Past Performance and Procurement Outcomes

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

May 18, 2023

Abstract

Balancing cost considerations with the quality and safety of contract execution is a major concern in procurement. We empirically analyze the introduction of a vendor rating system and the associated switch from price-only auctions to scoring auctions weighting price-and-past-performance. We find that compliance with safety and quality standards improves from 25 percent to 90 percent of audited parameters. Improvements involve all parameters and suppliers, are long-lasting and are reflected in a higher quality of the buyer output. However, we also show that these improvements come at virtually no cost for the buyer. We explain how this empirical puzzle can be an equilibrium outcome if firms vary in more than one dimension, and support this interpretation with a structural analysis.

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: SITE-Stockholm School of Economics, EIEF and Tor Vergata, *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 **1**

1.1 Literature 6

2 Institutional Details **8**

3 Data and Descriptive Evidence **11**

4 Reduced Form Analysis **18**

4.1 Internal performance measures 18

4.2 External performance measures 21

4.3 Evolution of winning bids 22

4.4 Supplier entry 23

5 Stylized model **25**

5.1 Univariate types. 27

5.2 Bivariate types. 28

6 Empirical model **30**

6.1 Model Primitives and Identification. 32

6.2 Estimation and simulation. 32

6.3 Counterfactuals 33

7 Cost-effectiveness Analysis **34**

8 Conclusions **37**

1 Introduction

The need to prevent corruption led lawmakers around the world to ensure that open auctions, where bidders receive equal treatment, are used as often as possible, even if supplier past performance for similar contracts differs considerably. But competitive auctions work poorly in the context of incomplete contracts characterizing a large part of the procurement:

bolstering price competition might come at the cost of inadequate ex-post performance.[1] Balancing these forces is therefore an endemic problem of public procurement.

If performance can be observed and recorded, the problem can be solved by linking the award of the contract to past performance, in the format of the scoring (SR) auction.[2] This could give an advantage to firms with a better organizational structure and work ethics, at the same time, stimulating today's performance through future rewards. However, it is not clear whether the immediate costs of providing such stimuli at the bidding stage would be outweighed by the hidden gains in terms of ex-post performance.

In this paper, we empirically analyze the introduction by a large utility company of a vendor rating system (i.e., the recording of past performance) combined with a scoring rule. This company, Acea spa (the buyer), provides water and power to a vast area in central Italy that includes its capital, Rome. In 2007, this firm implemented a new vendor rating system to improve contractual performance in the execution of the construction jobs that it awards for the maintenance and upgrade of the electricity grid. A unique feature of this setting is that it allows us to observe the evolution of quality and price between the announcement and the enforcement of the system, to which we refer as the *grace period*, separately from the response to the subsequent switch in the auction rules. Together with the availability of a control group, this provide us with a credible empirical strategy to identify the causal effects of the reform on bidder decisions involving entry, bidding, and quality provision.

Our analysis combines together several data sources. First, complete audit data on the 136 parameters measuring quality and safety standards in suppliers' contract execution are observed for 10 years: from the introduction of the new audits in 2007 to 2017. Second, from the public regulatory authority, we obtained measures about the quality of the output for both the buyer and comparable utility companies. Third, for all the utility companies in the country, we also observe the procurement auctions held between 2004 and 2017, thus offering a clear benchmark for the evolution of procurement costs. Finally, we complement

[1] A classic reference is Spulber (1990) which shows that in the construction sector, where contracting is typically imperfect, open competition spurs adverse selection and ex post opportunism of contractors. For a recent contribution, see Lopomo, Persico and Villa (2023) and the other references discussed in section 2.

[2] By scoring auctions we refer to auctions that award the contract on the basis of a scoring system evaluating multiple criteria, for instance weighting the price offer and the technical offer. Scoring auctions are frequently used in procurement and have been previously analyzed in the literature (see, among others, Che (1993); Asker and Cantillon (2008, 2010); Andreyanov (2019)).

the previous data with supplier balance sheet data.

We begin with the detailed analysis of the grace period. First, as it is natural, we ask whether and how the introduction of the vendor rating system induced an improvement in contract execution. Our empirical strategy here exploits the timing with which the reform was announced: in 2007, Acea's auditors started to record the scores on the parameters inspected, but only a few months later suppliers were informed that these audits would be used to determine a "reputation index" (RI) as part of a new system of scoring rule auctions giving 75 percent of the score to price and 25 percent to the RI.[3] We analyze how compliance with the parameters monitored evolved during the grace period, in response to the timing of five public announcements made by Acea to its suppliers and find clear evidence of a substantial change in suppliers' behavior: compliance in the parameters audited increased from 25 percent before the first announcement to more than 80 percent before the first auction took place under the new SR auction. Over the following months, performance improves further, stabilizing at 90 percent and remaining at this high level thought the rest of the decade of data in our sample.

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. Moreover, these improvements recorded by the internal audits have a corresponding upgrade in the buyer output quality: 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. Interestingly, the quality of the water distribution service, also provided by Acea, but not interested in the reform analyzed here, did not improve or worsen.

The analysis of bidding behavior reveals how firm competition changed in response to the buyer reform. If we partition the grace period into two phases corresponding to when compliance initially grows, and when it flattens out at high levels, then the first phase is associated with a noticeable decrease in prices, and the second to their rebound. We explain the initial price drop with the need to win more auctions, in order to get a chance to be

[3]As discussed below, we the word "reputation" is the one chosen by Acea for its vendor rating system, which is fully based on observable, audited parameters. This has not to be confused with the usage in the economics literature of reputation as a belief over a player's type Bar-Isaac and Tadelis (2008).

audited. In the second phase, the price rebound can be explained by the fact that most bidders have acquired their desired reputational index. The cumulative effect is ambiguous, so we cannot say that the costs for the buyer have increased at this point.

Proceeding to the events that took place after the grace period, one could expect that the firms that have invested into quality, would finally take advantage of the new auction mechanism by dramatically increasing their bids. To the contrary, we find that price decline when the SR auction is implemented, even though firms' reputation index remains high and stable.[4] This is puzzling, because it contradicts the conventional wisdom that an increase in quality must come at a cost for the buyer.

We can trace this intuition to a rigorous argument, that the firms must be compensated for their investment costs, since their expected utility is constant (across different auction formats) by the revenue equivalence principle. This argument, however, relies on a seemingly harmless assumption made throughout the literature, that firms only vary in their ability for cost reduction (i.e. efficiency). We show that if firms vary in two dimensions, for example, efficiency and ability to produce quality, the revenue equivalence principle does not hold anymore, thus opening a possibility of transfer of informational rents from the firms to the buyer, and, hence, explaining the puzzle.

To disentangle the complex incentives produced by the scoring auction, we analyze and estimate a structural model that significantly departs from the existing literature, due to the presence of sunk costs, associated with past performance. To see how it works, consider an anticipated switch from the price-only to the scoring auction, like the one that took place at the end of the grace period. There are two main factors at play here. Note first that if a firm has invested in quality through past contracts, her maximal possible discount does not change for the current contract (i.e., the quality investment is a sunk cost, as is the case in any vendor rating system that scores past performance). On the other hand, the maximal possible discount of the winning firm will decrease through selection. Indeed, the scoring auction puts the efficient (low-cost) firms at a disadvantage, so that inefficient (high-cost but possibly high-quality) firms can occasionally win. Similarly, we can argue that the quality

[4]This is likely due to the combination of irreversible investments into quality, as well as the fact that the reputation index is a moving average of the results of audits of past contracts, most of which took place during the grace period.

of the winning firm will increase, and even more so if she choose to raise it strategically. Thus, the first factor is the selection (i.e. ranking of firms by the auction mechanism), and it contributes to both quality and buyer's costs.

The second factor is more subtle and is related to the strategic shading of bids. Note that, by the nature of the scoring auction, the high-cost-high-quality firms are pitted against the low-cost-low-quality ones, making them bid more aggressively (shade less) than in the price-only auction. At the same time, some of the low-cost-high-quality firms may bid less aggressively. However, if costs and quality are positively correlated, we may expect more aggressive bidding on average. It will then decrease the cost for the buyer and can even outweigh the increase associated with the first factor.

The estimates of the model support this thesis. Indeed, the model predicts that the switch to the scoring auction is associated with a decrease in the winning firm's informational rents (i.e., her auction profits), by roughly 12 percent (€1.1 thousand per auction), on average. This partially offsets the increase in the winning firm's costs, so that the cumulative effect on the discount is statistically ambiguous. Thus, the switch came at essentially no cost for the buyer. On the other hand, for the firms, the switch was very costly, even without taking into account the investments made to increase their compliance.[5]

Finally, we ask is what is the cost-effectiveness of the reform. We combine the cost estimates with the ones on compliance on the audited parameters, separately for two sets of parameters concerning either quality or safety. This is achieved by mapping the estimates of the quality improvements to a measure of the buyer output (blackouts duration) and the ones on safety to the probability of deadly accidents. We estimate that the benefit from increased contract compliance ranges between €3.5 and €5.3 million per year in terms of safety improvements; while for the reduction in blackouts, the reform implies a benefit of €6.6 million. Since the costs for the buyer did not increase, these estimates represent a lower bound on the benefits produced by the reform making it highly cost-effective.

[5]This also rationalizes the strong opposition to the scoring rule voiced by bidders which ultimately led the buyer to dismiss the scoring rule auction.

1.1 Literature

This study offers the first, systematic evidence of the introduction of a vendor rating system in public procurement that we are aware of. At the highest level, its main lesson is that for contract procurement the gains from introducing a past-performance retrieval system coupled with scoring auctions may be higher than those from always bolstering price competition via price-only auctions. In the conclusion, we discuss the policy implications of our work, as well as its external validity and scalability. In this section, instead, we focus on how this study relates to various strands of the literature.

At the most general level, this study is related to two strands of the law and economics literature on agency problems. The first strand concerns ex-ante regulation vs. ex-post incentives. Shavell (1984), and the research line following from it, 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 extent of potential hazards.[6] 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.[7] The second strand is that of the efficiency-corruption trade-off in delegation within an organization, see Banfield (1975) as the classic reference. Price-only auctions represent rigid mechanisms where delegation to the agents (i.e., Acea's engineers) of the awarding and monitoring of the contracts is minimal. The introduction of a vendor rating requires delegating more powers to the auditors, thus risking that they will exploit it for personal gain. In our case, we find a wave of recent papers on public procurement.[8]

A second body of literature to which we contribute is the industrial organization studies of

[6] 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).

[7] 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.

[8] See Coviello, Guglielmo and Spagnolo (2018), Carril, Gonzalez-Lira and Walker (2020), Decarolis et al. (2020), Bosio et al. (2020), and Bandiera et al. (2021). More specifically on the issue of corruption, as it will be discussed below, Acea's approach involved not only fostering delegation but also containing corruption risks through a mechanism of rotation and random drawing on the pool of auditor-scored suppliers. Related to the issue of corruption, Bandiera, Prat and Valletti (2009) show that, in the context of Italian procurement, corruption concerns could be less of a priority than inefficient procurement.

auctions and competition. In particular, the very 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) and Burguet, Ganuza and Hauk (2012). This has led to the proposal by of new auction formats to optimally trade off price vs performance risk, 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 terms of various measures of contract performance 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] Furthermore, there is literature on scoring auctions that builds on the seminal results of Che (1993) and Asker and Cantillon (2010). This literature is closely connected to our study, but with important differences that we highlight when presenting our model.[11]

Another strand of the literature to which our paper contributes is that on 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 the sense that repeated interaction only increases the price/quality correlation when a quality rating system is present.

[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 play in improving contract performance in repeated public procurement under imperfect contracting. See, among others, Calzolari and Spagnolo (2009), Board (2011) and Andrews and Barron (2016).

[11] 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).

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, both private households and business establishments, in the Rome area. The firm is vertically integrated, owning and operating the majority 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).[12] As shown in Table 1, all of these firms spend significant resources every year on works aimed at preserving the operational efficiency of their power grid.

