M}|\alpha)\}_{m=1}^M$ can be thought of as the observed column of (sorted) score residuals $\{\hat{s}_{(m)}\}_{m=1}^M$, while $\{\hat{Q}_\rho(\frac{m}{M}|\alpha)\}_{m=1}^M$ can be thought of as the sought column of pseudo-types $\{\hat{\rho}_{(m)}\}_{m=1}^M$. At the same time, $\{\hat{q}_s(\frac{m}{M}|\alpha)\}_{m=1}^M$ can be obtained as

$$\{\sum_{k=1}^{M} K_h(\frac{m-k}{M})(\hat{s}_{(m+1)} - \hat{s}_{(m)})\}_{m=1}^M,$$

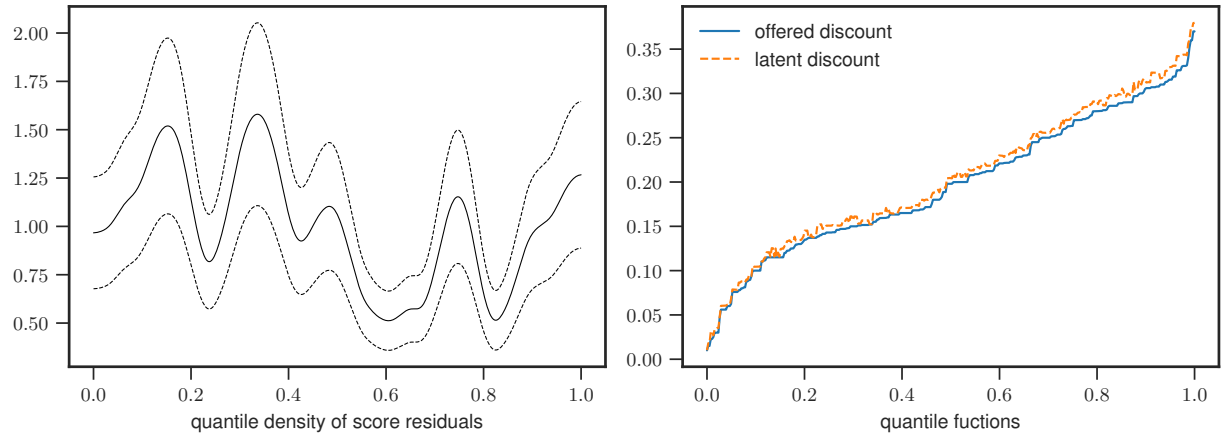
a non-parametric estimator of the quantile density, suggested in Jones (1992), see Andreyanov and Franguridi (2021) for details. We trim the distribution of residuals at 10% on each end and use a standard combination of a triweight kernel and Silverman rule-of-thumb bandwidth. Finally, $\hat{c}$ can be consistently estimated directly from the data, so the pseudo-sample can be constructed. See the results of estimation in Figure A.3.

While we could, in principle, construct a smooth estimator of $F_\rho(\cdot|\alpha)$ for every $\alpha$ and use it to evaluate each of the counterfactuals, we find it much easier to use the starting

Table A.1: Regression of the score on auction-day fixed effects with (1) additive and (2) multiplicative heterogeneity. With latter, the target variable is taken in logarithms.

| | *Dependent variable: score* | |
|---|---|---|
| | (1) | (2) |
| 2011-01-21 | 3.302** | 0.327*** |
| | (1.583) | (0.034) |
| 2011-01-26 | 5.734*** | 0.359*** |
| | (1.915) | (0.042) |
| 2011-02-15 | -7.209*** | 0.069* |
| | (1.831) | (0.040) |
| 2011-02-16 | 6.101*** | 0.369*** |
| | (1.515) | (0.033) |
| 2011-02-17 | 3.068*** | 0.299*** |
| | (1.152) | (0.025) |
| 2011-02-18 | 5.451*** | 0.348*** |
| | (1.692) | (0.037) |
| 2011-02-22 | 7.146*** | 0.389*** |
| | (1.316) | (0.029) |
| 2011-02-23 | 7.103*** | 0.396*** |
| | (1.160) | (0.025) |
| 2011-03-22 | 2.860** | 0.294*** |
| | (1.279) | (0.028) |
| 2011-03-23 | 13.682*** | 0.506*** |
| | (1.583) | (0.034) |
| 2011-04-28 | -1.973** | 0.198*** |
| | (0.984) | (0.021) |
| const | 45.264*** | 3.553*** |
| | (0.430) | (0.009) |
| Observations | 464 | 464 |
| $R^2$ | 0.211 | 0.201 |
| Adjusted $R^2$ | 0.193 | 0.184 |
| Residual Std. Error | 9.144 | 0.199 |
| F Statistic | 12.100*** | 11.410*** |
| *Note:* | *p<0.1; **p<0.05; ***p<0.01 | |

