



**HAL**  
open science

# Quality of Outsourced Services, Opportunism and Contract Design

Jean Beuve, Lisa Chever

► **To cite this version:**

Jean Beuve, Lisa Chever. Quality of Outsourced Services, Opportunism and Contract Design. Review of Law and Economics, 2017, 13 (3), 10.1515/rle-2014-0053 . hal-02139517

**HAL Id: hal-02139517**

**<https://hal.science/hal-02139517>**

Submitted on 24 May 2019

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Jean Beuve\* and Lisa Chever

# Quality of Outsourced Services, Opportunism and Contract Design

DOI 10.1515/rle-2014-0053

**Abstract:** This paper estimates the impact of contract design on the enforceability of quality in outsourced public contracts. Using panel data from 102 cleaning contracts, our results suggest that quality enforcement depends highly on the *ex ante* contracting process. Such findings are consistent with recent theoretical contributions which consider that the verifiability of agents' actions is endogenously determined by the principals' investments in drafting an explicit contract pertaining to the quality of the agent's output. Moreover, since those quality improvements are not accompanied by price increases, our results suggest that public contract managers have significant leeway to reduce opportunistic behavior for standard transactions.

**Keywords:** outsourcing services, quality, enforcement, contract design, monitoring

**JEL Classification:** D82, L15, L24

## 1 Introduction

In 2009, a public buyer operating in the field of social housing disqualified a candidate during an open call for tenders, arguing that the candidate was guilty of providing a low level of service quality in past cleaning contracts. The court, siding with the dismissed private operator, found that the argument used to disqualify the claimant was unlawful and sentenced the public buyer to reorganize the call for tenders. This judicial decision triggered an important organizational change on the public buyer's side. Confronted with a statutory ban on eliminating a firm's candidacy on the basis of bad past performance,<sup>1</sup> the public authority decided to address firms' opportunism by improving the completeness of its contracts. However, since the transaction costs incurred to

---

<sup>1</sup> See Spagnolo (2012) for an extensive discussion on this issue.

---

\*Corresponding author: Jean Beuve, Department of Economics, University of Paris I Panthéon Sorbonne, Paris, France, E-mail: jean.beuve@univ-paris1.fr

Lisa Chever, EPPP Chair – Sorbonne Business School, Paris, France

reduce contractual incompleteness can be substantial (Bajari and Tadelis, 2001), the extent to which this strategy will be relevant is not immediately obvious: it relies highly on the ability of better *ex ante* contracting to improve quality enforcement and on the magnitude of such an improvement.

To date, to deter firms' opportunism in public procurement, the literature has mainly focused on the disciplinary power of competition for the field. In Europe, for instance, public buyers are usually forced by regulation to use an open call for tenders (European Commission, 2004). This mechanism is most often considered as efficient: the overwhelming majority of empirical work on this issue indicates that this procedure allows for reduced costs. Nevertheless, its impact on quality is puzzling, suggesting that those costs savings might be achieved at the expense of quality.

Evidence indeed indicates that public buyers do not apply penalty clauses, even when substantial damage occurs (Spagnolo, 2012): firms may therefore have incentives to post aggressive bids and then reduce their efforts at the execution stage. Previous empirical literature on this subject focuses on the enforcement costs to explain the limited use of penalties (see, e.g., Girth (2012); Coviello et al. (2013)). However, in our case, the public buyer tends to apply them, but a low level of service quality still persists, suggesting that those clauses have a low incentive effect.

We complete previous empirical analyses on quality enforcement issues by investigating the role played by the *ex ante* contracting process in the deterrence of *ex post* opportunism. This approach is in line with the theoretical literature that examines situations in which '*the verifiability of the agent's actions is endogenously determined by the principal's investments in drafting an explicit contract pertaining to the quality of the agent's output*' (Kvaløy and Olsen, 2009). In this literature, investing in the *ex ante* contracting process can reduce contractual incompleteness and thus increase the court's ability to verify that the agent's actions are in line with the intent of the contract. This may be especially efficient in moral hazard settings where the supplier has some private information and needs some incentives to provide efforts (Kvaløy and Olsen, 2016).

To investigate this issue, we use an original database coming from Paris Habitat-OPH, the local public buyer sanctioned by the administrative court in 2009. We have access to monthly information related to 102 allotted contracts signed with a set of private operators in the sector of house cleaning over a four-year period. We look at the impact of the change in contractual completeness (which occurred in April 2010) on performance indicators. We also investigate whether performance controls and penalties provide stronger incentives after the change in contract design. In other words, we use the data we have on the

cleaning industry to test the hypothesis that verifiability is endogenous, *i. e.*, depends on contractual design.

Our findings suggest that bearing the costs of reducing contractual incompleteness significantly improves the delivered quality: more complete contracts not only reduce the average level of moral hazard (direct impact) but also optimize the monitoring of the contract (indirect impact). In addition, we give evidence that quality improvements are made possible without prices rising. As a consequence, our analysis illustrates how an exogenous ‘bad’ event (the court decision<sup>2</sup>) finally leads to an unexpected and profitable organizational change: the transaction costs incurred to reduce contractual incompleteness result in a long-term reduction of firms’ opportunism, do not increase the contract price and make it possible to cut down the costs of contract monitoring. We believe our results contribute to the emerging empirical literature on the issue of public contract enforcement (Girth, 2012; Coviello et al., 2013). Previous studies have investigated the relationship between the institutional environment and the contract enforcement. We depart from them by considering a strategy that is at the discretion of public managers and in line with the theoretical literature on endogenous verifiability (see, *e. g.*, Kvaløy and Olsen (2009; 2016)).

The rest of the paper is organized as follows. Section 2 presents the conceptual framework and puts forward the hypothesis we aim to test. Section 3 then gives some details about the institutional context, the data and our empirical strategy. The results from our estimations are provided in Section 4. Finally, we discuss those results and potential recommendations for the outsourcing of public services in Section 5.

## 2 Conceptual framework

Numerous quantitative studies have compared the costs of in-house public management and costs of private management when contracts are awarded through competitive tendering. An overwhelming majority of those studies conclude that outsourcing achieves a reduction in government expenditures in a great variety of sectors, such as the refuse collection industry (Domberger et al., 1986; Dijkgraaf and Gradus, 2004), road maintenance (Blom-Hansen, 2003),

---

<sup>2</sup> The judicial decision of the administrative court that sanctioned the public buyer was bad news from his viewpoint in that it would oblige the public buyer to evaluate all candidacies in the future, even those coming from poorly performing firms during past contracts. The use of the term “exogenous” is due to the fact that the decision of the court is by definition independent, *i. e.* the public buyer had no possibility to prevent or influence it.

vehicle and warehousing maintenance (Domberger et al., 2002), cleaning and housekeeping (Domberger et al., 2004; Milne and Wright, 2004) and even prisons (Cabral and Saussier, 2013). Competition for the field thus effectively prevents rent extraction by encouraging operators to bid competitively. Nevertheless, abandoned rent on costs can be recovered on the quality of the service (*quality-shading hypothesis*). Currently, although the effect of outsourcing on quality is of fundamental importance for the efficient organization of public services, empirical studies that examine quality issues exist in far fewer numbers than those on cost savings. Moreover, the few existing studies provide mixed evidence. Although some of them suggest that service quality becomes stagnant or rises when contracting out (Domberger et al., 1995; 2004; Savas, 1977), others reach the opposite conclusion of reduced quality following contracting out (Evatt Research Centre, 1990; Hartley and Huby, 1986; Ascher, 1987). An interpretation of those mixed results relies on contract specification and monitoring: as put forward by the Australian Industry Commission (1996), quality issues are primarily a result of poor application of the process of outsourcing rather than outsourcing *per se*.