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 the maintenance, upgrade, and replacement of 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. Up 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 involved exclusively the electricity sector, leaving out the water sector and the contracts related mostly to maintenance of the public illumination (IP) and electricity distribution (DIST) systems.[13] 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 materials and machinery used to the adherence to the worksite safety specifications and legal status

[12] The external validity of what can be learned from a firm-level reform is a typical concern in the literature (Bloom et al. (2014)). In our case, it is thus reassuring to observe that Acea is similar to both some other major operators active in the US, such as the multi-utility companies of the four US cities mentioned above, and to the other companies providing the same services in Italy, as discussed below.

[13] For Acea this was motivated by the intent to have a benchmark against which to evaluate the effectiveness of the new system. 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.

of all workers (the full list is reported in Appendix Table A.4). The logic followed by Acea was indeed that of covering 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 a typical inspection, the auditors could score on average 34 parameters.

2. To aggregate the audit outcomes into a supplier-level measure, the reputation index (RI), each parameter was associated with a weight, ranging from 2 to 10, and the RI was calculated as a weighted average across a predefined time span:

$$RI = \frac{\sum_{i=1}^{m} \sum_{j=1}^{136} p_{ij} u_j}{\sum_{j=1}^{136} u_j}, \tag{1}$$

   with $p_{ij} \in \{0, 1\}$ indicating the score in each of the $j \in \{1, .., 136\}$ parameters, with $u_j \in \{2, 3, ..., 10\}$ being the weight attached to parameter j and m being the set of audits considered

3. At each point in time, the the audits considered for the calculation of the RI are only those in the previous 12 months. Hence, RI ranges from 0 to 1 and entails no differential discounting of the m audits.

4. To limit the risk of corruption and biased evaluations, a randomization was implemented at two levels. First, every week it was randomly drawn which work sites to inspect out of those were suppliers were actively working. The same worksite could thus be audited multiple times, or never. Second, the composition of the 3-member auditor teams was randomly draws from the pool of Acea auditors (about 12 engineers).

5. Scoring rule auction. The RI was used as a component of the following scoring formula, that determined the winner of the SR auction as the bidder with the highest score:

$$S = \omega(1 - \frac{\text{Price offered}}{\text{Reserve price}}) + (1 - \omega)RI, \tag{2}$$

   where ω is the weight assigned to the price relative to that assigned to the RI. The switch from price-only to SR auctions entailed a change from $\omega = 1$ to $\omega = 0.75$.

6. For new entrants and for firms with very sparse audit data, Acea decided to calculate the RI only if at least 7 audit visits had been done in the previous 12 months.[14] If this condition was not met, 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 never enforced due to due to the inefficiency of the Italian court system (Giacomelli and Menon (2016) and Coviello et al. (2018)).

The timing of the reform plays an important role for the analysis that follows. It is summarized in Figure 1. The vertical red lines mark the most significant dates: the new audit system begins in 2007 and continues throughout the whole sample period. The SR auction is implemented from mid 2010 to mid 2011, price-only auctions are used at all other times. Between when the new audit system begins and when the first SR auction is implemented, Acea conducted a series of meetings with its suppliers to explain the reform.

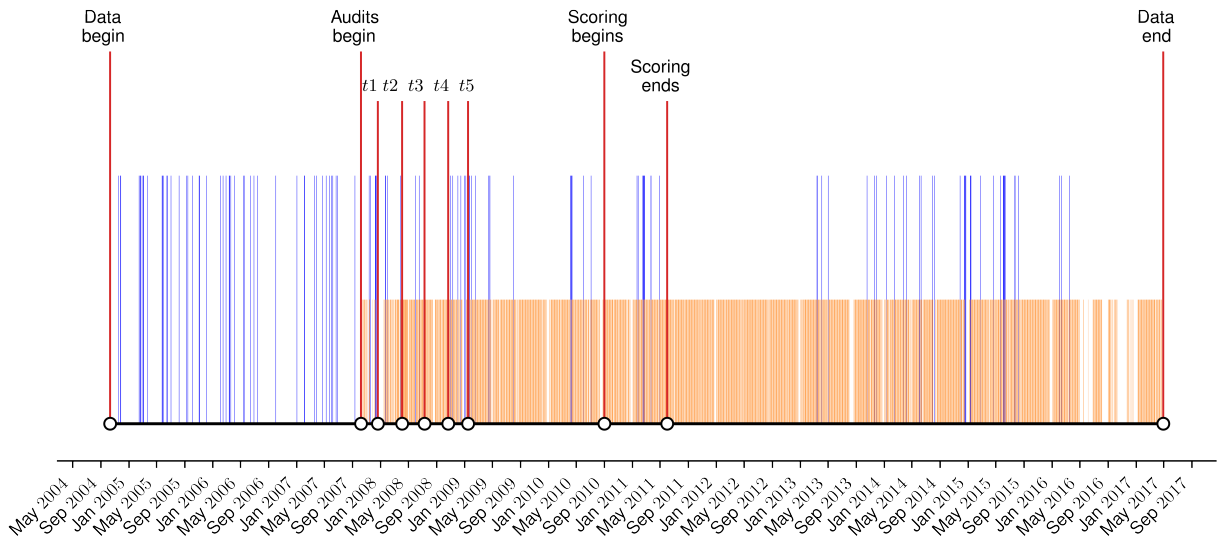
A crucial element of the timing of the reform is that initially Acea concealed its true motivation for the switch to digitalized audits. It was only three months after the new audit system had started that Acea announced to its suppliers in a public meeting held on December 20, 2007 (corresponding to the vertical line marked as $t1$ in Figure 1) the intention to switch its contract procurement system from price-only auctions to price-plus-performance auctions. Both in this first meeting with its suppliers and in 4 follow-up meetings held in the following 13 months (marked as $t2$,...,$t5$), Acea explained this new system and showed simulations of how a firm would benefit from higher RI. It also updated each firm by (privately) informing it of its current RI, as well as (publicly) disclosing the distribution of RI across all suppliers.

The delayed adoption of the SR was in part motivated by the need to collect enough audit data, but also by an internal legal dispute on whether the SR auction was compliant with the Italian law and EU Procurement Directives.[15] These legal considerations, together

[14]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 served to limit a "cold start" problem as a barrier to entry, as discussed in Butler et al. (2020).

[15]Although the law encourages the use of SR auctions, the parameters in the scoring formula should to pertain exclusively to elements of the bids and not features of the suppliers.

Figure 1: Timeline



Note: Long, blue ticks represent auctions; short, 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.

with the supplier complaints discussed below, led to the dismissal of the SR in 2011.

In the analysis that follows we focus mostly on what happened during the period between the adoption of the new audit system and the first scoring rule (recall that we refer to this as the grace period) and on the period of adoption of the SR auction. In the grace period, supplier competed to win contracts under price-only auctions but were already building their stock of RI. Clearly, in this period the RI could not act as a barrier to entry since bids were just price discounts. The period that followed the SR auction entailed 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. We refer to this latter system as *hybrid price-only auction*.

3 Data and Descriptive Evidence

The analysis is based on three main sets of data. The first comes from Acea and contains audit data covering the internal performance measures recorded through the new auditing system. The second combines data from Acea and Telemat, a large provider of public tender

Figure 2: Lasting increase in performance

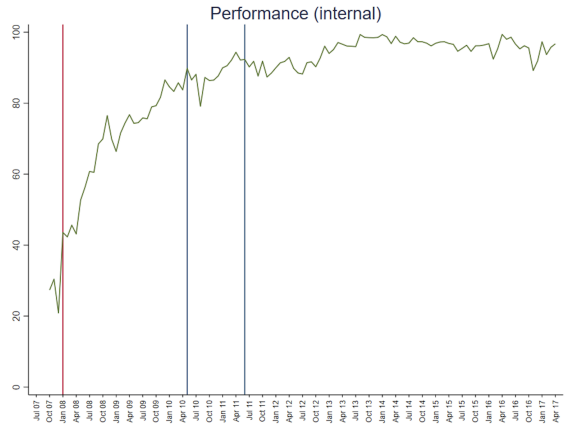
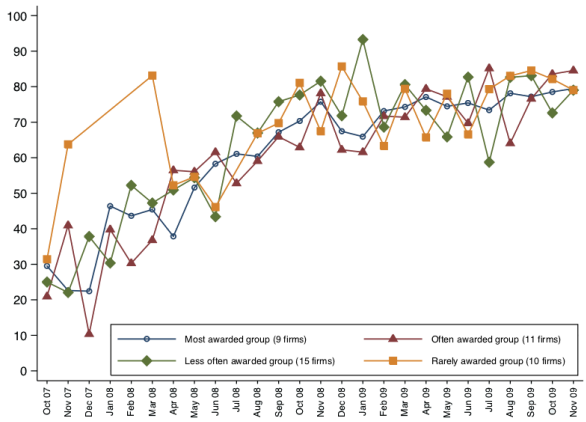


Figure 3: Performance by supplier type



data, and contains auction data, covering bidding and other auction-related information. The third comes from the public authorities supervising the power and water sectors and contains external performance measures.

The first dataset contains all of Acea's audits under the new system, from its introduction in October 2007 until April 2017. We will refer to the results of these audits as internal performance measures.

There are 302,634 scores assigned to each parameter inspected during 8,974 audits involving 634 contracts and 73 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 at all or 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 a 10-year period. 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 whilst, at the opposite end of the spectrum, parameters in the "Works site regularity" and "Works site safety" categories are systematically assessed.

The table also reports the average share of compliant parameters (i.e., those scored with a 1 over all those scored with either zero or 1). The share is reported separately for each of

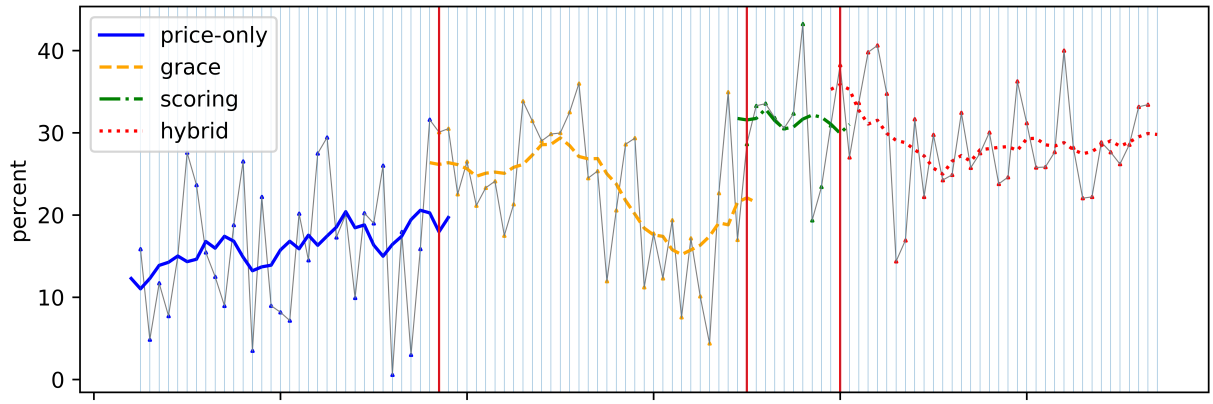
the 12 categories and for four time periods: before the suppliers were informed of the true motivation for the digitalized audits (Pre $t1$), after they received this information but before the introduction of scoring auctions (Post $t1$), during the scoring auctions (SR Period) and after the hybrid price-only auctions (Post SR). Across nearly all categories, there is a sharp increase between the Pre $t1$ and Post $t1$ periods. The increase is more moderate in the latter periods. For instance, for the two most audited classes “Works site regularity" and “Works site safety", the increase between the first two periods is stunning: from 10 percent to 61 percent and from 31 percent to 75 percent respectively. By contrast, the change observed between the latter periods is more modest: from 84 percent to 94 percent and from 92 percent to 96 percent respectively.