Figure A.3: Quantile density and functions




pseudo-sample $(q_{(m)}, \hat{\rho}_{(m)}, \hat{\gamma}_{(m)})$ to obtain a counterfactual pseudo-sample $(q_{(m)}, \hat{s}'_{(m)}, \hat{\gamma}_{(m)})$. The counterfactual winner in the auction is, therefore, the firm with the highest counterfactual score $\hat{s}'$.

## C.4 Simulations

Consider a sample analog of equation (12), evaluated at an evenly spaced grid:

$$\hat{Q}_s(u|\alpha') = \hat{Q}_\rho(u|\alpha') - \frac{\int_0^{m/M} \hat{C}(x) d\hat{Q}_\rho(x|\alpha')}{\hat{C}(u)}, \quad u \in \{\frac{m}{M}\}_{m=1}^M. \tag{14}$$

Again, $\{\hat{Q}_\rho(\frac{m}{M}|\alpha')\}_{m=1}^M$ can be thought of as the (nonparametrically estimated) column of pseudo-types, adjusted to reflect the change in the scoring rule:

$$\{\hat{\rho}'_{(m)}\}_{m=1}^M = \{\hat{\rho}_{(m)} + (\alpha' - \alpha) q_{(m)}\}_{m=1}^M.$$

Furthermore, we can approximate the integral with a sum:

$$\int_0^{m/M} \hat{C}(x) d\hat{Q}_\rho(x|\alpha') \approx \frac{1}{M} \sum_{m=1}^{M} \hat{C}(\frac{m}{M})(\hat{\rho}'_{(m)} - \hat{\rho}'_{(m-1)}),$$

and, of course, the $\hat{C}$ function can be estimated directly from the data. Finally, the counterfactual scores can be obtained as $\{\hat{Q}_s(\frac{m}{M}|\alpha')\}_{m=1}^M$ and the counterfactual discounts as

$\{\hat{s}'_{(m)} - \alpha' q_{(m)}\}_{m=1}^{M}$.

# D External Validity

This section focuses on the issue of external validity. Following List (2020), we discuss how our study complies with four necessary conditions for external validity. These conditions (SANS conditions) are useful in identifying whether the results obtained from a narrow, specific reform (or experiment) like that carried out by Acea are sufficiently likely to hold in broader contexts. These four conditions are (i) representativeness of the sample with respect to the full population and representativeness of the sample with respect to the relevant variables for the study; (ii) attrition rates and reasons for attrition and noncompliance; (iii) naturalness of the setting, choices, tasks and time frame observed; and (iv) scalability of the results, cost-benefit of the policy proposed at scale and conditions that would affect the outcomes.

- **Selection**: the issue of sample selection has three key dimensions in our work: (i) the selection of the buyer (Acea), (ii) the selection of the firms participating in the calls for tenders, finally (iii) the selection of the types and number of audits performed. Regarding the former, Acea s.p.a. is a multi-utility company offering electricity and water to about 1.6 million customers in the Rome area in central Italy. The firm is vertically integrated, owning and operating most of its generation, transmission, and distribution systems. From this point of view, it is very similar to some of the largest US power operators, such as the Los Angeles Department of Water and Power (LADWP), ComEd (Chicago), BGE (Baltimore), and PECO (Philadelphia). For the external validity of what can be learned from the Acea case, it is reassuring to observe in Table 1 that Acea is indeed similar to some other major operators active in the US along many margins such as size (in terms of employees and revenues) and resources spent to preserve the operational efficiency of its power grid. Moreover, as discussed in the text when presenting the difference-in-differences analysis, Acea is also similar to other companies providing the same services in Italy.