From a theoretical perspective, we can distinguish two main explanations about why the outsourcing process can entail quality issues. On the one side, a first strand of the literature focuses on the role of contractual incompleteness (see, e. g., Grossman and Hart (1986)): when dealing with some transactions, writing complete agreements might be prohibitively costly. The resulting contractual blanks prevent parties from safeguarding their respective interests. On the other side, even in the case of complete contracts (*i. e.*, contracts for which performances are measurable), *ex post* opportunistic behaviors could be due to asymmetries of information between parties (Jensen and Meckling, 1976): because operators may not share the buyer's goals but may be more familiar with the details of the task, they may have both motive and opportunity to behave in ways that maximize their own interest at the expense of the public authority. Solutions lie in the provision of outcome- or behavior-based contracts.

In the specific case of cleaning activity, contracting on quality is supposed to be rather simple.<sup>3</sup> Consequently, contractual incompleteness should not be an

---

<sup>3</sup> As an illustration, Brown and Potosky (2005) sent a survey to public managers about the transaction cost dimension of a variety of basic local government services to build a typology of 'ease of measurement' for service performance. More precisely, they asked managers to determine this measurement easiness on a five-point scale, giving that '*a service is easy to measure if it is relatively straightforward to monitor the activities required to deliver the service and to identify performance measures that accurately represent the quantity and the quality of the service*'. They also state that for easy-to-measure services, "*government officials can easily write a contract and clearly specify the activities and outcomes for the vendor to perform and*

issue. Moreover, those contracts include incentive schemes that should align parties' interests and limit informational rents. Despite this, quality issues are often raised: in the data we explore about cleaning activities, the persistence of quality issues are illustrated by frequently imposed penalties, users' complaints, breaches and early terminations of contracts. Explanations of such a paradox can be found in the literature on endogenous verifiability. Indeed, the seminal papers of Townsend (1979) and Dye (1985) highlight that costly contracting and imperfect enforcement are important vehicles to understand the nature of transactional relationships. Whereas the classic moral hazard approach assumes perfect enforcement (see, e. g., Holmström (1979)) and models of incomplete contracting consider that contracting is prohibitively costly so legal enforcement is impossible (see, e. g., Grossman and Hart (1986)), more recent contributions overstep those assumptions by dealing with costly contracting and endogenous verifiability (Kvaløy and Olsen, 2009; 2016). The time and effort spent on the contracting process determines the level of verifiability. Contracting on quality is indeed challenging. Even if service quality can be identified in terms of performance characteristics, such an assessment may require subjective judgment rather than the mere accumulation of data (Jensen and Stonecash, 2005). For instance, in the specific case of cleaning services, the only way quality can be measured is through personal observation, and what constitutes a high standard of cleanliness may vary from one observer to another (Domberger and Jensen, 1997). In such a situation, improving contract details can help in reducing ambiguities between parties about the intent of the contract and thus increase its enforceability. This threat might then help to overcome the service providers' temptation to reduce efforts.

The data currently available to us in this study are close to the framework of Kvaløy and Olsen (2016), where contracting is costly and enforcement is probabilistic (the probability that the incentive contract will be enforced by a court of law is determined by the costs spent on contracting). In their paper, better contract specification leads to higher-powered incentives. Nevertheless, as far as we know, there is no hard empirical proof of such statements. This paper contributes to the literature precisely by filling this gap. In other words, we take the example of the cleaning sector to test the following theoretical hypothesis:

*Hypothesis: Verifiability is endogenous, i. e., it depends on the contractual design.*

---

achieve". In this way, they identify very easy-to-measure outsourced activities, such as payroll, commercial solid waste collection and street and house cleaning (score < 2), and very difficult-to-measure ones, such as child welfare programs, drug and alcohol treatment and the operation of mental health programs (score > 4). According to such a typology, contracts established to outsource cleaning services might be rather complete and quality-shading might be rather scarce in this sector.

## 3 Data and empirical strategy

### 3.1 Institutional framework

Over the last few decades, outsourcing activities to external providers has become a fairly common practice for governments. As a consequence, the way it influences the costs of public services has been the focus of academics' and practitioners' interests. The outsourcing of public activities is regulated by European directives, which are then transposed into national law through the 'French Public Procurement Code'. As soon as service contracts reach the EU thresholds (approximately 200 k€), both sets of regulations constrain public buyers to use the traditional open call for tenders.<sup>4</sup> This procurement process is made of different steps. First, the buyer defines its needs and launches publicity. Second, firms send documents related to their candidacy (their references, their number of employees, their competences, etc.) and their offer. Third, the buyer analyzes the different bids. If they are satisfying, the associated offers are also analyzed. Finally, the winner is the 'most economically advantageous offer'. The most economically advantageous offer is the one that best meets all contract award criteria, quality and price set by the public buyer. The choice of the economically most advantageous offer is in contrast to the choice of the lowest bidder, based solely on financial considerations. Award criteria are identified, selected and grouped generally into two categories: those that fall under quality of supply and those that fall under cost. Always related to the object of the contract, they must be objective, operational and non-discriminatory. The main objective of this procedure is to avoid an aggressive bidding strategy as the public buyer is not obliged to select the lowest bidder. In our data, we do not know how the selection criteria are weighted during the selection process, but we know that those criteria and their respective weights are exactly the same for all contracts (which are all awarded through the same procedure). In theory, competitive mechanisms incite bidders to reveal their private information (Demsetz, 1968). Nevertheless, this statement is true only if, in particular, it is possible to contract on and monitor the quality of the service. Otherwise, the *ex ante* competition does not prevent opportunistic behaviors, and firms can decrease the delivered quality.

---

<sup>4</sup> Although derogation is possible in specific cases, it is still the most used mechanism. According to an EU report, 68% of service contracts awarded between 2006 and 2010 followed an open procedure (see European Commission (2011), p. 12).

## 3.2 Cleaning contracts

In our set of house-cleaning contracts, contract specification entails a precise description of quality standards, in terms of direct service provision (detailed description of the tasks, the resources to be used and the calendar of interventions) as well as in terms of relational aspects (compliance with deadlines when answering to the buyer's requirements and when providing contractual documents, such as service manuals and periodic activity reports). Concerning monitoring, the contract specifies how the quality will be evaluated and by whom. The public buyer we study implemented a detailed evaluation grid to minimize the issue of subjective judgment and to allow comparisons based upon an objective quality-scoring identically constructed across all cleaning contracts. Evaluations are made during 'contradictory controls', *i. e.*, controls in the presence of the representative of the public buyer and the responsible for technical and administrative matters of the cleaning company, organized once per month. The contract holder is notified 48 hours in advance.<sup>5</sup> The completion of the evaluation grid leads to a final mark out of 100. This scoring then allows contractually defined incentive mechanisms. Indeed, if the obtained mark is less than 80, a second contradictory control is planned 48 hours later and penalties are imposed to the contract holder. Those penalties increase if the mark obtained during the second control is still less than 80. The goal of such penalties is to provide coercive means at the disposal of the public buyer to enforce contractual specifications and to punish any breach of contractual obligations. Moreover, the public buyer is contractually authorized to impose penalties when the cleaning companies fail to fulfill their obligations in terms of mandatory documents supplied. Finally, contracts also contain a cancellation clause that can be applied in the event of repeated failures and/or when accumulated penalties exceed a contractually predetermined threshold.