The striking increase in compliance for the audited parameters is clearly shown by Figure 2: full compliance for all parameters, by all the firms, audited in a given month would set the blue line equal to 100, but we see that the compliance level in 2007 is only between 20 percent and 30 percent. The vertical, red line marks the date of the first announcement: it is evident how performance improves after this date. Moreover, the long time series of available data provides a rare opportunity to observe the long-lasting impact of this reform which entails average compliance settling at around 90 percent. As discussed at the end of this section, this is the case even after a legal controversy led to the dismissal of the price-and-rating system and its replacement with a hybrid mechanism. Furthermore, Figure 3 shows that all suppliers improved their performance, albeit with different timing. By pooling suppliers into 4 groups depending on the frequency with which they win, the positive trend in compliance is evident for all of them. The higher performance by those suppliers winning less often should not be surprising: these are the firms bidding less aggressively, thus winning less, but delivering higher quality.

The second dataset contains data on the awarding of public procurement auctions. Figure 4 shows the evolution of the winning discount in the Acea auctions. Since, as shown in Figure 1, the auctions are less regular than audits, we use days when auctions were held as observations.

By combining internal Acea data with data from a private provider of data on public contracts (Telemat), we obtained a dataset covering the universe of auctions held between

Figure 4: Evolution of discounts over time.



Note: average (per auction day) discount of the winning firm and it's rolling window average (over 5 auction days). Timeline is split into four periods: price-only (before t1), grace (after t1), scoring and hybrid.

2005 and 2016 for the type of maintenance jobs involved in Acea's reform.[16] The data include the object of the contract, the reserve price, the award price 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. For a subset of auctions, we integrate the data with the information on losing bids and on the subsequent life of the contracts using data from the authority supervising public contracts (ANAC).

Table 3 reports summary statistics for the auction data, dividing them into three panels. The top panel describes the data during the pre-announcement period (i.e., 01/2005-11/2007). The middle panel covers the data after the first announcement ($t1$), but before the SR was implemented. The bottom panel presents statistics for the later periods after the SR was introduced. The first two panels report the data for both Acea and the control group, the last one reports data for Acea only, but separately for the SR and post-SR periods. The main outcome variable for the price analysis below is the winning discount.[17]

[16]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, by using textual search methods, we were able to separate $OG10$ 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., all of those with a reserve price below €10,000 or above €2.5 million).

[17]Bids are percentage discounts relative to the reserve price publicized in the call for tenders. This reserve price is unlikely to be affected by Acea's reform because 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 every year by the

The comparison of the top and bottom panels of Table 3 indicates that the average winning discount in Acea's auctions declines, from 21.73 percent to 18.99 percent, while it grows in the Control group's auctions, from 21.30 percent to 22.95 percent. This suggests that the prices paid by Acea might have increased after the first announcement.

Panel (a) of Figure 5 shown the evolution over time of the winning discount for both Acea and the control group buyers. The date of the first announcement ($t1$) is marked in the figure by the red, vertical line. The figure also reveals a more nuanced pattern for the winning discounts after $t1$ relative to what is visible from the statistics in Panel (b): discounts first increase and then sharply decrease (soon after $t5$). The very different behavior in the control group suggests that this is likely due to Acea's reform and not to changes in market conditions. The following analysis will establish these effects formally.

Regarding the other variables reported in Table 3, 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.[18]

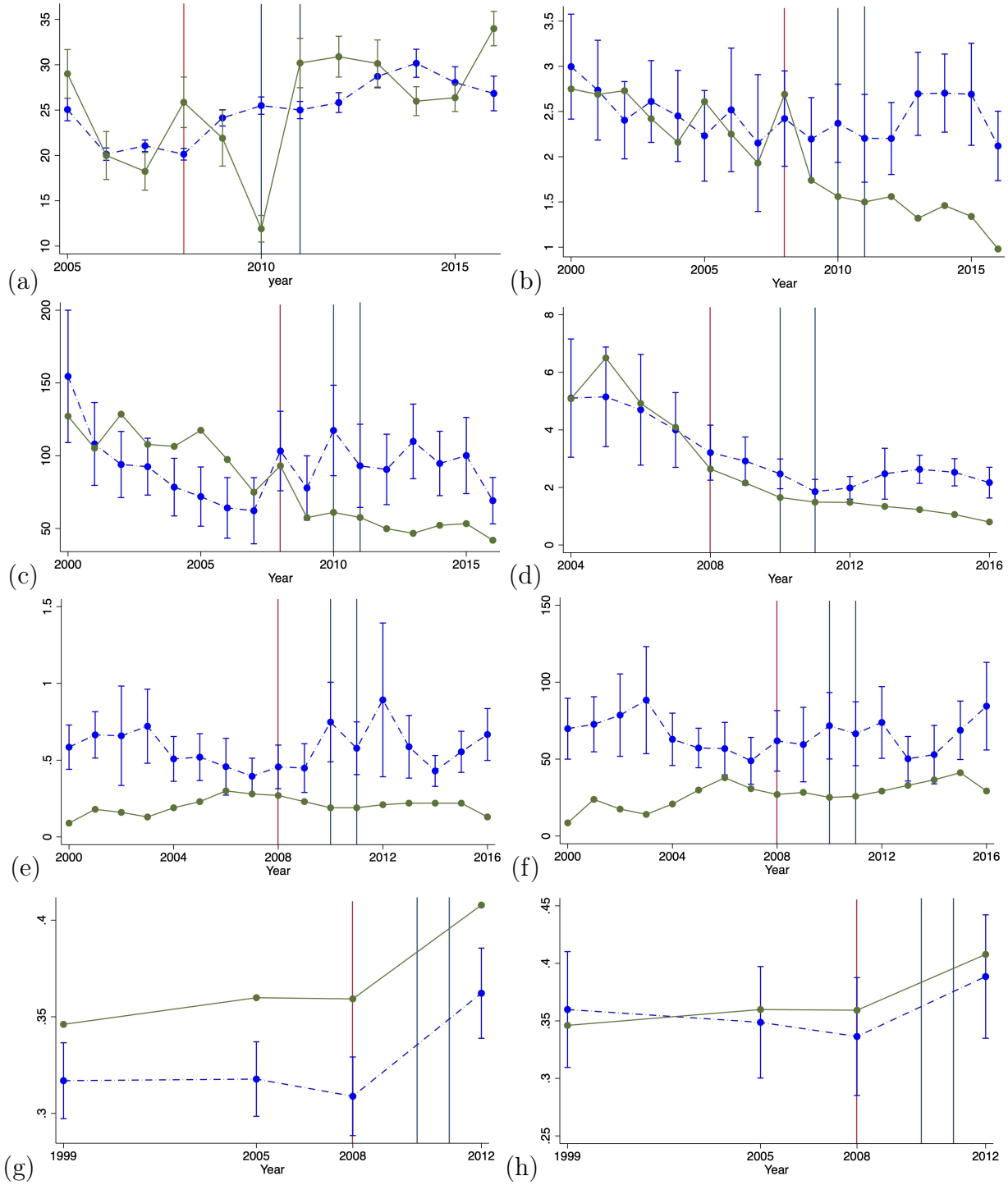
Finally, Panel (c) reports statistics for the period from the introduction of the SR onward. The Difference-in-differences strategy presented next focuses exclusively on the sample periods of Panels (a) and (b). The statistics in Panel (c) are nevertheless interesting to get a sense of the longer-run impacts of the reform. In particular, we observe that during the 35 auctions using the SR procedure, there is a sharp increase in the discounts relative to the earlier period and that this higher discount level is preserved during the following hybrid price-only system. As shown in panel (a) of Figure 5, this increase takes place in the Control group too and, hence, likely reflects some broader trend in the market. Finally, notice that the reserve price is higher in the post-SR period relative to those in the SR period: this is part of a trend in Acea's contracting in order to concentrate its demands into fewer, larger lots.

Our third data is related to external performance measures. In Italy, electricity and water are both partially-regulated sectors. For electricity, although only power transmission

region where Acea operates and used exclusively by contracting authorities to calculate reserve prices.

[18]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.

Figure 5: Evolution of Discounts and External Performance Measures



Note: (a) winning discount (b) long-lasting blackouts (c) duration long-lasting blackouts (d) short-lasting blackouts (e) programmed power cuts (f) duration programmed power cuts (g) water leakage (h) water leakage (pop. over 1 million). Acea (in green) and other utilities (in blue). The red line indicates the date of Acea's first announcement. The blue, vertical lines indicate the scoring rule period. Dashed graph corresponds to the control group.

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 year 2000 to 2016 and cover all low-voltage power distributors, including Acea. Herein, the main indicators of firm performance are constituted by the number and duration of blackouts and programmed power cuts. The top six rows of Table 4 report summary statistics for these external performance measures, none of which is part of the RI parameters.

The external performance measures allow a comparison of Acea's performance to that of other similar firms. In Figure 5, 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 less than 3 hours), (c) duration long-lasting blackouts (in minutes), (d) number of short-lasting blackouts, (e) programmed power cuts and (f) duration programmed power cuts. The observed pattern is similar across most of the measures: after $t1$, Acea's performance gradually improves in both absolute and relative terms. For instance, this is clearly the case for the most socially impactful measure, the number of blackouts experienced by Acea's customers: as the plot shows, this number declines both 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. In the next section, we will explore some additional features linked to Acea's suppliers' behavior that contribute to explaining the slower improvement in external performance measures.

Finally, the last two plots in Figure 5 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 sectors thus serves the role of a

placebo. External performance measures for the water sector have been obtained from the environmental census of the Italian Statistical Institute (Istat). 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 simply aggregated up the water leakage data for all the counties served by Acea.[19] The bottom rows in Table 4 report summary statistics for the water sector, while the bottom panels of Figure 5 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.

4 Reduced Form Analysis

The descriptive evidence so far shows that Acea's reform improved contract performance over the following 10 years. A careful empirical analysis is nevertheless needed to answer three questions crucial to deriving more general implications from this reform. First, what triggered the performance improvement, and, in particular, was it driven by a response to the *announcement* of the scoring auction? Second, what was the effect on prices of the changes in performance? Third, was the improvement in performance confined to the internal measures, or did it also affect the external performance measures? These are interrelated but distinct questions that we will address through different combinations of the data described earlier and with different empirical strategies.

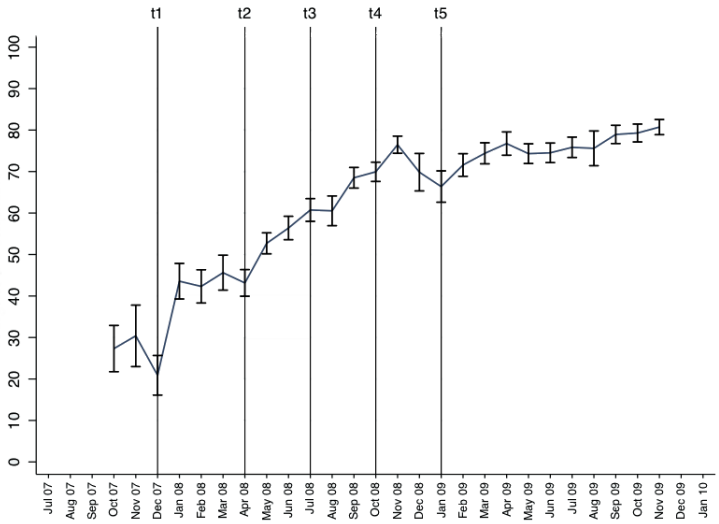
4.1 Internal performance measures

As discussed earlier, most of the performance improvements observed over the long run took place during the first years (see Figure 2). In Figure 6, we focus on this earlier period,

[19] 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.

zooming into the dynamics of performance after the new audit system was introduced but before the switch to scoring auctions. We also add to Figure 6 vertical bars marking each one of Acea's announcements, $t1, ..., t5$. We can visually observe how performance jumps upward after each announcement – except $t3$ – and how it's growing dynamic reduces its speed soon after $t5$. Moreover, the variance declines over time, as shown by the 95 percent confidence interval for the monthly mean.