The new scoring rule based on past performance is tested by Acea on all its contractors in the electricity sector. The water sector, in which Acea also operates, is used as an internal control for the in-house policy evaluation. The comparability of the observed sample to other suppliers working with the public sector worldwide is ensured by legal standards. All suppliers in Italy must be certified by an external body to participate in public procurement. Similar systems exist in the United States and the rest of the world[36] and ensures comparability across different sectors and countries.

Acea operates in two segments of public works, electricity, and water. The past performance system is implemented only in the former, while the latter is used as an (internal) control. All the contractors participating in the auctions are observed. There is no reason to believe contractors in the electricity and water segments would be structurally different from contractors in other public works segments. All segments involving large public works with an element of risk involved would benefit from the policy proposed due to: (i) increased quality and (ii) increased safety. Therefore, we expect the results to hold across segments in terms of increased performance and reduced costs of safety hazards.

- **Attrition**:

  Concerning attrition, we see mixed evidence in the data. On the one hand, the number of bidders in each auction appears to grow when transitioning into the scoring period, but the change is continuous. On the other hand, if we follow the identities of firms, we observe modest attrition. More specifically, we see 4 *exiters*, 19 *stayers*, and 2 new entrants, see Table 5.

  Data shows that the discounts change dramatically across periods but not across firms. The difference is small if we compare the discounts of entrants, stayers, and dropouts in the grace period. Similarly, if we compare the stayers to entrants in the scoring period, the difference is negligible.

[36] In Italy, we refer to the OG and OS qualifications for suppliers working with the public sector. In the United States, a similar unique system is provided through the System for Award Management https://sam.gov

Overall, we do not see evidence of structural differences between the firms leaving and entering the market.

- **Naturalness**: The trade-off between price and contract performance is a common problem in both public and private procurement. However, this is heavily emphasized in the public sector to avoid the misallocation of public funds due to corruption and discretional assignment of public works. This has led to very different legislation and applications of the law, across different countries.

  Some legal systems allow more flexibility in the assignment of public contracts, among these the Past Performance Information Retrieval System (PPIRS) in the United States is a case in which the evaluation of past performance is a requirement for the assignment of public works. The use of a centralized scoring system for private suppliers ensures that there are no anomalies in the scores received by a certain company when working for a public body.[37]

  The European legislation is more strict in this sense.[38] Nonetheless, some countries, such as the United Kingdom have been able to implement a scoring system for private firms working with the public sector similar to the one present in the United States.

- **Scalability**: is essential to assess wether the results presented would hold at a national or international level. A critical concern in this sense is the protection of public procurement works against corruption.

  All the results shown would not hold if public administrators are able to bend the rules and assign public works to their "preferred" contractors. Our setting is in fact only able to sever the informal ties between inspectors and contractors, it does not act on the ties between the senior management of the firm and the contractors. In this case, the US PPIRS offers a great additional level of transparency to the public procurement auctions, which would be essential to scale the program at a national or international

[37]The PPIRS program has been recently withdrawn and integrated into the one-stop-shop System for Award Management (see https://sam.gov/) where all firms conducting economic activity with the Government of the United States are required to register.

[38]See *Directive 2014/24/EU of the European Parliament and of the Council*, URL: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex%3A32014L0024

level.

A similar system in Europe would be advisable to assess the effects of a wider use of past performance for the assignment of public procurement works. It would also harmonize the rules for public procurement in different countries and ensure the comparability of the works within the whole European Union, both for evaluating firm's performance in different countries and for monitoring purposes.

Overall, we conclude that the results presented are solid ground for future policies that scale public procurement systems based on past performance at a national or international level.

# E Additional Results

In this appendix section, we present several additional results supplementing the various analyses presented in the main text.