Regarding such contractual arrangements, one could imagine that the public buyer we study is well protected against opportunism. This is not what we observe in the data. As previously said, frequently imposed penalties, user complaints, breaches and early terminations of contracts persist despite the use of open auctions, the definition of quality standards and the monitoring of the contract. To tackle those problems, the public buyer has modified its formal

---

<sup>5</sup> It is important to note that such a delay does not enable the private operator to react by 'falsifying' the quality evaluation through a short-term effort just before the control. Indeed, a lot is composed by between 544 and 3,066 accommodations and controls are run for random buildings. Furthermore, cleanliness is mainly derived from the length of the effort over time and a short, even intense, burst of activity might be not sufficient to obtain a good evaluation.



contracts in three directions. From the specification point of view, the new versions of the contract are more precise regarding the obligations of the operator. For instance, the contract describes more technically the way cleaning must be performed and includes a glossary of cleaning operations. From the monitoring point of view, the new contractual arrangement increases the level of details provided to the cleaner about how the evaluation is made and adds the possibility of unplanned and non-contradictory controls by public agents. Finally, concerning incentives, the new contract adds new categories of penalty clauses and increases their amounts. Aside from this formal transformation, the public buyer has also become more rigorous in the application of penalties. Table 1 highlights all the differences between old (launched before April 2010) and new contracts (launched after April 2010).

**Table 1:** Old versus new contract design.

|   | Panel A – Old contract   | Panel B – New contract   |
|---|--|--|
| Tasks descriptions and contractual requirements | 62 tasks and 3 levels of frequency (daily, weekly, monthly)                                | 118 tasks and 6 levels of frequency (daily, weekly, monthly, quarterly, semiannual, annual)  |
| Evaluation for quality                          |  | Unchanged  |
| Performance obligations                         |  | Unchanged  |
| Mandatory documents                             |  | Unchanged  |
| Controls  | Contradictory controls (once per month /at the discretion of the public buyer)             | Contradictory controls (once per month /at the discretion of the public buyer) + <b>Unplanned and not contradictory controls</b> (at the discretion of the public buyer)   |
| Penalties                                       | 60 euros if quality score < 80 /300 euros if the 2nd following quality score is still < 80 | <b>3 %</b> of the price (fixed part) if quality score < 80 / <b>6 %</b> of the price (fixed part) if the 2nd following quality score is still < 80 + <b>new penalties</b> for late delivery of mandatory documents |
| Award procedure                                 |  | Unchanged  |

### 3.3 Data

The public buyer we study organizes its cleaning activity by establishing a geographic allotment. Indeed, the buildings located in the same area correspond

to a given lot  $j$  whose characteristics remain rather constant over time. This allows us to follow a lot over time and to assess whether different types of contract are associated with different quality levels. Over the study period of 48 months (between July 2008 and June 2012), our unbalanced panel database<sup>6</sup> of 49 lots experienced a total of 102 contracts. More precisely, 45 had been renewed one time,<sup>7</sup> whereas 4 had been renewed two times. We know that these contracts are shared among 13 firms and are managed by 6 different departments on the buyer's side (each department is in charge of a geographical area). We also have information regarding the monthly level of implemented quality, the control frequency, the applied penalties, the tendering phase and the price of the contract. Definitions and summary statistics of our variables are presented in Table 2 for the whole sample and by distinguishing variables before and after the change. We call panel *A* the 50 contracts launched before April 2010 and Panel *B* the 52 launched after this date.

### 3.3.1 Dependent variable

We are mainly interested in explaining the variable  $Quality_{ijt}$ . It measures the level of quality delivered by firm  $i$  on lot  $j$  at time  $t$ . Quality indicators are based on a scale of 0 to 100.<sup>8</sup> The average level of quality is 88.91, which is approximately 9 points beyond the threshold implying penalties. The quality scores of panel *B* are significantly higher by two points than panel *A* (the means are significantly different according to parametric and non-parametric mean-comparisons tests – see Table 2). Figure 1 reports the average value of quality at each period and shows that  $Quality_{ijt}$  significantly increased after April 2010. Since most of the ongoing contracts belong to panel *B* after April 2010, this observation strongly corroborates our intuition that the level of service quality increased after the modifications of the contract design. Moreover, Figure 1 shows that the positive impact of the new contract design did not immediately occur after its introduction. Indeed, firms' performances did not jump just after the introduction of the new design but rather progressively increased. The fact that the effect has taken time to unfold suggests that the new regime requires learning. As we

---

<sup>6</sup> The panel database is unbalanced because some quality indicators are missing. Additional details are given when presenting the data. Additional econometric specifications are also run to deal with the biases this issue might generate.

<sup>7</sup> It means that we observe a lot through two different contracts and sometimes through two different suppliers.

<sup>8</sup> This mark is communicated to the supplier, but is not public information.

Table 2: Variable definitions and summary statistics.

| Variables – definitions              |  | Whole sample |       |      |        |      |       |      |       |       |      | Panel A |      | Panel B |       | Mean test** |            |                |
|--------------------------------------|--|--------------|-------|------|--------|------|-------|------|-------|-------|------|---------|------|---------|-------|-------------|------------|----------------|
|                                      |  | m            | std   | min  | max    | N    | m     | std  | min   | max   | N    | m       | std  | min     | max   | N           | Parametric | Non-parametric |
| <i>Quality<sub>ijt</sub></i>         | Level of quality supplied by firm <i>i</i> on lot <i>j</i> at time <i>t</i> (from 0 to 100)  | 88.91        | 6.86  | 42.5 | 100    | 1382 | 87.75 | 6.64 | 44.50 | 100   | 637  | 89.91   | 6.89 | 42.50   | 100   | 745         | 0.000      | 0.000          |
| <i>Observe<sub>ijt</sub></i>         | Takes a value of 1 if the <i>Quality<sub>ijt</sub></i> was measured and 0 otherwise  | 0.61         | 0.49  | 0    | 1      | 2248 | 0.64  | 0.48 | 0     | 1     | 1000 | 0.60    | 0.49 | 0       | 1     | 1248        | 0.053      | 0.054          |
| <i>NewDesign<sub>ijt</sub></i>       | Takes a value of 1 if the contract for lot <i>j</i> is awarded after April 2010 and 0 otherwise  | 0.56         | 0.50  | 0    | 1      | 2248 | 0     | 0    | 0     | 1     | 1000 | 1       | 0    | 1       | 1     | 1248        | –          | –              |
| <i>Decision<sub>t</sub></i>          | Takes a value of 1 after the decision of the administrative court of Paris, i. e., after May 2009 and 0 otherwise  | 0.78         | 0.42  | 0    | 1      | 2248 | 0.50  | 0.50 | 0     | 1     | 1000 | 1       | 0    | 1       | 1     | 1248        | –          | –              |
| <i>Penalties<sub>ijt-1</sub>*</i>    | Overall value of penalties paid by firm <i>i</i> for lot <i>j</i> at time <i>t-1</i> (in euros)  | 564          | 1.531 | 0    | 13 790 | 2195 | 495   | 1261 | 0     | 8023  | 999  | 620     | 172  | 0       | 13790 | 1196        | 0.057      | 0.085          |
| <i>ControlFreq<sub>ijt-1</sub>*</i>  | Number of times the buyer controlled the quality supplied by firm <i>i</i> on lot <i>j</i> at time <i>t-1</i> / Maximal number of times it could have done it                                      | 0.54         | 0.28  | 0    | 1      | 2195 | 0.64  | 0.27 | 0     | 1     | 999  | 0.45    | 0.27 | 0       | 1     | 1196        | 0.000      | 0.000          |
| <i>Price<sub>ijt</sub></i>           | Winning bid of firm <i>i</i> at time <i>t</i> for the fixed part of lot <i>j</i> / Number of accommodations and by the number of months (in euros). This a unit price per month, per accommodation | 13.48        | 3.57  | 8.94 | 31.91  | 102  | 12.26 | 2.42 | 9.45  | 19.91 | 1000 | 14.41   | 4.05 | 8.94    | 31.91 | 1248        | 0.001      | 0.000          |
| <i>NbOffers<sub>jt</sub></i>         | Number of offers for lot <i>j</i> at time <i>t</i>   | 6.23         | 2.90  | 1    | 14     | 102  | 4.44  | 1.73 | 1     | 8     | 1000 | 8.12    | 2.62 | 2       | 14    | 1248        | 0.000      | 0.000          |
| <i>NbAccommodations<sub>jt</sub></i> | Number of accommodations for lot <i>j</i> at time <i>t</i>   | 1.846        | 594   | 544  | 3 066  | 102  | 1832  | 565  | 544   | 2930  | 1000 | 1858    | 627  | 544     | 3066  | 1248        | 0.916      | 0.714          |