Figure 6: Average Compliance



Note: The graph shows the monthly average compliance with the internal parameters (audits data). The average is calculated across all the scores recorded in all the audits taking place in the month of reference, weighting each parameter by its weight in the RI. The vertical lines identify each announcement date.

This graphical evidence illustrates what clearly emerged during the reform: suppliers began improving their compliance with the audited parameters well before the scoring rule was introduced and Acea's announcements had a key role in driving this behavior. To formally show the connection between performance changes and announcement timing, Table 5 reports the results of Bai-Perron tests for the presence of structural breaks in the time series of the compliance measure in the same time window as Figure 6. 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). The test results are a clear indication that $t1$ is a breakpoint. As regards the other break

dates, all tests allowing for an unspecified number of breaks identify a break near $t5+1$ (i.e., 1 month after $t5$). This is also quite revealing since, by the fifth meeting, suppliers had found out that average compliance had reached a fairly high level across all active suppliers and parameters. As discussed below, this likely changed the strategic environment in the auctions, through a change in the perceived value of further improvement in compliance.

In Table 6, 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 5) 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 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, we also present additional results exploring the composition of the set of parameters audited and firms inspected.

Finally, an important question is whether we can consider the improved compliance to be the result of strategic decisions by suppliers to improve their performance. Indeed, in experimental settings, the mere change in the environment might trigger forms of *Hawthorne effect* (or observer effect, Levitt and List (2011)). Hence, the mere change from paper-based to digitalized audits might have led suppliers to improve their performance. To rule out this possibility, we can compare how the probability of observing compliant parameter changes between the audits held before and after $t1$. The estimates (reported in the Appendix in Table A.1) show that parameters receiving a higher weight in the announced scoring formula pass from being the ones more likely to be non-compliant before $t1$, to being the most likely to be compliant after $t1$. Furthermore, the parameters more likely to be compliant post $t1$ are those that experts consider faster to adjust.[20] These results are indicative of suppliers effectively changing their behavior.

[20]With the help of expert engineers, we created an indicator variable, *quick*, taking the value of 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 work nearby. The adequacy of the machinery, instead, is an example of a parameter with a quick equal to zero. While clearly arbitrary, this dummy variable is helpful to test the reasonableness of the performance response observed in our data.

4.2 External performance measures

The availability of external performance measures is useful to assess whether the previous findings are the mere result of multitasking by suppliers who refocus their efforts on the set of audited parameters. We evaluate the impact of Acea's announcement at $t1$ on the six external performance measures introduced earlier. This is relevant both 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 measures that are highly socially valuable. The estimation strategy is again a DID based on the following equation:

$$O_{ft} = a_f + b_t + \beta_4 D^{Acea}_{t \geq t_1} + \gamma X_{ft} + \epsilon_{ft}, \tag{3}$$

where O_{ft} is one of the performance outcomes that we observe at the level of buyer, f, and year, t. Notice that in Equation (3) the unit of observation is a buyer, while in Equation (**??**) it is a contract. 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 β_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 7 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 earlier: 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, these estimates confirm the presence of long-lasting performance improvements.

4.3 Evolution of winning bids

In this section, we analyze the evolution of winning bids. In particular, to causally estimate how the initial jump in compliance (associated with the announcement at $t1$) affects the winning auction discount, we consider employing a difference-in-differences (DID) strategy. The units of analysis are the auctions held by Acea (treated group) and by other utility firms (control group). A baseline DID model analogous to that of equation (3), but with contract-level data would be:

$$D^{w}_{ift} = a_f + b_t + \beta_1 D^{Acea}_{t \geq t_1} + \gamma X_{ift} + \epsilon_{ift}, \tag{4}$$

where D^w 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. $Treatment$ is a dummy variable equal to one for the contracts awarded by Acea from $t1$ onward and zero otherwise. The coefficient of interest is β_1, the effect of the announcement on the winning discount, conditional on fixed effects for the entity awarding the contract (a_f) and time (b_t), and on other covariates (X) involving contract characteristics.

In practice, our setting entails dividing the sample period not just in pre and post treatment, but in five periods: before $t1$, the announcement period (from $t1$ to $t5+1$), the post announcement period (from $t5+1$ to the first SR auction), the SR auction period, the post SR period. This subdivision follows both the institutional changes and the break points detected by the Bai-Perron test, especially $t1$ and $t5+1$, when the performance growth slows down. We thus estimate a DID model with four treatment dates. Table 8 presents these baseline estimates. We present the estimates of four specifications differing in the set of covariates and control group observations. For the latter, 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, 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: after the announcement the discounts increase significantly by 6-7 percent of the reseve price, they then rebound in the post announcement phase returning to a level comparable to that beforere the announcement. Afterwards, they

increase again in the SR period (by 3-5 percent) and, even more so in the post-SR auction phase (by 6-8 percent).

Our explanation for these findings is that while firms improved their compliance with the performance measures, they also competed more fiercely to win auctions. Only suppliers with ongoing contracts can be scored and accumulate RI points to be used under the forthcoming SR award system. However, as all firms reached a high score, two forces push toward lower discounts: first, they need less to get additional scores and, second, increasing performance when its level is already high is very costly. Hence, discounts declined in this phase. The estimates indicates a less pronounced decline than that shown by Figure 5 because the regressions control for auction characteristics.[21] That figure also reveals that winning discounts increased once again during the short period in which SR was introduced, to about 30 percent, and remained relatively high afterward.

However, we shall stress that the more transparent interpretation of these reduced form estimates must be restricted to the period before the SR auction was implemented because it is only for this period that we can study the price effects of higher quality without the additional effects produced by the implementation of the SR on bidding and entry behavior. To study the more in depth the SR period, we thus resort below to a structural analsysis.

4.4 Supplier entry

In this subsection we analyze whether the price and performance effects discussed above can derive from a different selection of contractors induced by the reform. We focus on entry decisions. While the summary statistics show that the number of bids submitted remains stable and approximately equal to 11 both before and after $t1$, the set of bidders changed in Acea's auctions: while there are 34 suppliers placing at least one bid both before and after $t1$, there are other 36 suppliers who place at least one bid before $t1$, but no bid afterward. We refer to the latter group of firms as "*exiters*" and to the former as "*stayers*." There are also 3 new entrants placing bids only after $t1$, but never before then. This implies that

[21] The control variable driving most of this difference is a dummy for whether the auction is an "average bid auction." This is a form of modified first-price auction incentivizing low discounts. See the discussion below.

Figure 7: Exit dates

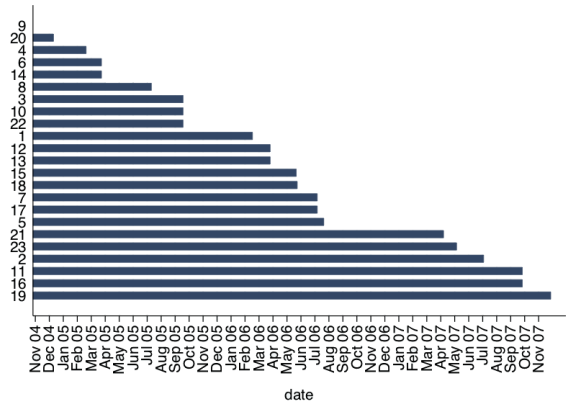
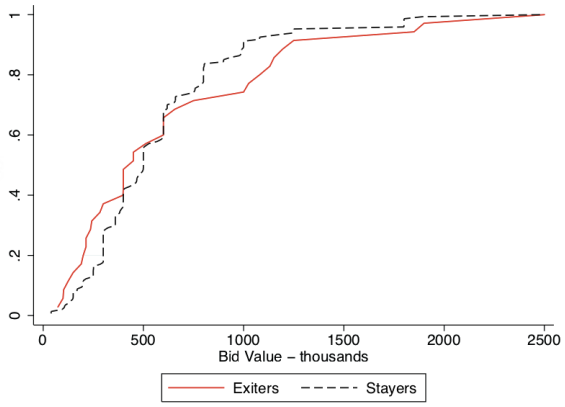


Figure 8: Bid CDFs



the average number of bids placed per bidder doubles; from 0.16 (i.e, 11/70) to 0.30 (i.e., 11/37). This increased participation is due to the *stayers*, not too unusually high bidding frequencies for the 3 new entrants and is likely driven by the same incentive to earn RI that we discussed when analyzing the evidence on winning bids. As regards *exiters* and new entrants, however, their mere presence potentially indicates that the reform might have also triggered some selection effects.

If we focus on *exiters*, however, the data provide only weak evidence of possible selection effects. In particular, Figure 7 shows the timing of the exits (numbers on the vertical axis represent anonymized firms)[P] does not seem clearly linked to $t1$. This figure reports the last date at which each of the *exiters* (represented by the numerical identifiers on the vertical axis) placed a bid. The smooth path of exits indicates more of a gradual process than a sharp drop at $t1$.

Furthermore, as illustrated by Figure 8, if we compare the cdf of winning bids by both *exiters* and *stayers* (in the pre-$t1$ auctions), we do not observe significant differences.

Finally, even in terms of characteristics, *exiters* do not seem to be substantially different from *stayers*. Table 9 reports summary statistics for the subset of *exiters* and *stayers* that we could match to the Infocamere database, the Italian firm registry.[22] Statistics for the *exiters* are reported in the first four columns of panel (a), followed by statistics for the *stayers* in the following four columns. Along most dimensions, *exiters* are smaller than *stayers*; this is the case for revenues, profits, and capital. The average number of employees is also lower, but in

[22]The registry covers nearly all Italian firms, for a description see Conley and Decarolis (2016).

this case, the median is nearly identical. For both groups, the wide variation in characteristics among firms means that the differences in the averages are not statistically significant and it is not obvious how to interpret the results. Thus, to benchmark them we present in panel (b) the analogous statistics obtained for the suppliers active in the auctions of the multi-utility company of the city of Turin. This is the multi-utility company that awards most contracts within the DID control group. Analogously to what was done for the Firm, we partition its suppliers into those bidding both before and after $t1$ (*stayers*) and those bidding only before $t1$ (*exiters*). The comparison of the two groups leads to similar conclusions to those found for the Firm's suppliers: the average revenues, profits, and capital are higher among *stayers*. But the data are again characterized by many extreme observations and the result is reversed for revenues and profits when looking at the median.

5 Stylized model

In this section, we propose a stylized model of a multidimensional (i.e., scoring) auction, where non-price characteristics of the firm (i.e., her quality) are related to her past performance. The firm can invest in her quality by performing better. However, since performance is measured in past contracts, from the current viewpoint, the associated costs are effectively sunk. This is in stark contrast to 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[23]. Let $0 \leqslant \theta_i \leqslant 1$ be firm i's efficiency parameter, such that her costs of executing the auctioned contract, i.e., her *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

$$C_i(q) = \frac{(q - \underline{q}_i)^2}{2\beta}, \tag{5}$$

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

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

and d_i is her discount (i.e., the difference between the reserve price and her bid).[24] 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 $\sigma : \rho \to s$. Denote the equilibrium distribution of score as $F_s(.)$, then each firm maximizes

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

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

$$(\rho - s)\frac{\partial G}{\partial s}(s) - G(s) = 0, \tag{6}$$

$$\alpha G(s) - \frac{\partial C_i}{\partial q}(q) = 0. \tag{7}$$

Equation (6) 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

$$\sigma(\rho) = \int_{\underline{\rho}}^{\rho} z dH(z)/H(\rho), \quad H(\rho) = \sum_{i=2}^{N} p_n F_\rho^{n-1}(\rho) \tag{8}$$

where $F_\rho(.)$ is the equilibrium distribution of pseudo-type and $\underline{\rho}$ is the lowest participating pseudo-type. Equation (7) 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. [25]

Our first observation is that there must be full participation (i.e., no endogenous entry).