The estimates in Table A.2 explore the behavior of suppliers when they become aware of the new scoring auction. We do so by focusing on the audit data in the period before the introduction of the scoring rule and further partitioning this sample into two subsamples: audits held before and after $t1$. For each of these subsamples, we estimate a series of probit regressions performed at the level of each individual audited parameter. We estimate the following probit model for the probability of the score being 1 (i.e., compliant) on features of parameters, contracts and suppliers:

$$Pr(compliant) = \Phi[t + f + \alpha\ weight + \theta\ quick + \gamma_j \sum_{j=2}^{12} category_j], \tag{15}$$

where $\Phi$ is the normal cdf, *compliant* is the score (0 or 1) taken by the parameter audited, $t$ and $f$ are fixed effects for the year and contractor, *weight* is the weight associated with the parameter, *quick* is a dummy for whether the parameter can be adjusted within one month at a small cost and $category_j$ are dummies for the category to which the parameter belongs.

We are particularly interested in the coefficient on *weight* as this has the potential to reveal the strategic nature of supplier responses. Table A.2 shows the probit marginal effects

Table A.2: Probability of Compliant Parameter

| | <t1 | | | | >t1 | | | |
|---|---|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) |
| Weight | -0.024*** | -0.022*** | -0.022*** | -0.023*** | 0.012*** | 0.009*** | 0.006*** | 0.006*** |
| | (0.003) | (0.005) | (0.005) | (0.005) | (0.000) | (0.000) | (0.000) | (0.000) |
| Quick | | 0.075** | 0.075** | 0.065** | | 0.015*** | 0.009*** | 0.009*** |
| | | (0.025) | (0.025) | (0.025) | | (0.002) | (0.002) | (0.002) |
| C2-Documentation | | -0.406*** | -0.406*** | -0.424*** | | -0.179*** | -0.093*** | -0.091*** |
| | | (0.038) | (0.038) | (0.038) | | (0.003) | (0.003) | (0.003) |
| C4-Works Execution | | -0.538*** | -0.538*** | -0.534*** | | -0.065*** | -0.057*** | -0.056*** |
| | | (0.042) | (0.042) | (0.043) | | (0.003) | (0.002) | (0.002) |
| C8-Underground works | | -0.328*** | -0.328*** | -0.309*** | | -0.225*** | -0.114*** | -0.108*** |
| | | (0.034) | (0.034) | (0.035) | | (0.003) | (0.002) | (0.002) |
| C10-Personnel | | -0.315*** | -0.315*** | -0.327*** | | -0.137*** | -0.130*** | -0.126*** |
| | | (0.048) | (0.048) | (0.048) | | (0.003) | (0.002) | (0.002) |
| C11-Works site regularity | | -0.683*** | -0.683*** | -0.681*** | | -0.193*** | -0.157*** | -0.151*** |
| | | (0.036) | (0.036) | (0.037) | | (0.002) | (0.002) | (0.002) |
| C12-Works site safety | | -0.394*** | -0.394*** | -0.406*** | | -0.110*** | -0.098*** | -0.095*** |
| | | (0.037) | (0.037) | (0.038) | | (0.003) | (0.002) | (0.002) |
| Year FEs | | | ✓ | ✓ | | | ✓ | ✓ |
| Firm FEs | | | | ✓ | | | | ✓ |
| N | 3813 | 3053 | 3053 | 3019 | 361338 | 256720 | 256720 | 256630 |

*Note: This table reports the marginal effects of probit regressions. The dependent variable is the score on the parameter: 1 if compliant and 0 if not compliant. The first four columns regard the subsample of scores assigned in the audits held before $t1$, while the latter four columns regard audits that occurred after $t1$.*

for two separate samples: audits held in the period before $t1$ (first four columns), and audits held after then (last four columns). We find that the sign of the coefficient on *weight* changes from negative to positive. Thus, after $t1$, suppliers become more compliant in those parameters with the strongest potential to bolster their RI. This switch in the coefficient sign is evident across all specifications, as we move from a baseline model, controlling only for *weight*, and we expand the model to incorporate parameter, contract and firm features.[39]

Regarding the other coefficients in Table A.2, the one on *quick* is useful to assess the potential for collusion between suppliers and monitors. Indeed, performance might be improving because the repeated interaction allows the parties to learn how to collude under the new system. However, this interpretation of the data would seem less plausible if the improvements were concentrated on those parameters that should be faster to effectively