\*Variables built based on information related to the period from January 2008 through June 2008.

\*\*P-value of the difference between means of Panels A and B. The non-parametric independent two-group comparisons correspond to the Mann-Whitney Test.

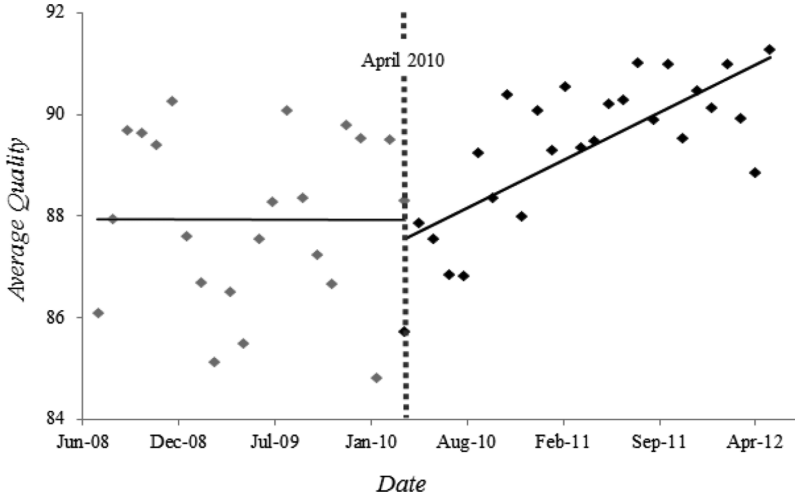


Figure 1: Quality evolution.

will explain in more detail below, the effect of the change introduced by the public buyer can be decomposed into a direct impact due to the new contract design and an indirect impact due to more efficient controls and penalties. Consequently, private firms should commonly realize that the new design is also better enforced by the public buyer to adjust their level of performance, *i. e.* the level of delivered quality.<sup>9</sup>

Two important points have to be made here. First, the quality measure did not change between the old and new contracts. Indeed, there is one single evaluation grid to measure quality, and this evaluation grid is stable over time. Second, this quality score is obtained through *ex post* observations of delivered quality compared to performance obligations. In other words, the quality score does not depend on the completion of the tasks detailed in the contract.<sup>10</sup> Those tasks are described as one way to reach the expected quality, but they are not defined as mandatory and they are not subject to specific controls.

<sup>9</sup> As discussed below, this learning effect also prevents us from using regression discontinuity estimations.

<sup>10</sup> The contract indeed describes the tasks to be performed to obtain a good level of performance. However, the obtained quality score does not depend on whether these tasks have been realized: The obtained quality score depends on the final result, not on the means used to obtain it. This precision is crucial to be sure that the effects associated with our variable *NewDesign* (see section 3.3.2. **Main independent variables**) are indeed due to the rise in contract completeness and not to a change in the way quality is measured. We thank the anonymous referee who raised this crucial point.

### 3.3.2 Main independent variable

We want to assess the impact of contractual completeness on moral hazard. The precise information we have enables us to disentangle a direct impact due to the new contract design from an indirect impact due to more efficient controls and penalties.

#### 3.3.2.1 Contract design

We first constructed the variable  $NewDesign_{jt}$ , which is a binary variable taking a value of 1 if the contract running on lot  $j$  at time  $t$  belongs to Panel  $B$  and 0 otherwise. This variable captures the exogenous shock affecting all the contracts launched from April 2010. As previously said, it resulted in specifying the contracts in greater detail, in reinforcing the penalty clauses and in including a new control procedure. We interpret those changes as an increase of contractual completeness. According to the literature on endogenous verifiability, such a change may increase contract enforceability. We may then expect a positive impact of the variable  $NewDesign_{jt}$  on the level of service.

#### 3.3.2.2 Controls

The variable  $ControlFreq_{ijt-1}$  corresponds to the number of times the quality delivered by firm  $i$  on lot  $j$  at time  $t$  has been controlled, divided by the maximal number of times it could have been controlled. On average, it is equal to 0.54. It is higher in Panel  $A$  than in Panel  $B$ , which reflects that our buyer decreased the controls frequency while reinforcing contractual requirements and imposing higher penalties. One interpretation would be a substitution effect between the larger amounts of penalties and the necessity to carry out assiduous controls. In other words, if the expanded threat of punishment disciplines firms, then regular controls are less useful. We expect that a firm managing a frequently monitored contract (*i. e.*, having a larger  $ControlFreq_{ijt-1}$ ) will feel more intensely under the scrutiny of the buyer, leading us to anticipate a positive impact of the variable  $ControlFreq_{ijt-1}$  on our dependent variable.

#### 3.3.2.3 Incentives

$Penalties_{ijt-1}$  is the second variable that accounts for contract monitoring. It corresponds to the full amount of penalties paid by firm  $i$  on lot  $j$  at time  $t$ . Approximately 731 additional euros per contract have been paid. When

considering the contract size,<sup>11</sup> this amount is rather low and reflects that penalties are a last resort. Nevertheless, a significant difference of 150 euros of penalties distinguishes Panel A from Panel B, which sheds some light on our buyer's decision to reinforce the sanction mechanisms. This evolution appears even more clearly in Table 3, which summarizes the value of penalties paid each year by the cleaning contract suppliers. Since penalties account for small amounts, and since reaching a fixed amount of penalties leads to a breach of the contract, we claim that it is the accumulated value of paid penalties that may have an incentive effect on the suppliers' decision to improve their current performances. We thus expect a positive impact of the variable  $Penalties_{ijt-1}$  on  $Quality_{ijt}$ . The variables  $ControlFreq_{ijt-1}$  and  $Penalties_{ijt-1}$  capture some heterogeneity regarding the contract monitoring. However, when the contract enforceability increases, we expect that monitoring of firms' outcomes will provide stronger incentives to deliver high quality. Therefore,  $ControlFreq_{ijt-1}$  and  $Penalties_{ijt-1}$  may have a larger impact after the change in contract design.