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

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

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

$$\eta_\alpha(x) := \sum_{n=2}^{N} p^n \xi_\alpha^{n-1}(x,y), \quad \xi_\alpha(x) := Prob(\theta_i + \alpha \underline{q}_i \leqslant x), \tag{9}$$

and assume that they are continuously differentiable and strictly monotone. Since, in equilibrium, $G(s) = H(\rho)$, it follows from (5), (7) and (9) 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 ρ_i for each $(\theta_i, \underline{q}_i)$. Moreover, the pseudo-type ρ_i is necessarily monotone in θ_i.

With the pseudo-type at hand, we can compute equilibrium quality and score via equations (7) and (8), 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. *If $\xi_\alpha(.)$ is continuously differentiable and strictly increasing in the support of $(\theta_i, \underline{q}_i)$, then there exists a unique symmetric equilibrium, with full participation and strictly monotone strategy σ.*

We split the remaining analysis into two separate cases.

5.1 Univariate types.

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

Recall that the score is monotone in ρ and therefore θ. Furthermore, since all firms have the same parameter $\underline{q}$, the firm with the highest θ wins, independently of α. This leads to several important conclusions. First, the winner does not depend on α thus winner's quality will increase by (7). Second, the firm's interim expected profits are fixed by the revenue

equivalence principle, and she is therefore 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 her total profits net of the investment costs, thus her expected discounts are lower in the scoring auction than in the price-only auction.

Proposition 2. *In the symmetric equilibrium with univariate types, the expected quality is higher, and the expected discounts are lower, in a scoring auction than in a price-only 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, θ_i distributed uniformly on $[0, 1]$, and $\beta = 1$ in the table 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.

5.2 Bivariate types.

Let the firms vary in both θ_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 informational rents.

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

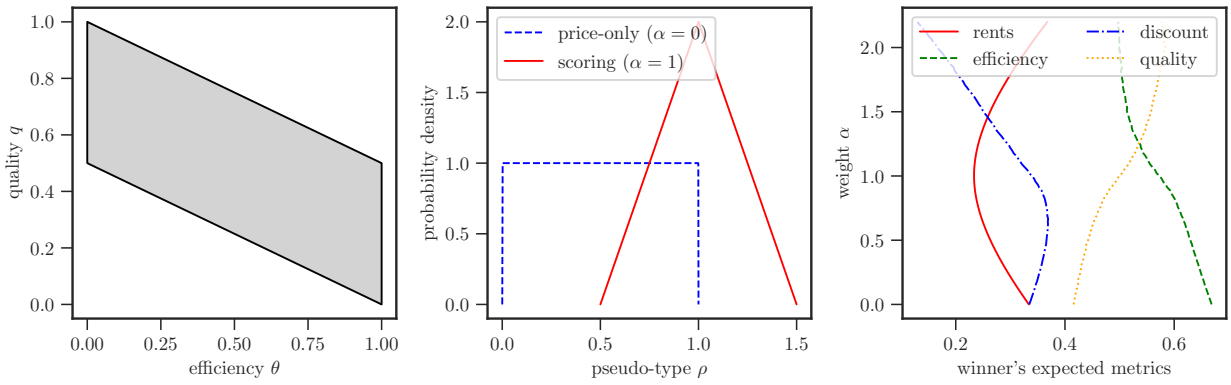


Figure 9: Bivariate types example.

However, the strategies (8), 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 informational rents are likely to decrease, due to a more aggressive equilibrium strategy σ. Thus, while a switch to the scoring auction, through selection, necessarily leads to a decrease in the efficiency of the winning firm, it may lead to an increase in discounts, if the informational rents of the winning firm decrease 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 θ' and quality $\underline{q}'$ wins, but in the scoring auction another firm with type θ'' 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, the expected quality is higher, but the expected discounts can be higher or lower, in a scoring auction than in a price-only auction.*

To illustrate the idea, we pick a distribution such that q and θ are slightly negatively correlated in the scoring auction. Namely, we take two firms and a uniform distribution of

(θ, q) in the region defined by $0 \leqslant \theta \leqslant 1$ and $1/2 \leqslant q + \theta/2 \leqslant 1$, see Figure 9 (left).[26] 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 smaller informational rents.

design (weight)	$\partial H(\rho)/\partial\rho$	$H(\rho)$	expected winner's efficiency	expected winner's info. rent	expected winner's discount
price only ($\alpha = 0$)	$1,\ \rho \in (0,1)$	ρ	$2/3$	$1/3$	$1/3$
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 comparison between $\alpha = 0, 1$ in the Table above, see Appendix B for the derivation. One can see that, while the winning firm's efficiency (type θ) has decreased through selection, the informational rents have 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 α, see Figure 9 (right), holding firm's quality fixed (which corresponds to β approaching 0). Both expected quality and expected discount are increasing in the α weight, in the range between 0 and roughly .6.

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

6 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.

[26] To rationalize the observed distribution of (θ, q), one has to verify that $\rho - \alpha^2\beta H(\rho)$ is increasing for all ρ in the support. For $\alpha = 1$ it would suffice that $\beta < 1/2$.

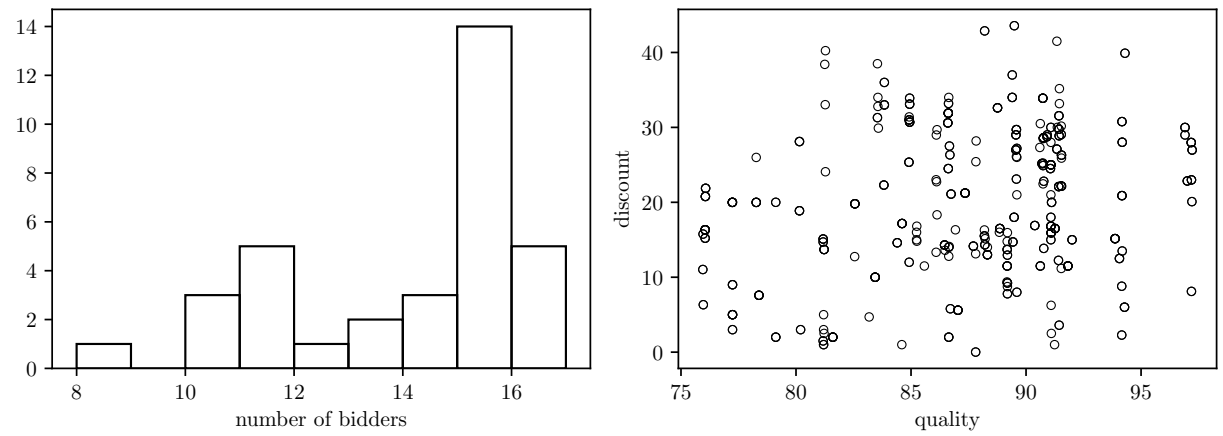


Figure 10: Distribution of data

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[27] 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

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

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.

6.1 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 ρ, score as s, and the best possible discount the firm can offer for the contract as θ.

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 θ. Since we observe the joint distribution of (s, q), through the optimality conditions (6), we observe the joint distribution of (ρ, q) and, therefore, the marginal distribution of θ is identified. The joint distribution of $(\theta, \underline{q})$ is not identified, since β - the parameter of investment costs - is unknown.

We stress that, even with unknown β, we still have access to the monetary outcomes in the price-only auction ($\alpha = 0$), most importantly, discounts. In fact, these particular counterfactuals do not depend the shape of the cost function, and are therefore robust to misspecification.

6.2 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 γ. 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.3, to obtain the estimates of fixed effects γ.

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 (6), see Andreyanov and Franguridi (2021) for details. For any value of β 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.

6.3 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 informational rents (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 of discounts (27% as opposed to 26.81%), but this change is statistically insignificant. We stress that all of these results do not depend on the choice of β.

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 β approaches 0. Thus, for a new scoring weight, α' we only have to calculate the

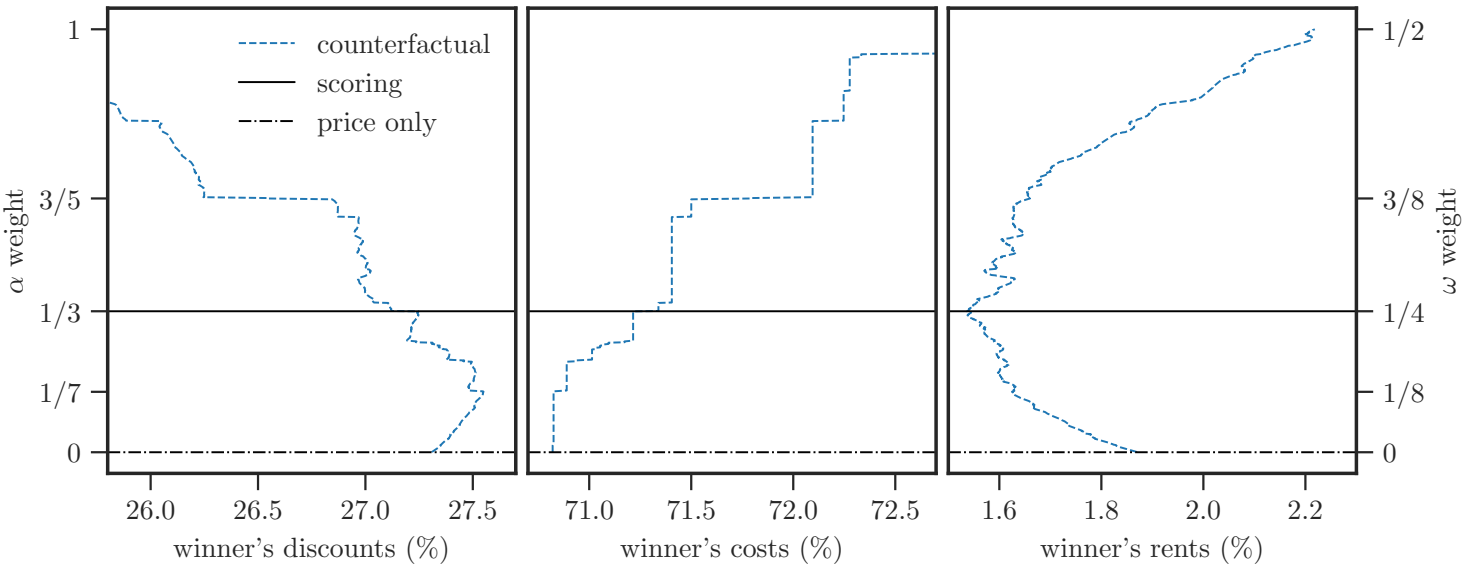


Figure 11: Counterfactuals

new pseudo-type distribution $F_\rho(\cdot|\alpha')$ and re-evaluate the informational rents. 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, for the expected winner's informational rents, the relationship is not monotonic. This leads to the observed discounts in the scoring auction being very close to those in the counterfactual price-only auction, as the increase in costs is partially offset by the decrease in rents.

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

7 Cost-effectiveness Analysis

We conclude our analysis with a back-of-the-envelope cost-effectiveness analysis comparing outcomes under Acea's reform and under the status quo absent any reform. An exhaustive cost-benefit (or welfare) analysis would require assigning a monetary value to the increased

compliance on all parameters. In the spirit of the cost-effectiveness approach, we focus on a subset of specific outcomes.

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. We thus convert the estimate in column 2 of Table 7 into a measure of the number of hours 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),[28] we associate a cost of blackouts of 2.5 euro/hour for residential customers and of 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 are most likely covering safety parameters.[29] Construction and maintenance jobs for electricity generation and transmission are among those with the highest incidence of workplace accidents, including deadly accidents.[30] The occurrence of such accidents has costs for both society and Acea, and the 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.[31] 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

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

[29] This subset of parameters is identified with an * in Table XXX. The selection of this subgroup of parameters was decided by the Acea engineers.

[30] 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. In the U.S., the Bureau of Labor Statistics recorded a total of 5,587 fatal electrical injuries between 1992 and 2013, an average of 254 fatal electrical injuries each year. Death was due either to electrocution or to fires caused by electricity, see Campbell and Dini (2015).

[31] 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/.