[39]All estimates in Table A.2 are based on the subset of parameters that are audited at least once both before and after $t1$. The results remain qualitatively the same for the post-$t1$ sample if all audits are included.

adjust. With the help of expert engineers, we created a dummy variable, *quick*, that is equal to 1 if the transition from a score of not compliant to one of compliant can be reasonably achieved within a one month time frame without incurring extraordinary costs. For instance, examples of parameters with *quick* equal to 1 are those involving the adequacy of "personal protection tools" (mostly helmets) or the presence of signs warning of ongoing works nearby. Instead, the adequacy of the machinery is an example of a parameter with *quick* equal to zero. While clearly arbitrary, this dummy variable is helpful to test the reasonableness of the performance response observed in our data. Indeed, the finding that the coefficient on *quick* is positive (and that its significance increases post $t1$) is suggestive of suppliers effectively changing their behavior. This interpretation is further strengthened by what we report below with regard to the behavior in the auctions.

While it is impossible to fully rule out the possibility of collusion/corruption, the system of random rotation of auditors and of random selection of the sites to inspect was explicitly meant to curtail these types of risks. Indeed, Acea never expressed to us concerns about episodes of corruption or collusion during the period our data cover.

Table A.3 an augmented version of Table 4. The table reports the estimates of a series of probit regressions for the probability of passing the audits in the various periods. Namely: $< t1$ indicates the period right before the $t1$ (i.e., 20 December 2007). $t1$, $t2$, $t3+1$, $t5+1$ are all the breakpoints identified in the time series analysis during the announcement phase (20 December 2007 to 18 May 2010). We progressively control for several confounding factors: the winning bid's aggressiveness (*Win bid - Average bid*), the number of participants in the auctions, a series of contract-specific controls, and firm fixed effects. The time coefficients are generally large and significant, showing an increasing trend from $< t1$ to *hybrid*. The trend is more pronounced when we include firm-specific fixed effects. This shows not only that there is an improvement over time in the scores, but also that the improvement is mostly within the firms.

Finally, Figure A.4 analyzes the jump in discount at the moment of the switch, but controlling for (i.e., partialling out) the number of bidders and the reserve price. The jump of the residualized discount is slightly smaller than in the original exercise, 11 percentage points as opposed to 14, when we average over the 3 auction days before and after the switch.