### 3.3.3 Control variables

We use additional controls to capture heterogeneity across contracts. Indeed, some dimensions, such as the degree of competition and the price or scope of the contract, can affect the level of the delivered quality. The variables associated with these dimensions and their expected impacts on quality are described below.

#### 3.3.3.1 Geographical allotment

Our variable  $NbAccommodations_{jt}$  captures the number of accommodations that are included in lot  $j$  at time  $t$ . We indeed aim to exploit the panel nature of our

**Table 3:** Penalties per year.

| Year       | 2008     | 2009     | 2010     | 2011     | 2012**   |
|------------|----------|----------|----------|----------|----------|
| Penalties* | 21 646.5 | 12 347.4 | 37 297.2 | 60 614.5 | 20 800.0 |

\*in Euros; \*\* until June.

<sup>11</sup> The final prices of the contracts we study are made of two parts: A fixed part and a variable part; this last part depends on the buyer's needs, which are unknown at the awarding stage. We only have precise information about the winning bid for the fixed part. It accounts for at least 70% of the final price and equals around 395 642 euros).

data by following the same lot over time. To correctly perform it, the lots have to remain stable.

Nevertheless, we cannot rule out that our buyer will build, buy or sell buildings, affecting the characteristics of the lots. The variable  $NbAccommodations_{jt}$  enables us to account for those types of changes. Since our buyer allots its cleaning activity because larger lots are expected to be more difficult to manage, we suspect that the number of accommodations negatively impacts the level of delivered quality.

### 3.3.3.2 Competition

The variable  $NbOffers_{jt}$  stands for the number of offers received by the buyer for lot  $j$  at time  $t$ . An increase in competition is supposed to be beneficial: it incites firms to reveal their private information and to lower their rents. However, more competition might also encourage aggressive bids at the expense of quality. Therefore, the impact of competition on quality is not consensual.

In addition, the degree of competition may be an endogenous variable, *i. e.*, a variable varying for unobserved reasons that also affect outcome variables (like quality).<sup>12</sup> To clarify the nature of the variable  $NbOffers_{jt}$ , Figure 2 describes the relationship between the number of offers and the year the contract is awarded. All the contracts starting from 2010 include the new design. We can see that the number of offers tended to increase in 2010 and 2011. Although it seems surprising that strengthening the contract design would generate a boom in the number of potential suppliers, practitioners argue that this change sent to firms the signal that the buyer was unsatisfied with its current main suppliers and aimed to find new and better performing firms. This signal may have renewed the set of interested firms, creating a check on the degree of competition. If this increase in competition actually enabled a renewal of the set of suppliers (19% of the contracts launched in 2010 are awarded to new firms) by attracting better performing firms, adverse selection might have been reduced. However, if more competition ended up in encouraging aggressive bids, it may have resulted in damaging quality. Finally, the variable  $NbOffers_{jt}$  is problematic for two reasons: its impact on quality is difficult to anticipate and we do not know whether we should consider it as an independent variable.

### 3.3.3.3 Prices

We finally built the variable  $Price_{ijt}$ , which corresponds to the bid of the winning firm  $i$  for lot  $j$  at time  $t$  divided by the number of accommodations. This variable

---

<sup>12</sup> See, *e. g.* Coviello and Mario Mariniello (2012) or Amaral et al. (2013).

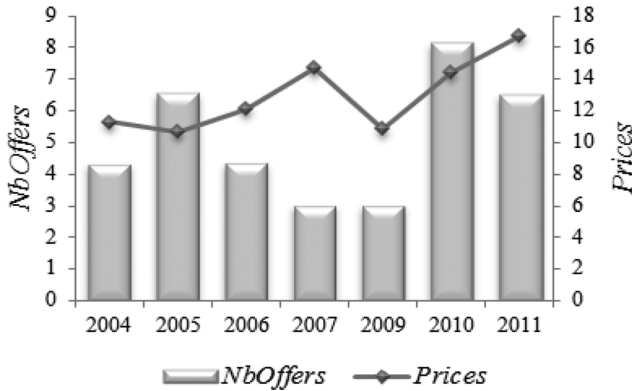


Figure 2: Number of offers and price evolution.

imperfectly captures the competitiveness of the winning offers.<sup>13</sup> Although bid competitiveness might signal firms' efficiency, low prices might also reflect a strategic or a naive behavior of candidates.<sup>14</sup> In addition, on the same grounds as those put forward when presenting the variable  $NbOffers_{jt}$ , the variable  $Price_{ijt}$  is likely to be endogenous. Once again, we try to clarify by analyzing its evolution over time (see Figure 2). We can see that winning bids have tended to increase since 2010. Thus, it seems that the new contract design is associated with less competitive bids: firms will compensate for the costs they incur to fulfill the increasing expectations of the buyer by posting higher prices. Even though the potential trade-off between price and quality is a matter of concern in this paper, considering  $Price_{ijt}$  as an independent variable may be an issue.

### 3.3.3.4 Outcome variables

We have to decide whether to introduce  $NbOffers_{jt}$  and  $Price_{ijt}$  as control variables. On the subject of 'bad controls', Angrist and Pischke (2008: 64, chap. 3) give the following reasoning: "Some variables are bad controls and should not be included in a regression model even when their inclusion might be expected to change the short regression coefficients. Bad controls can be defined as variables that are themselves outcome variables in the notional experiment at hand. Consequently, bad controls might just as well be dependent variables too. On the other hand, good controls are variables that we can think of as having been fixed

<sup>13</sup> As previously mentioned, the final prices of the contract are made of a fixed part and a variable part. The variable  $Price_{ijt}$  is built thanks to information we have about the winning bid for the fixed part, which accounts for at least 70% of the final price.

<sup>14</sup> "Naive behavior" refers to the winner's curse issue (see, e.g. Hong and Shum (2002)).



at the time the regressor of interest was determined". On the one hand, the price and the degree of competition may have been affected by the change in the contract design. They could be considered as outcome variables. On the other hand,  $NbOffers_{jt}$  and  $Price_{ijt}$  are fixed before the quality is delivered. We could thus decide to use them as control variables. However, we can reasonably assume that prices and (at least, part of) the quality are simultaneously determined by the agent at the awarding stage. We do not have this problem with the number of offers: it is fixed before the delivery of quality and it results from rivals' decisions. As a consequence, we have decided to use  $NbOffers_{jt}$  as a control variable and  $Price_{ijt}$  as an outcome variable; we separately explore the determinants of the latter variable in a second step of our analysis.