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 3). Suppose that the same rate of compliance observed across audits applies to every working day, then the 55 percent improvement in parameter compliance discussed in sub-section A above 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,[32] the estimated benefit ranges between 3.5 and 5.3 million euro/year.[33]

Finally, regarding the cost, the baseline estimates in Table 8 imply no changes in the winning discounts. Since the winning discount in the auctions is the most relevant cost feature of the reform,[34] this no-effect on prices obviously implies that the reform was highly cost-effective relative to the status quo. This conclusion is robust to worst-case scenario analysis. For this, we calculate the reform's effect on the average winning discount directly from the descriptive statistics by taking the difference between the pre-announcements average (21.73 percent) and the year 2010 average (11.91 percent).[35] This gives a reduction in the winning discount of 9.82 percentage points. At an average contract value of €516.1 thousand and considering 43 contracts per year, the total yearly cost increase is then €2.18 million. Hence, we can conclude that, even under a worst-case scenario, the benefits from adoption would

[32]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 2007 nominal euro. We shall also remark that our approach is quite conservative because for benefits we have employed the lowest OECD estimate of the value of a statistical life. 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 accruing from reductions in non-fatal accidents associated with better safety practices and all improvements in quality associated with increased compliance with the quality parameters.

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

[34]Indeed, according to Acea, carrying out the audits under the new system is no more costly than doing them under the paper-based system.

[35]As shown by Figure 5, using the 2010 average discount as representative of the discount level after the reform is the worst-case scenario. If we were to consider the average across the whole period between $t5+1$ and the scoring rule introduction, the level would be higher at 16.19 percent.

exceed the costs.

8 Conclusions

This paper has studied the merits of using past performance audits to spur greater efforts from contractors when 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 are reflected in higher service quality by the utility. Regarding prices, we find some evidence of an initial drop followed by a moderate price increase. Overall, price effects appear negligible when compared to the substantial improvement in performance, as confirmed by a cost-effectiveness analysis involving the duration of blackouts. As in List (2006), the two elements (measurable past performance and repeated interactions) are complementary in the success of the policy. However, we also highlight the sophisticated equilibrium effects produced by the revised procurement system. By quantifying the loss in firm markups, we offer a rationalization for why a system that was successful on many grounds was nevertheless opposed by bidders and ultimately dismissed by the buyer.

The empirical evidence in this paper points to the very large benefits of implementing rating mechanisms in public procurement for the government and taxpayers.

Once the merits of this kind of rating mechanism in improving contractor performance are proven, many 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 both Europe and the US on the use of the past performance of contractors 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.[36] Interestingly, the EU follows a very different system, essentially barring the use of past performance with the exception of extremely severe types of misbehavior sanctioned by the judiciary (Gordon and Racca, 2014). Indeed, the use of mechanisms based on past performance has been one of the most contentious issues in the debate leading up to the 2004 and 2014 EU Procurement Directives.[37] To this debate, our results offer a clear empirical illustration of the potential benefits of a rating mechanism based on objective and clearly targeted past performance measures.

Finally, we shall explicitly discuss the issues of external validity and scalability, to which we devote a specific section in the appendix. 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 clearly satisfied in our setting 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 and have been shown to be of limited relevance for the sellers; (iii) naturalness of the setting, choices, tasks and time frame observed is also clearly satisfied 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 in this sense are the ability to develop a large-scale monitoring system and the corruption risk. For both problems, however, the implementation of a system like the PPIRS in the

[36]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).

[37]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.

US shows that large-scale monitoring of contractors is feasible (and more so every day with the deployment of IoT) and that corruption risk can be contained via transparency of the public procurement auctions. 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 firms' performance in different countries and for monitoring purposes.

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). For the total Turnover (bln $), the values in parenthesis refer to power only.*

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

Parameter Category	Share Compliant Parameters				Number of observations
	Pre t1	Post t1	SR period	Post SR	
Documentation	0.33	0.65	0.84	0.93	53,121
Equipment and machinery	0.70	0.93	0.96	0.95	44,266
H.T. works site controls	.	0.79	0.93	0.97	2,507
Personnel	0.32	0.67	0.91	0.96	21,513
Works execution	0.19	0.84	0.97	0.98	30,663
Works site regularity	0.10	0.61	0.84	0.94	59,531
Works site safety	0.31	0.75	0.92	0.96	78,338
Works on joints	1	0.96	1	1	1,746
Customer relationship mgnt	1	0.94	.	1	85
Air works	.	0.98	1	1	146
Underground works	0.40	0.69	0.91	0.89	10,450
Works on transformer station	1	1	1	1	268

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: Summary Statistics for the Auctions Data

Panel (a): Pre-announcements (01/2005-11/2007)

	Acea			Control		
	Mean	SD	N	Mean	SD	N
Winning Discount	21.73	10.51	172	21.30	10.19	2020
Winning Bid	516.1	428.6	172	445.5	522.4	2020
Length (days)	401.6	179.1	172	327.8	340.4	1788
Num. Bids	10.69	4.305	172	-	-	-
Public Illumination	0.180	0.386	172	0.266	0.442	2020
Central Region	1	0	172	0.202	0.402	2020
Municipal Firm	1	0	172	0.390	0.488	2020

Panel (b): Post-announcements & before SR period (12/2007-03/2010)

	Acea			Control		
	Mean	SD	N	Mean	SD	N
Winning Discount	18.99	10.40	138	22.95	11.60	2247
Winning Bid	516.1	313.5	138	384.9	468.1	2247
Length (days)	385.9	146.7	138	354.1	1106.8	1741
Num. Bids	11.21	4.337	138	-	-	-
Public Illumination	0.232	0.424	138	0.265	0.442	2247
Central Region	1	0	138	0.197	0.398	2247
Municipal Firm	1	0	138	0.395	0.489	2247

Panel (c): SR and hybrid price-only periods (04/2010-12/2016)

	SR period			Post SR		
	Mean	SD	N	Mean	SD	N
Winning Discount	28.76	7.292	35	28.16	6.631	159
Winning Bid	513.2	260.5	35	884.6	616.1	159
Length (days)	421.3	98.24	35	421.3	183.6	159
Num. Bids	13.52	2.336	35	12.42	4.568	159
Public Illumination	0.629	0.490	35	0.245	0.432	159

Note: selected summary statistics for the auction data. "Control" sample consists of auctions held by CAs other than Acea. Panel (a) covers auctions held before $t1$ by both Acea and Control units; Panel (b) covers auctions held at or after $t1$ (and before the switch to SR) by both Acea and Control units; Panel (c) covers Acea's auctions held under either the SR (left panel) or the hybrid price-only (right panel) systems. The definition of the variables is as follows: Winning Discount is the discount (over the reserve price) offered by the winning supplier, Winning Bid is the price bid by the winning supplier, 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 the contract type is classified by Acea 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 the Acea's auctions.

Table 4: Summary Statistics for the Regulators' Reports (External Performance Measures)

VARIABLES	(1) Mean	(2) St. Dev	(3) Median	(4) Min	(5) Max	(6) N	(7) Source
Long-lasting blackouts ($num/LVlines$)	2.43	2.50	1.76	0	24	1,433	ARERA
Blackouts duration ($min/LVlines$)	94	134	49.40	0	960	1,419	ARERA
Short-lasting blackouts ($num/LVlines$)	2.70	3.90	1.84	0	62	1,286	ARERA
Programmed power cuts ($num/LVlines$)	0.6	1.24	0.30	0	29.50	1,431	ARERA
Duration programmed power cuts ($min/LVlines$)	65.60	114	31.20	0	989	1,428	ARERA
Low voltage users (*thousands*)	365	815	6.42	0	4,664	1,642	ARERA
Water Leakage (%)	0.33	0.09	0.32	0.15	0.74	257	ISTAT
Water users (*thousands*)	893	1,054	491	119	4,341	257	ISTAT

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 5: Breakpoints in the Internal Performance Measure

	F-stat breaks	5 unknown breaks
Number of breaks	4	5
Dates of the brakes:		
Date 1	t1	t1
Date 2	t2	t2
Date 3	t3+1	t3+1
Date 4	t5+1	t5+1
Date 5	-	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 6: Acea's Announcements and Supplier Compliance

	(1)	(2)	(3)	(4)
t1	0.200***	0.195***	0.172***	0.456**
	(0.044)	(0.042)	(0.049)	(0.200)
t2	0.065**	0.060**	0.058**	0.082
	(0.031)	(0.030)	(0.029)	(0.132)
t3+1	0.122***	0.115***	0.138***	-0.107
	(0.024)	(0.023)	(0.023)	(0.109)
t5+1	0.082***	0.067***	0.055***	0.013
	(0.018)	(0.017)	(0.020)	(0.068)
safety parameters	0.164***	0.125***	0.038	0.042
	(0.046)	(0.046)	(0.048)	(0.049)
IP-type contract	-0.051***	-0.054**	-0.039	-0.027
	(0.017)	(0.024)	(0.063)	(0.063)
Firm Fixed Effects	No	Yes	Yes	Yes
Contract Fixed Effects	No	No	Yes	Yes
Month Fixed Effects	No	No	No	Yes
N	963	963	963	963

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 in parentheses.* * $(p < 0.10)$, ** $(p < 0.05)$, *** $(p < 0.01)$.

Table 7: 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) WaterLeakage (full sample)	(7) WaterLeakage (above 1m)
β_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
Buyer & Year FE	YES	YES	YES	YES	YES	YES	YES
R-squared	0.843	0.574	0.826	0.720	0.788	0.816	0.890
Sample	All	All	All	All	All	All	Reduced

*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 either 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 8: Baseline Price Estimates

	(1)	(2)	(3)	(4)
ACEA * announcements	6.32***	6.74***	7.15***	6.48***
	(1.43)	(1.00)	(1.48)	(1.05)
ACEA * post-announcements	0.63	1.65	1.35	1.61
	(4.77)	(4.59)	(4.48)	(4.62)
ACEA * SR	4.16***	3.45**	5.34**	3.00**
	(1.38)	(1.37)	(2.42)	(1.39)
ACEA * post-SR	7.93***	6.86***	7.59***	6.46***
	(1.73)	(1.77)	(2.03)	(1.83)
ABA	-14.03***	-15.58***	-15.75***	-15.47***
	(5.01)	(5.28)	(5.52)	(5.35)
Observations	26822	8874	1762	7416
R^2	0.56	0.64	0.59	0.65
Buyer&Year FE	Yes	Yes	Yes	Yes
Size, region, category	No	Yes	Yes	Yes
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 errors clusters by year and CA are reported in parentheses.* * $(p < 0.10)$, ** $(p < 0.05)$, *** $(p < 0.01)$.

Table 9: Summary stats: Exiting and Incumbent firms

Panel (a): Contractors Entering Acea's Auctions

	Exiters				*Stayers*			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Mean	p50	SD	N	Mean	p50	SD	N
Revenues	8,283	2,458	14,615	24	8,934	5,660	9,401	16
Profits	-21	6	697	24	32	5	73	16
Capital	391	36	788	24	998	47	2699	16
Number of Employees	10.3	5	11.1	24	51.7	4.50	180.4	16
Number of Managers	4.96	2	7.57	24	3.38	2	2.55	16
Proportion Female Managers	0.07	0	0.11	24	0.12	0	0.26	16
Public Company	0.96	1	0.21	23	0.88	1	0.34	16

Panel (b): Contractors Entering Turin's IRIDE Auctions

	Exiters				*Stayers*			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Mean	p50	SD	N	Mean	p50	SD	N
Revenues	7,121	4,795	7,127	18	50,860	2,645	152,410	15
Profits	30	15	256	18	736	9.69	2,283	15
Capital	298	40	505	26	10,319	40	43,370	19
Number of Employees	9.04	9.50	5.53	26	15.1	8	15.8	19
Number of Managers	4.35	3	2.96	23	8.11	5	9.45	19
Proportion Female Managers	0.03	0	0.06	26	0.09	0	0.15	19
Public Company	0.71	1	0.46	24	0.72	1	0.46	18

Firm-level summary statistics. Panel (a) refers to the contractors active in Acea's auctions, while panel (b) refers to the contractors bidding in the auctions of Turin's multi-utility company (IRIDE). Across all multi-utilities in the DID control group, this is the one for which we observe most contracts during the sample period. For both Acea and IRIDE, we indicate as *exiters* those contractors observed bidding at least once before $t1$, but never after then, and as *stayers* those bidding at least once both before and after $t1$. For each of the 4 sets, the columns Mean, p50 and SD report the average, median and standard deviation taken across all firms in the set. Column N reports the number of firms considered. Acea characteristics considered are averaged over the years 2006-2010. They are: revenues, profits and capital (all expressed in €1,000), the number of all dependent workers (Number of Employees and Number of Managers), the fraction of female managers over all managers (Proportion of Female Managers) and the share of public companies.