Table A.3: Probit Audit Passed

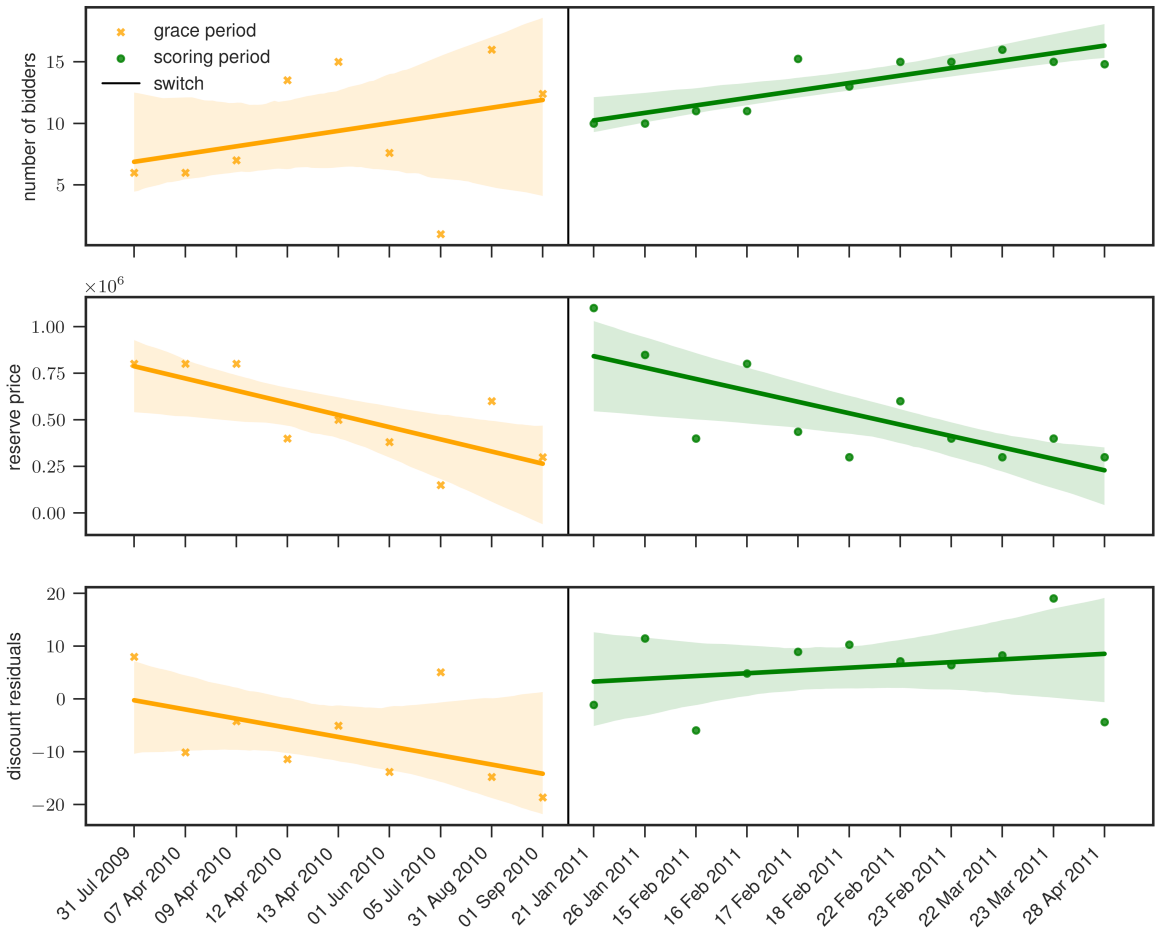
| | (1) | (2) | (3) | (4) | (5) |
|---|---|---|---|---|---|
| VARIABLES | Evaluation | Evaluation | Evaluation | Evaluation | Evaluation |
| <t1 | -0.899*** | -0.616*** | -0.618*** | -0.750*** | -0.659*** |
| | (0.097) | (0.124) | (0.124) | (0.111) | (0.114) |
| t1 | -0.203*** | 0.017 | 0.016 | -0.154** | -0.077 |
| | (0.034) | (0.081) | (0.081) | (0.061) | (0.065) |
| t2 | 0.103*** | 0.290*** | 0.290*** | 0.123** | 0.198*** |
| | (0.021) | (0.077) | (0.077) | (0.053) | (0.057) |
| t3+1 | 0.669*** | 0.852*** | 0.851*** | 0.666*** | 0.745*** |
| | (0.024) | (0.077) | (0.077) | (0.057) | (0.061) |
| t5+1 | 0.927*** | 1.093*** | 1.092*** | 0.866*** | 0.941*** |
| | (0.021) | (0.076) | (0.076) | (0.058) | (0.061) |
| scoring | 1.437*** | 1.547*** | 1.544*** | 1.006*** | 1.070*** |
| | (0.050) | (0.089) | (0.090) | (0.094) | (0.096) |
| hybrid | 1.787*** | 1.899*** | 1.897*** | 1.102*** | 1.160*** |
| | (0.010) | (0.070) | (0.071) | (0.086) | (0.088) |
| Win bid - Avg bid | 0.000 | -0.002 | -0.002 | 0.000 | 0.004* |
| | (0.001) | (0.002) | (0.002) | (0.002) | (0.002) |
| Number of offers | | | 0.001 | | -0.011*** |
| | | | (0.002) | | (0.003) |
| Observations | 123,173 | 123,138 | 123,138 | 102,516 | 102,516 |
| Supplier FEs | | ✓ | ✓ | ✓ | ✓ |
| Auction Controls | | | | ✓ | ✓ |

Standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

*Note: The table reports a series of probit regressions on the probability of passing the audit of a single parameter (1 = passed, 0 = failed). The regressions include several controls, namely: supplier fixed effects, number of offers in the auction, distance between winning bid and mean bid, and contract specific controls.*

Figure A.4: Discontinuity at the switch.