### 3.4 Model specifications

We are interested in assessing whether better specified contracts make it possible to improve the level of quality. We can investigate this question because the buyer we study built the quality indicator  $Quality_{ijt}$ . Therefore, we first estimate the following model (1):

$$Quality_{ijt} = \beta_1 + \beta_2 \cdot NewDesign_{jt} + \beta_3 \cdot Y_{jt} + W_j + \varepsilon_{ijt} \quad (1)$$

where  $NewDesign_{jt}$  is our first main variable of interest capturing the change in the formal contract, and  $Y_{jt}$  is a vector of variables capturing the characteristics of lot  $j$  at time  $t$ . We abstract unobservable biases due to the nature of the lots by adding lot fixed effects ( $W_j$ ). This first model is a simple "before-after" estimation: it assesses the impact of switching from panel  $A$  to panel  $B$ . Since our main goal is to investigate the impact of new contracts on moral hazard issues, we second run a model with firm fixed effects,<sup>15</sup>  $Z_i$

$$Quality_{ijt} = \beta_1 + \beta_2 \cdot NewDesign_{jt} + \beta_3 \cdot Y_{jt} + W_j + Z_i + \varepsilon_{ijt} \quad (2)$$

Equation (2) enables us to have a more conservative estimation regarding the impact of the new contract design on moral hazard issues. Indeed, if the coefficient associated with the variable  $NewDesign_{jt}$  decreases when switching from eqs (1) to (2), it means that part of the increase in quality comes from the

---

<sup>15</sup> It could also be worthwhile to use robust standard errors clustered at the firm level. Nevertheless, we are limited here by the nature of our data. Even if there is no specific rule, some papers emphasize that the minimum number of clusters for accurate inference is around 50 with roughly equal cluster size (Kézdi (2004), Wooldridge (2001)). Because our dataset contains only 13 firms (see section 3.3 Data), we cannot use clusters at the firm level.

selection of more efficient firms. In eq. (2), the variable  $NewDesign_{jt}$  captures only some changes in the level of quality that are related to moral hazard issues. Then, we can disentangle the impact of the formal contract itself from the frequency of the controls and the penalties by running a third model with the vector  $X_{ijt} = (ControlFreq_{ijt-1}, Penalties_{ijt-1})$ :

$$Quality_{ijt} = \beta_1 + \beta_2.NewDesign_{jt} + \beta_3.Y_{jt} + \beta_4.X_{ijt} - 1 + W_j + Z_i + \varepsilon_{ijt} \quad (3)$$

Finally, we run a fourth model to test whether the formal contract determines the efficiency of the controls and the applied penalties. To do so, we add an interaction term between the variable  $NewDesign_{jt}$  and the demeaned value of the variables related to contract enforcement ( $\hat{X}_{ijt} = X_{ijt} - \bar{X}_{ijt}$ ).<sup>16</sup> We obtain the last equation:

$$Quality_{ijt} = \beta_1 + \beta_2.NewDesign_{jt} + \beta_3.Y_{jt} + \beta_4.X_{ijt} - 1 + \beta_5.(NewDesign_{jt} * \hat{X}_{ijt-1}) + W_j + Z_i + \varepsilon_{ijt} \quad (4)$$

In this model, whereas the vector of coefficients  $\beta_4$  captures the influence of controls and penalties before the changes in contract design,  $\beta_5$  captures their influence after the change. Moreover, demeaning the variables  $X_{ijt}$  in the interaction term makes it possible to assess the marginal impact of penalties and controls after the change in contract design.

All of these specifications have been tested through Ordinary Least Squares (OLS) estimations. However, Figure 1 suggests that the regression discontinuity design (RDD) method might be appropriate. In such a case, the assignment variable would be the date of the contract awarding: contracts awarded before April 2010 are not “treated”, whereas contracts awarded after April 2010 are “treated”. Then, assuming that contracts just below the cutoff (awarded just before April 2010) are very similar to those just above the cutoff (awarded just after April 2010), RDD would allow us to assess the impact of the treatment (being a more or less complete contract) on the outcome (the level of provided quality). However, we do not use RDD for several reasons, first and foremost, because of the aforementioned learning effect. The impact of the new contract design took time to unfold and, consequently, the comparison of contracts signed very shortly after and before the cutoff point would fail to capture the influence of the new regime on the delivered quality. Second, RDD needs to assume that contracts around the cutoff are very similar, whereas we do not

<sup>16</sup> See pages 68-69 of chapter 4 in Wooldridge (2001) for the explanation regarding the construction of the interaction term.

make such an assumption. In contrast, we assume that observable and unobservable characteristics of contracts may differ around the cutoff point. For this reason, we use OLS estimations and we add control variables to isolate the effect of observable characteristics (the number of received offers, etc.) from the variable in which we are interested. Moreover, we use a Heckman model to evaluate the effects of unobservable characteristics (like the caretaker's diligence to carry out a control).

## 4 Results

### 4.1 The determinants of quality

Table 4 exhibits our baseline results for the effect of contractual completeness on quality enforcement. In Model 1, which is a simple before-after, we find that the change in contract design has a significant and positive impact on quality.

**Table 4:** How to implement quality?

|  | <b>Model 1</b><br><i>Quality<sub>ijt</sub></i> | <b>Model 2</b><br><i>Quality<sub>ijt</sub></i> | <b>Model 3</b><br><i>Quality<sub>ijt</sub></i> | <b>Model 4</b><br><i>Quality<sub>ijt</sub></i> |
|--|--|--|--|--|
| <i>NewDesign<sub>ijt</sub></i>                                   | 2.526***<br>(0.605)                            | 2.006***<br>(0.721)                            | 2.104***<br>(0.709)                            | 1.433*<br>(0.738)                              |
| <i>ControlFreq<sub>ijt-1</sub></i>                               |  |  | 1.529 +<br>(0.959)                             | -4.283***<br>(1.494)                           |
| <i>NewDesign<sub>ijt</sub></i> * $\widehat{ControlFreq}_{ijt-1}$ |  |  |  | 8.601***<br>(1.813)                            |
| <i>Penalties<sub>ijt-1</sub></i>                                 |  |  | 0.000<br>(0.000)                               | -0.000 +<br>(0.000)                            |
| <i>NewDesign<sub>ijt</sub></i> * $\widehat{Penalties}_{ijt-1}$   |  |  |  | 0.001***<br>(0.000)                            |
| <i>NbOffers<sub>ijt</sub></i>                                    | 0.063<br>(0.159)                               | 0.018<br>(0.196)                               | 0.088<br>(0.199)                               | 0.062<br>(0.196)                               |
| <i>NbAccommodations<sub>ijt</sub></i>                            | -0.002***<br>(0.001)                           | -0.001<br>(0.001)                              | -0.001<br>(0.001)                              | -0.002 +<br>(0.001)                            |
| <i>Constant<sub>ijt</sub></i>                                    | 91.332***<br>(1.505)                           | 85.040***<br>(2.288)                           | 83.855***<br>(2.474)                           | 88.308***<br>(2.696)                           |
| <i>Lot</i>   | Yes  | Yes  | Yes  | Yes  |
| <i>Firm</i>  | No   | Yes  | Yes  | Yes  |
| <i>N</i>   | 1 359  | 1 359  | 1 359  | 1 359  |
| <i>Adj-R<sup>2</sup></i>   | 0.24   | 0.27   | 0.27   | 0.29   |

Robust standard errors in parentheses; \*\*\*  $p < 0.01$ , \*\*  $p < 0.05$ , \*  $p < 0.1$ , +  $p < 0.15$ .