Table 10: Regression of the reputation index on offered discount (1) for the strongest firm, (2) among 3 strongest firms, and (3) among 5 strongest firms.

	Dependent variable: reputation index		
	(1)	(2)	(3)
2011-01-21	10.968***	8.957***	8.879***
	(0.493)	(0.642)	(0.534)
2011-01-26	5.044***	6.171***	5.759***
	(0.545)	(0.790)	(0.645)
2011-02-15	3.925***	6.026***	5.866***
	(0.747)	(0.847)	(0.735)
2011-02-16	11.177***	9.795***	9.377***
	(0.496)	(0.702)	(0.594)
2011-02-17	6.943***	7.309***	7.339***
	(0.382)	(0.562)	(0.470)
2011-02-18	5.254***	7.012***	6.809***
	(0.598)	(0.774)	(0.661)
2011-02-22	12.020***	9.965***	9.928***
	(0.583)	(0.726)	(0.617)
2011-02-23	12.184***	9.541***	9.450***
	(0.572)	(0.589)	(0.495)
2011-03-22	4.804***	6.500***	6.300***
	(0.587)	(0.655)	(0.551)
2011-03-23	12.120***	9.092***	9.184***
	(0.700)	(0.894)	(0.761)
2011-04-28	6.259***	6.609***	7.185***
	(0.325)	(0.512)	(0.434)
const	90.699***	86.976***	86.076***
	(0.633)	(0.743)	(0.677)
offered discount	-0.204***	-0.094**	-0.104***
	(0.037)	(0.038)	(0.034)
Observations	34	102	170
R^2	0.899	0.288	0.227
Adjusted R^2	0.849	0.201	0.173
Residual Std. Error	0.785	1.996	2.160
F Statistic	17.814***	3.312***	4.212***
Note:			*p<0.1; **p<0.05; ***p<0.01

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 (%)	87.774 (0.101)	89.939 (0.404)	-	-
discount (%)	19.476 (0.164)	26.811 (0.273)	19.516 (0.163)	27.034 (0.214)
firm's cost (%)	79.433 (0.157)	71.586 (0.287)	79.433 (0.157)	71.133 (0.247)
firm's rent (%)	1.091 (0.021)	1.604 (0.106)	1.052 (0.022)	1.832 (0.109)
firm's markup (%)	1.432 (0.03)	2.277 (0.162)	1.394 (0.03)	2.64 (0.168)
firm's cost (€)	390235 (3445)	366273 (1476)	390235 (3445)	363136 (1135)
firm's rent (€)	5474 (124)	8239 (496)	5331 (125)	9363 (539)

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.

Bandiera, Oriana, Andrea Prat, and Tommaso Valletti. 2009. "Active and Passive Waste in Government Spending: Evidence from a Policy Experiment." *The American Economic Review*, 99(4): 1278–1308.

Bandiera, Oriana, Michael Carlos Best, Adnan Qadir Khan, and Andrea Prat. 2021. "The Allocation of Authority in Organizations: A Field Experiment with Bureaucrats*." *The Quarterly Journal of Economics*, 136(4): 2195–2242.

Banfield, Edward C. 1975. "Corruption as a Feature of Governmental Organization." *Journal of Law and Economics*, 18(3): 587–605.

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

Beuve, Jean, and StÃ©phane 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.

Bloom, Nicholas, Renata Lemos, Raffaella Sadun, Daniela Scur, and John Van Reenen. 2014. "The New Empirical Economics Of Management." *Journal of the European Economic Association*, 12(4): 835–876.

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.

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

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.

Conley, Timothy G., and Francesco Decarolis. 2016. "Detecting Bidders Groups in Collusive Auctions." *American Economic Journal: Microeconomics*, 8(2): 1–38.

Coviello, Decio, Andrea Guglielmo, and Giancarlo Spagnolo. 2018. "The effect of discretion on procurement performance." *Management Science*, 64(2): 715–738.

Coviello, Decio, Luigi Moretti, Giancarlo Spagnolo, and Paola Valbonesi. 2018. "Court Efficiency and Procurement Performance." *The Scandinavian Journal of Economics*, 120.

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.

Levitt, Steven D., and John A. List. 2011. "Was There Really a Hawthorne Effect at the Hawthorne Plant? An Analysis of the Original Illumination Experiments." *American Economic Journal: Applied Economics*, 3: 224–238.

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, StÃ©phane. 2000. "Transaction costs and contractual incompleteness: the case of ÃlectricitÃ© 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 cost-effectiveness analysis, the value of a statistical life figures come from the OECD (https://www.oecd.org/environment/mortalityriskvaluationinenvironmenthealthandtransportpolicies.htm), while those for the economic cost of 1 hour of blackout, separately for business and residential customers come from Table 11 whats this table?[P] 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 at the same time, she can first optimize over q conditional on s. This is equivalent to plugging equation (7) 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 θ', and chooses 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 $\int_0^\theta x dx = \theta^2/2$ by the envelope conditions. It does not depend on either α or the sunk nature of costs because the firm with the highest type always wins, see, 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 informational rent, 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 from $s_3 - s_4$, where $\{s_i\}_{i=1}^4$ are four scores, randomly picked in every auction, to validate an additive model of heterogeneity.[38] 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.

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 number of bidders. To test whether the distributions are identical, the statistic is formed

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

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.

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

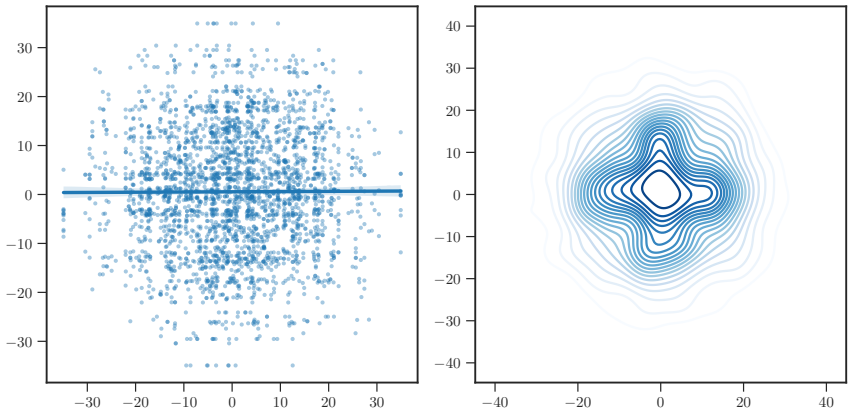


Figure A.1: Scatterplot and contourplot of score differences $s_1 - s_2$ and $s_3 - s_4$.

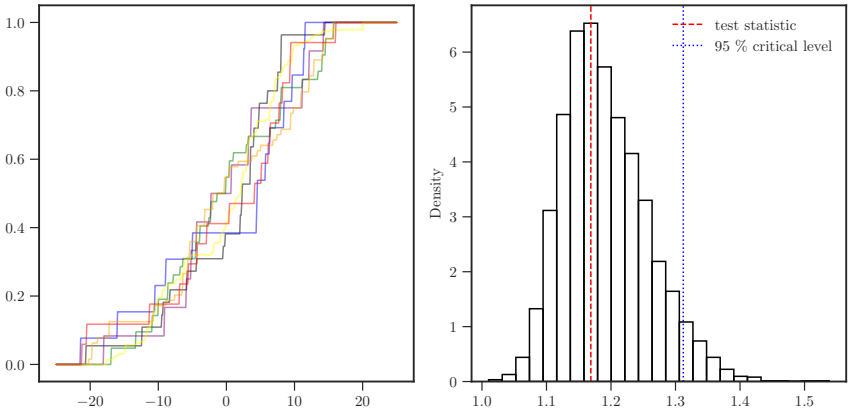


Figure A.2: Empirical CDF's of score residuals (left) with different number of bidders and the distribution of the δ statistic (right) obtained via sub-sampling.

C.2 Quantile approach

To make the optimality conditions more palatable, we will rewrite them in the quantile form. 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

$$Q_\rho(u|\alpha) = Q_s(u|\alpha) + q_s(u|\alpha)c(u), \tag{10}$$

and the envelope conditions as

$$Q_s(u|\alpha) = Q_\rho(u|\alpha) - \frac{\int_0^u C(x)dQ_\rho(x|\alpha)}{C(u)}. \tag{11}$$

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 get rid of the auction fixed effects. Note that this process does not change the ranking of firms within 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 δ_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 an ascending order and denote the new sample as $(q_{(m)}, \hat{s}_{(m)}, \hat{\gamma}_{(m)})$. We aim at using the identifying equation (10) 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 (10), evaluated at an evenly spaced grid:

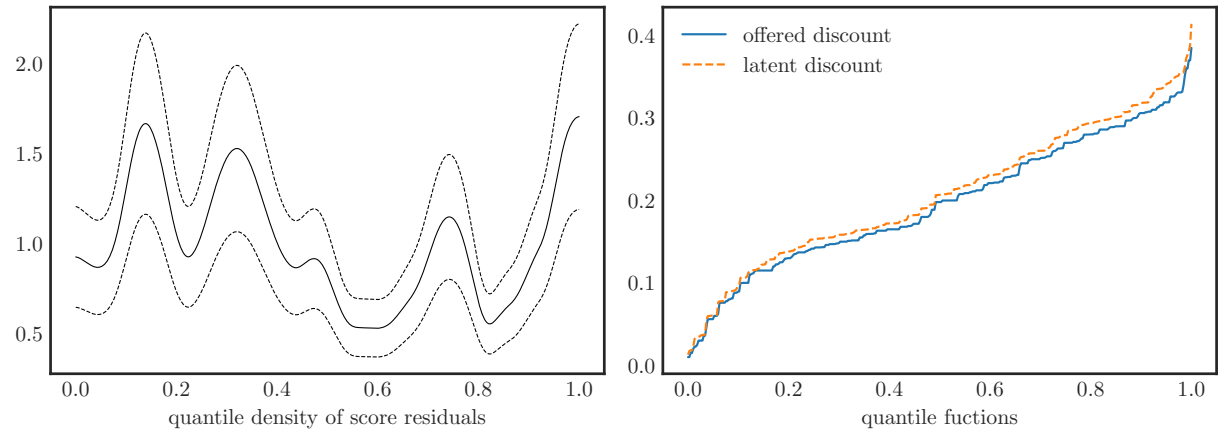
$$\hat{Q}_\rho(u|\alpha) = \hat{Q}_s(u|\alpha) + \hat{q}_s(u|\alpha)\hat{c}(u), \quad u \in \{\frac{m}{M}\}_{m=1}^M. \tag{12}$$

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, and so the pseudo-sample can be constructed. See the results of estimation in Figure A.3.