*Note: Top figure - number of bidders. Middle figure - reserve price. Bottom figure - residualized discounts, with the logarithms of the number of bidders and reserve price partialled-out. Data is averaged by auction day.*

# F Full list of internal performance measures

Table A.4: Internal Performance Measures

| Parameter | Category | Weight |
|---|---|---|
| Appliances conditions* | Vehicles | 9 |
| Assembly appliances with respect to original design | Cabinet Works | 7 |
| Assembly electromechanical equipment | Aerial Works | 7 |
| Assembly other equipment | Aerial Works | 7 |
| ATM presence* | Documents | 10 |
| Bend radius of wires execution | Cabinet Works | 7 |
| Binder quality | Underground Works | 4 |
| Binder reconstruction - thickness | Underground Works | 7 |
| Binding execution | Aerial Works | 9 |
| Braces compliant with original design | Aerial Works | 5 |
| Braces sealing | Aerial Works | 5 |
| Burying material | Underground Works | 7 |
| Cabin interferences | Cabinet Works | 3 |
| Cleanliness in assembly stages | Joints Exexution | 6 |
| Clothing availability* | Works Safety | 8 |
| CLS thickness, with respect to prescriptions | Underground Works | 7 |
| Columns centering during direct burying | Aerial Works | 4 |
| Concession and/or permits* | Documents | 4 |
| Concrete transport documents* | Documents | 3 |
| Concreting pipe | Underground Works | 4 |
| Connection grounding - cabin | Cabinet Works | 8 |
| Construction signs* | Works Regularity | 4 |
| Correct cable finding | Joints Exexution | 6 |
| Correct installation equipotential box | Joints Exexution | 7 |
| Correct installation equipotentiality | Joints Exexution | 7 |
| Correct schemes continuity recovery | Joints Exexution | 7 |
| Display of execution plate | Joints Exexution | 5 |
| Disposition DSE(CEL) actuated through notes/minutes* | Works Verifications | 9 |
| Document of transport/quality of concrete | Underground Works | 8 |
| DPI availability* | Works Safety | 10 |
| DPI usage* | Work Execution | 10 |
| Drag and deflection | Aerial Works | 8 |
| Duct characteristics | Underground Works | 4 |
| Duct disposal | Underground Works | 4 |
| Electrical connections executions | Cabinet Works | 9 |
| Electrical risk checks* | Work Execution | 8 |
| Emergency personnel appointment* | Works Safety | 10 |

| | | |
|---|---|---|
| Emergency personnel presence* | Works Safety | 10 |
| Equipotential connection* | Work Execution | 10 |
| Extrados height of upper tube | Underground Works | 8 |
| Fencing of construction site* | Works Regularity | 10 |
| Fencing of deposits* | Works Regularity | 5 |
| Fencing of excavations* | Works Regularity | 9 |
| Fencing of machine operator* | Works Regularity | 8 |
| Fill-in commercial documents | Users Management | 6 |
| Filling material compliant | Underground Works | 8 |
| Fire extinguisher* | Works Safety | 9 |
| Floor plan of the project* | Documents | 4 |
| Floor plan of the services* | Documents | 7 |
| Following the sequences | Joints Exexution | 5 |
| Gas detector* | Works Safety | 9 |
| Gas-operated welding instruments | Joints Exexution | 5 |
| Graphics* | Documents | 5 |
| Groot bed thickness | Underground Works | 5 |
| Ground loop compliant with original design | Aerial Works | 9 |
| Grounding connection | Aerial Works | 9 |
| Grounding of appliances* | Work Execution | 10 |
| Grounding of plants* | Work Execution | 10 |
| Grounding works compliant with cabinet | Cabinet Works | 8 |
| Hydraulic brus-cutter | Joints Exexution | 6 |
| Hydraulic press | Joints Exexution | 7 |
| Identification* | Personnel | 10 |
| Insulated brush-cutting | Joints Exexution | 6 |
| Insulating appliances availability* | Works Safety | 9 |
| Interferences | Underground Works | 7 |
| Interferences | Cabinet Works | 7 |
| Interferences - Stretching cables | Aerial Works | 6 |
| Material supplies | Underground Works | 6 |
| Medical aid kits* | Works Safety | 10 |
| Milling - thickness | Underground Works | 7 |
| Modification of vehicles and pedestrial circulation* | Works Regularity | 9 |
| Observing prescriptions for cable-laying work | Underground Works | 7 |
| OTMs conditions* | Vehicles | 8 |
| Paintings executions | Cabinet Works | 2 |
| Plant delivery documents (PCL)* | Documents | 10 |
| Positioning of cross-bars, shelves and so on | Aerial Works | 6 |
| Positioning of metal braces | Aerial Works | 7 |
| Potential dangers during works* | Work Execution | 8 |
| Preliminary notification present and displayed* | Works Verifications | 8 |
| Proper clothing usage* | Work Execution | 7 |
| Qualifications according to norms CEI* | Personnel | 10 |