When switching from Model 1 to Model 2, we add firm fixed effects. The coefficient associated with the variable  $NewDesign_{jt}$  slightly decreases. We interpret this to be a sign that the reduction of the moral hazard issue is the main driver of the quality improvement. We then add the variables  $ControlFreq_{ijt-1}$  and  $Penalties_{ijt-1}$  in Model 3. The coefficient associated with the variable  $NewDesign_{jt}$  remains positive and significant. It corroborates the central idea of the paper: improving the contract details reduces the suppliers' temptation to shirk on quality. In other words, the formal contract itself acts as a discriminating device. However, in Model 3, we do not find that the variables related to *ex post* monitoring influence the level of quality. By distinguishing their respective impacts before and after the change in contract design, model 4 allows us to further investigate this result. In model 4, we indeed see that the formal contract itself has a sizeable influence on the efficiency of the *ex post* monitoring: our specification exhibits a negative effect of  $ControlFreq_{ijt-1}$  and  $Penalties_{ijt-1}$  before the change in contract design, whereas their effect becomes positive and significant after the change in contract design. In addition, although the degree of competition does not influence the delivered quality, we unsurprisingly find that smaller lots are associated with higher performances.

## 4.2 Robustness checks

### 4.2.1 Testing selection bias

Quality control should be done on a monthly basis, but the examination of the data reveals that controls are not always performed. As a consequence, our dataset suffers from missing information corresponding to situations in which controls were not made. Since the decision to make a control is highly decentralized, left to the discretion of the caretaker, whose motives are difficult to make out, the reasons we may invoke to justify this sample selection are multiple and the way it affects our estimates is difficult to anticipate.

The variable  $Observe_{ijt}$  is a dummy variable: it takes a value of 1 if the quality is controlled and 0 otherwise.  $Quality_{ijt}$  was not measured in 39% of all cases. To tackle this issue, we can use a two-step Heckman method (Heckman, 1979). Provided that we can explain why quality is observed or not, we can extrapolate the missing quality indicators as if they had been observed. Thereafter, the model indicates whether the bias due to sample selection is severe, and it accounts for the bias effect on both the dependent and independent variables. The first step of the procedure corresponds to the selection equation, which models the probability of being observed. The second step

corresponds to the corrected outcome equation: it explains the level of quality, given it is observed. However, to be over-identified, the model requires at least one instrument to be included in the first step but not in the second step. This variable must be correlated with the variable  $Observe_{ijt}$  (instrument relevance condition) but not with any unobservable factor that could influence the variable  $Quality_{ijt}$  (exclusion restriction condition).

We suspect  $ControlFreqOthers_{ijt-1}$  is a valid instrument. It measures the rate of control at  $(t-1)$  on a sub-sample of observations.<sup>17</sup> Regarding the relevance condition, we assume that people having both the same job and the same employer may observe and influence each other through a spillover effect: one caretaker's diligence in carrying out a control depends on the observation of the others caretaker's diligence, which is captured through their control rate at  $(t-1)$ . Therefore, so as to respect the exclusion restriction, we build our instrument on a sub-sample of observations: we exclude the observations related to firm  $i$  and those related to the department managing lot  $j$ . Indeed, a firm is likely to observe the past control frequency of its territorial department and/or the past control frequency of its other ongoing contracts: this information may influence its behavior.<sup>18</sup> In contrast, these firms should not be aware, at least in the short term, of the caretakers' propensity to perform controls in the other departments regarding other firms. As a consequence, caretakers' private information should not influence a firm's incentives toward quality.

The results of the two-step Heckman estimates are presented in Table 5. Whatever the specification, we can see that our instrument is significant ( $p < 0.01$ ), satisfying the relevance condition. The variable  $Lambda$  captures the impact of the first stage on the second stage. Given that  $Lambda$  is negative and sometimes significant, we conclude that the probability of being observed is higher when the quality is lower, validating the existence of a sample selection bias. However, our main findings remain extremely stable. There is one notable change in model 8: the coefficient associated with the variable  $NewDesign$  slightly decreases. Since this model does not seem to be significantly affected by the sample selection ( $Lambda$  is

---

<sup>17</sup> To calculate the value of the variable  $ControlFreqOthers_{ijt-1}$ , we look at the total number of controls performed at  $t-1$  (in the other departments and with the other firms) and we divide this number by the total number of controls that could have been performed at  $t-1$  (in the other departments and with the other firms).

<sup>18</sup> For instance, a firm may be more prone to increase the quality of services if it observes that the caretakers tend to increase their controls.

Table 5: Dealing with sample selection.

|   | Model 5<br><i>Quality<sub>ijt</sub></i> | Model 6<br><i>Quality<sub>ijt</sub></i> | Model 7<br><i>Quality<sub>ijt</sub></i> | Model 8<br><i>Quality<sub>ijt</sub></i> |
|---|---|---|---|---|
| <i>NewDesign<sub>jt</sub></i>                               | 2.187***<br>(0.611)                     | 1.655**<br>(0.688)                      | 1.685**<br>(0.713)                      | 1.184 +<br>(0.734)                      |
| <i>ControlFreq<sub>ijt-1</sub></i>                          |   |   | 0.540<br>(1.061)                        | -5.192***<br>(1.643)                    |
| <i>NewDesign<sub>jt</sub> * ControlFreq<sub>ijt-1</sub></i> |   |   |   | 9.238***<br>(1.818)                     |
| <i>Penalties<sub>ijt-1</sub></i>                            |   |   | 0.000<br>(0.000)                        | -0.000<br>(0.000)                       |
| <i>NewDesign<sub>jt</sub> * Penalties<sub>ijt-1</sub></i>   |   |   |   | 0.001***<br>(0.000)                     |
| <i>NbOffers<sub>jt</sub></i>                                | 0.128<br>(0.133)                        | 0.074<br>(0.161)                        | 0.116<br>(0.164)                        | 0.076<br>(0.163)                        |
| <i>NbAccommodations<sub>jt</sub></i>                        | -0.002***<br>(0.001)                    | -0.001<br>(0.001)                       | -0.001<br>(0.001)                       | -0.002 +<br>(0.001)                     |
| <i>Constant<sub>ijt</sub></i>                               | 88.576***<br>(1.387)                    | 87.731***<br>(2.333)                    | 86.836***<br>(2.603)                    | 91.324***<br>(2.802)                    |
|   | <i>Observe<sub>ijt</sub></i>            | <i>Observe<sub>ijt</sub></i>            | <i>Observe<sub>ijt</sub></i>            | <i>Observe<sub>ijt</sub></i>            |
| <i>ControlOthers<sub>ijt-1</sub></i>                        | 2.616***<br>(0.162)                     | 2.681***<br>(0.165)                     | 2.481***<br>(0.174)                     | 2.768***<br>(0.180)                     |
| <i>NewDesign<sub>jt</sub></i>                               | 0.374***<br>(0.110)                     | 0.460***<br>(0.124)                     | 0.526***<br>(0.127)                     | 0.651***<br>(0.130)                     |
| <i>ControlFreq<sub>ijt-1</sub></i>                          |   |   | 0.656***<br>(0.170)                     | 2.303***<br>(0.277)                     |
| <i>NewDesign<sub>jt</sub> * ControlFreq<sub>ijt-1</sub></i> |   |   |   | -2.550***<br>(0.335)                    |
| <i>Penalties<sub>ijt-1</sub></i>                            |   |   | -0.000<br>(0.000)                       | -0.000<br>(0.000)                       |
| <i>NewDesign<sub>jt</sub> * Penalties<sub>ijt-1</sub></i>   |   |   |   | -0.000<br>(0.000)                       |
| <i>NbOffers<sub>jt</sub></i>                                | -0.088***<br>(0.023)                    | -0.093***<br>(0.029)                    | -0.085***<br>(0.030)                    | -0.107***<br>(0.031)                    |
| <i>NbAccommodations<sub>jt</sub></i>                        | 0.000<br>(0.000)                        | 0.000<br>(0.000)                        | 0.000<br>(0.000)                        | 0.000<br>(0.000)                        |
| <i>Constant<sub>ijt</sub></i>                               | -0.439<br>(0.312)                       | -0.506<br>(0.488)                       | -0.895*<br>(0.501)                      | -2.049***<br>(0.537)                    |
| <i>Lambda</i>   | -1.923**<br>(0.775)                     | -1.873*<br>(0.746)                      | -1.726**<br>(0.815)                     | -0.861<br>(0.753)                       |
| <i>Lot</i>  | yes                                     | yes                                     | yes                                     | yes                                     |
| <i>Firm</i>   | no                                      | yes                                     | yes                                     | yes                                     |
| <i>N</i>  | 1 359                                   | 1 359                                   | 1 359                                   | 1 359                                   |