Figure A.3: Quantile density and functions



While we could, in principle, construct a smooth estimator of $F_\rho(\cdot|\alpha)$ for every α, and use it to evaluate each of the counterfactuals, we find it much easier to use the starting 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 (11), evaluated at an evenly spaced grid:

$$\hat{Q}_s(u|\alpha') = \hat{Q}_\rho(u|\alpha') - \frac{\int_0^{m/M} \hat{C}(x) d\hat{Q}_\rho(x|\alpha')}{\hat{C}(u)}, \quad u \in \{\frac{m}{M}\}_{m=1}^M. \tag{13}$$

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 non compliance; (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 the majority of its generation, transmission

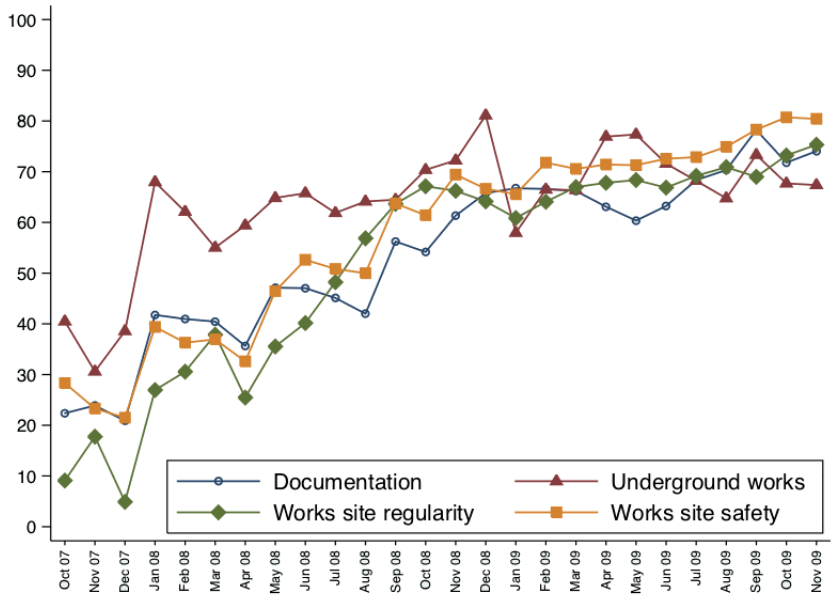
and distribution systems. From this point of view, it is very similar to some of the largest US power operators such as the Los Angeles Department of Water and Power (LADWP), ComEd (Chicago), BGE (Baltimore) and PECO (Philadelphia). For the external validity of what can be learned from the Acea case, 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 evaluation of the policy. The comparability of the observed sample to other suppliers working with the public sector throughout the world is ensured by legal standards. In fact, all suppliers in Italy need to be certified by an external body to participate in public procurement works. Symilar systems exist in the United States and in the rest of the world[39] and ensure comparability across different sectors and countries.

Finally, we explore the issue of selection in the audits performed. Throughout the announcement phase, we see an increase in the number of audits carried out by Acea, however the composition of the audits in the various categories follows the same trend, see Figure A.4, and their proportion remains stable over time. Additionally, the number of audits does not only increase in the announcement phase, but it fluctuates throughout the whole period analyzed (2007 to 2015). This is expected and mirrors the number of auctions held by Acea over time. As more (less) auctions are held, to ensure the same proportion of works are audited, Acea increases (decreases) the number of audits. In Figure A.5 we plot the trends of number of auctions and number of audits held throughout the years.

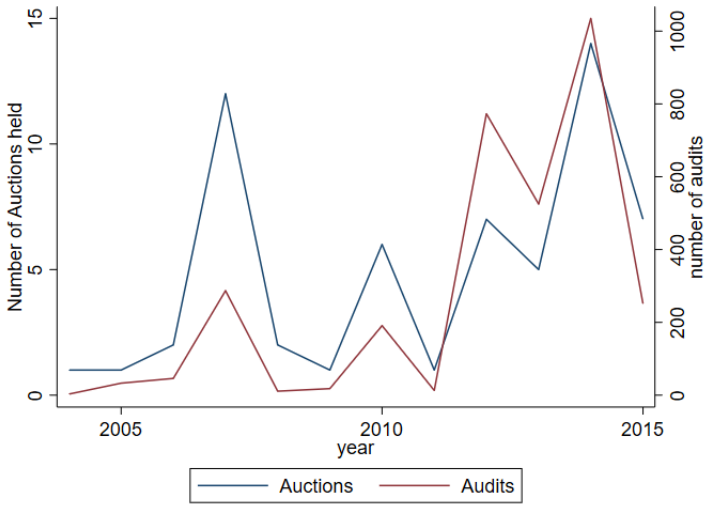
[39]In Italy we refer to the OG and OS qualifications for suppliers working with the public sector. In the United States, a symilar unique system is provided through the System for Award Management https://sam.gov

Figure A.4: Parameters Audited: Evolution of Compliance over Time



Source: Audits data. The lines show the progress of the reputation index calculated on a monthly basis for each of the four most audited Safety and Quality dimensions.

Figure A.5: Evolution of Yearly Auctions and Audits Held



Note: The figure represents the evolution of the number of audits (red) and the number of auctions (blue) held by Acea every year, between 2004 and 2015.

- **Attrition**: In the data we observe attrition in the bidders participating in the auctions. More specifically, we see 34 *exiters*, 36 *stayers* and 3 new entrants after the new scoring

rule is implemented $t1$. While it is true that *stayers* bid more aggressively,[40] we only see limited correlation between suppliers leaving the experiment and the implementation of the new scoring rule. The decrease in the number of suppliers is not sharp around $t1$, but follows a general declining trend, as shown in Figure 7.

The more aggressive bidding behavior of *stayers* does not imply more success in winning the auctions, nor *exiters* experience more losses on average. As illustrated in Figure 8, if we compare the cdf of winning bids by both *exiters* and *stayers* (in the pre-$t1$ auctions), we do not observe significant differences. Even in terms of characteristics, *exiters* do not seem to be substantially different from *stayers*. Table 9 reports summary statistics for the subset of *exiters* and *stayers* that we could match to the Infocamere database, the Italian firm registry.[41] Along most dimensions, *exiters* are smaller than *stayers*; this is the case for revenues, profits and capital. The average number of employees is also lower, but in this case the median is nearly identical. For both groups, the wide variation in characteristics among firms means that the differences in the averages are not statistically significant and it is not obvious how to interpret the results. Thus, to benchmark them we present in panel (b) the analogous statistics obtained for the suppliers active in the auctions of the multi-utility company of the city of Turin. This is the multi-utility company that awards most contracts within the DD control group. Analogously to what was done for Acea, we partition its suppliers into those bidding both before and after $t1$ (*stayers*) and those bidding only before $t1$ (*exiters*). The comparison of the two groups leads to similar conclusions to those found for Acea's suppliers: the average revenues, profits and capitals are higher among *stayers*. But the data are again characterized by many extreme observations and the result is reversed for revenues and profits when looking at the median.

Overall, the attrition rate of suppliers does not raise concerns for the following reasons: (i) the decrease in number of suppliers is not sharp at $t1$, but follows a smooth decreasing trend. (ii) We do not see any structural difference in the composition of the

[40] The average number of bids pre and post $t1$ is stable at 11 bids per auction, the number of bids per supplier almost doubles after $t1$ from 0.16 (i.e, 11/70) to 0.30 (i.e., 11/37)

[41] The registry covers nearly all Italian firms; for a description see Conley and Decarolis (2016).

leaving contractors from the ones remaining. (iii) We do not have any reason to think there are unobservable factors driving the probability of exiting from the experiment.

- **Naturalness**: The trade off between price and contract performance is a common problem to both public and private procurement. However this is heavily emphasized in the public sector to avoid 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.[42]

  The European legislation is more strict in this sense.[43] 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

[42]The PPIRS program has been recently withdrawn and integrated in 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.

[43]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 firms performance in different countries and for monitoring purposes.

Overall, we conclude that the results presented are a 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 a series of additional results supplementing the various analyses presented in the main text.

- The estimates in Table A.1 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:

$$Pr(compliant) = \Phi[t + f + \alpha\ weight + \theta\ quick + \gamma_j \sum_{j=2}^{12} category_j], \qquad (14)$$

where Φ 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.1 shows the probit marginal effects 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.[44]

Regarding the other coefficients in Table A.1, 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 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. However, it is relevant here that 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

[44] All estimates in Table A.1 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.

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.1: Probability of Compliant Parameter

	Pre-announcement				Post-announcement			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Weight	-0.026***	-0.024***	0.000	-0.025***	0.011***	0.013***	0.013***	0.013***
	(0.005)	(0.007)	(.)	(0.007)	(0.001)	(0.001)	(0.001)	(0.001)
Quick		0.077*	0.000	0.074*		0.066***	0.065***	0.066***
		(0.036)	(.)	(0.036)		(0.006)	(0.006)	(0.006)
C2-Documentation		-0.412***	0.000	-0.440***		-0.284***	-0.268***	-0.270***
		(0.053)	(.)	(0.055)		(0.010)	(0.010)	(0.010)
C3-Works Execution		-0.518***	0.000	-0.523***		-0.189***	-0.189***	-0.192***
		(0.062)	(.)	(0.064)		(0.010)	(0.010)	(0.010)
C7-Underground works		-0.302***	0.000	-0.296***		-0.291***	-0.288***	-0.286***
		(0.051)	(.)	(0.052)		(0.009)	(0.009)	(0.009)
C9-Personnel		-0.308***	0.000	-0.332***		-0.349***	-0.359***	-0.365***
		(0.069)	(.)	(0.069)		(0.011)	(0.011)	(0.011)
C10-Works site regularity		-0.673***	0.000	-0.680***		-0.449***	-0.443***	-0.441***
		(0.054)	(.)	(0.056)		(0.009)	(0.009)	(0.009)
C11-Works site safety		-0.381***	0.000	-0.405***		-0.272***	-0.272***	-0.275***
		(0.056)	(.)	(0.057)		(0.010)	(0.010)	(0.010)
Year Fixed Effects	No	No	Yes	Yes	No	No	Yes	Yes
Firm Fixed Effects	No	No	No	Yes	No	No	No	Yes
N	1702	1374	1374	1374	56085	44653	44653	44653

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 $t1$.

- In Table A.2 an augmented version of Table 6. The table reports the estimates of a series of probit regressions for the probability of passing the audits in the various periods. Namely: $tPre$ 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). tSR indicates the effective scoring rule period (18 May 2010 to 23 June 2011). $tPost$ indicates the period of hybrid system (from 23 June 2011 onwards). We progressively control for

several confounding factors: the winning bid's aggressiveness (*Win bid - Average bid*), the number of participants to the auctions, a series of contract specific controls, and firm fixed effects. The time coefficients are generally large and significant, and they show an increasing trend from $tPre$ to $tPost$. 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.

Table A.2: Probit Audit Passed

	(1)	(2)	(3)	(4)	(5)
tPre	-0.931***	0.231	0.245	-0.160	-0.157
	(0.097)	(0.156)	(0.157)	(0.170)	(0.172)
t1	-0.263***	0.775***	0.782***	0.327**	0.328**
	(0.035)	(0.127)	(0.127)	(0.142)	(0.143)
t2	0.063***	1.126***	1.129***	0.720***	0.722***
	(0.022)	(0.123)	(0.123)	(0.137)	(0.138)
t3+1	0.625***	1.724***	1.736***	1.348***	1.350***
	(0.025)	(0.124)	(0.125)	(0.140)	(0.141)
t5+1	0.869***	1.811***	1.827***	1.382***	1.384***
	(0.024)	(0.121)	(0.122)	(0.139)	(0.139)
tSR	1.347***	2.012***	2.049***	1.820***	1.824***
	(0.040)	(0.130)	(0.131)	(0.141)	(0.144)
tPost	1.777***	2.347***	2.382***	1.877***	1.882***
	(0.010)	(0.124)	(0.125)	(0.137)	(0.142)
Win bid - Avg bid	0.005***	0.004**	0.007***	0.009***	0.009***
	(0.001)	(0.002)	(0.002)	(0.002)	(0.002)
Number of offers			-0.007***		-0.001
			(0.002)		(0.004)
Observations	116,339	116,234	116,234	83,001	83,001
Supplier FE	NO	YES	YES	YES	YES
Auction Controls	NO	NO	NO	YES	YES
Number of Offers	NO	NO	YES	NO	YES

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.

Table A.3: 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	

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..*