| | | |
|---|---|---|
| Quality of CLS | Underground Works | 6 |
| Quality of works | Cabinet Works | 4 |
| Realization compliant with original design | Cabinet Works | 7 |
| Reels stan* | Work Execution | 7 |
| Repaintings executions | Cabinet Works | 2 |
| Respect planned meetings | Users Management | 8 |
| Sealing ducts in wells | Underground Works | 6 |
| Security and coordination plan presence* | Works Verifications | 10 |
| Security signs worksite* | Works Verifications | 9 |
| Sequences and installation | Joints Exexution | 5 |
| Sheet piling | Underground Works | 5 |
| Shrinking stages (thermo or auto) | Joints Exexution | 6 |
| Sign of machine operator* | Works Regularity | 8 |
| Size of excavations | Aerial Works | 6 |
| Slope of foundation upper surfaces | Aerial Works | 5 |
| Splicing technicians qualified* | Personnel | 10 |
| Squareness with axis | Aerial Works | 6 |
| Steady polymerization process | Joints Exexution | 3 |
| Straight alignment of supports | Aerial Works | 6 |
| Subcontractors operating plan presence* | Works Verifications | 10 |
| Supplies (cabinet) | Cabinet Works | 8 |
| Supplies (I.T.) | Cabinet Works | 8 |
| Supply materials 1 | Aerial Works | 8 |
| Supply materials 2 | Aerial Works | 8 |
| Supply materials 3 | Aerial Works | 8 |
| Supply materials 4 | Aerial Works | 8 |
| Supply materials 5 | Aerial Works | 8 |
| Support burying | Aerial Works | 6 |
| Support positioning | Aerial Works | 4 |
| Tent installation | Joints Exexution | 6 |
| Timely execution of the works | Users Management | 8 |
| Total height | Underground Works | 8 |
| Type and quantity of tubes compliant with original design | Underground Works | 3 |
| Type of cable | Underground Works | 4 |
| Vehicles conditions* | Vehicles | 8 |
| Vehicles documents* | Vehicles | 10 |
| Vehicles identification* | Vehicles | 7 |
| Vertical braces | Aerial Works | 6 |
| Visible badge* | Personnel | 7 |
| Visual examination of quality and execution | Underground Works | 5 |
| Warning signs (night)* | Works Regularity | 10 |
| Warning signs (proximity to site)* | Works Regularity | 10 |
| Warning signs (vertical and horizontal)* | Works Regularity | 9 |
| Warning tape | Underground Works | 6 |

| | | |
|---|---|---|
| Water tightness verification | Joints Exexution | 6 |
| Wear layer reconstruction - thickness | Underground Works | 7 |
| Width of excavation | Underground Works | 8 |
| Wire stripping 1 | Joints Exexution | 6 |
| Wire stripping 2 | Joints Exexution | 5 |
| Workplace cleanliness | Joints Exexution | 5 |
| Workplans* | Documents | 10 |
| Works awarding* | Documents | 9 |
| Works compliant with original design | Aerial Works | 7 |
| Works overseers presence* | Works Safety | 10 |
| Worksite journal updated* | Works Verifications | 7 |

*Note: Parameters marked with an * are those identified by Acea engineers as most closely related to safety features of the job execution..*