Robust standard errors in parentheses; \*\*\* p < 0.01, \*\* p < 0.05, \* p < 0.1, + p < 0.15.

no more significant), model 4 might be more relevant than model 8 for estimating eq. (4).

#### 4.2.2 Testing shock exogeneity

As previously mentioned, the modification of the contract design by the public buyer is not an *ex nihilo* decision. In contrast, it is a reaction to a judicial decision of the administrative court of Paris in June 2009 (see Figure 3 that reports the timing of the events).

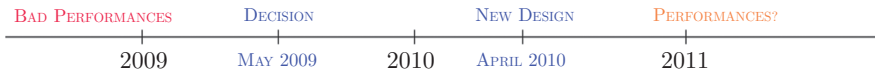


Figure 3: Timing of the events.

If this court's decision ever impacted the buyer's and/or firms' behaviors, we may have concerns about not having a quasi-natural experiment. To check that, we replicate our estimates by including the variable  $Decision_t$ , which is a dummy variable taking a value of 1 after May 2009. The results are presented in Table 6. Whatever the specification, we can see that the decision has no impact on the buyer's propensity to observe the quality or on the level of delivered quality: parties did not adapt their behavior to this decision. Moreover, our results remain perfectly stable.

### 4.3 Prices

To complement our analysis, we assess whether the new contract design has an impact on prices, *i. e.*, whether the public buyer has to arbitrate between price and quality. Therefore, we estimate the following equation:

$$Price_{ijt} = \alpha_1 + \alpha_2 \cdot NewDesign_{jt} + \alpha_3 \cdot PriceIndex_t + \alpha_4 \cdot Y_{jt} + W_j + Z_i + \zeta_{ijt} \quad (5)$$

This estimation allows us to assess the impact of the change in contract design, given the nature of the lot (we add the lot fixed effects,  $W_j$ ) and the degree of competition. We also aim at disentangling the impact of the identity of supplier  $i$  from the impact of the change in contract design. However, since we run our estimations on a (rather small) sample of 102 contracts, we cannot simultaneously add lot and firm fixed effects. We address this issue by testing various specifications: each one includes a

Table 6: Testing for the exogeneity of the change in contract design.

|   | Model 9<br><i>Quality<sub>ijt</sub></i> | Model 10<br><i>Quality<sub>ijt</sub></i> | Model 11<br><i>Quality<sub>ijt</sub></i> | Model 12<br><i>Quality<sub>ijt</sub></i> |
|---|---|--|--|--|
| <i>Decision<sub>t</sub></i>                                 | 0.310<br>(0.499)                        | 0.297<br>(0.497)                         | 0.233<br>(0.500)                         | 0.591<br>(0.499)                         |
| <i>NewDesign<sub>jt</sub></i>                               | 0.371***<br>(0.120)                     | 1.497**<br>(0.738)                       | 1.562**<br>(0.759)                       | 0.877<br>(0.777)                         |
| <i>ControlFreq<sub>ijt-1</sub></i>                          |   |  | 0.529<br>(1.062)                         | -5.319***<br>(1.645)                     |
| <i>NewDesign<sub>jt</sub> * ControlFreq<sub>ijt-1</sub></i> |   |  |  | 9.341***<br>(1.820)                      |
| <i>Penalties<sub>ijt-1</sub></i>                            |   |  | 0.000<br>(0.000)                         | -0.000 +<br>(0.000)                      |
| <i>NewDesign<sub>jt</sub> * Penalties<sub>ijt-1</sub></i>   |   |  |  | 0.001***<br>(0.000)                      |
| <i>NbOffers<sub>jt</sub></i>                                | -0.088***<br>(0.023)                    | 0.075<br>(0.161)                         | 0.115<br>(0.164)                         | 0.070<br>(0.163)                         |
| <i>NbAccommodations<sub>jt</sub></i>                        | 0.000<br>(0.000)                        | -0.001<br>(0.001)                        | -0.001<br>(0.001)                        | -0.002 +<br>(0.001)                      |
| <i>Constant<sub>ijt</sub></i>                               | 88.272***<br>(1.471)                    | 87.457***<br>(2.376)                     | 86.655***<br>(2.633)                     | 90.936***<br>(2.823)                     |
|   | <i>Observe<sub>ijt</sub></i>            | <i>Observe<sub>ijt</sub></i>             | <i>Observe<sub>ijt</sub></i>             | <i>Observe<sub>ijt</sub></i>             |
| <i>Decision<sub>t</sub></i>                                 | 0.007<br>(0.091)                        | -0.011<br>(0.091)                        | 0.006<br>(0.093)                         | 0.019<br>(0.097)                         |
| <i>ControlOthers<sub>ijt-1</sub></i>                        | 2.616***<br>(0.162)                     | 2.681***<br>(0.165)                      | 2.481***<br>(0.174)                      | 2.769***<br>(0.180)                      |
| <i>NewDesign<sub>jt</sub></i>                               | 0.371***<br>(0.120)                     | 0.466***<br>(0.134)                      | 0.523***<br>(0.136)                      | 0.641***<br>(0.139)                      |
| <i>ControlFreq<sub>ijt-1</sub></i>                          |   |  | 0.656***<br>(0.170)                      | 2.305***<br>(0.277)                      |
| <i>NewDesign<sub>jt</sub> * ControlFreq<sub>ijt-1</sub></i> |   |  |  | -2.552***<br>(0.335)                     |
| <i>Penalties<sub>ijt-1</sub></i>                            |   |  | -0.000<br>(0.000)                        | -0.000<br>(0.000)                        |
| <i>NewDesign<sub>jt</sub> * Penalties<sub>ijt-1</sub></i>   |   |  |  | -0.000<br>(0.000)                        |
| <i>NbOffers<sub>jt</sub></i>                                | -0.088***<br>(0.023)                    | -0.093***<br>(0.029)                     | -0.085***<br>(0.030)                     | -0.107***<br>(0.031)                     |
| <i>NbAccommodations<sub>jt</sub></i>                        | 0.000<br>(0.000)                        | 0.000<br>(0.000)                         | 0.000<br>(0.000)                         | 0.000<br>(0.000)                         |
| <i>Constant<sub>ijt</sub></i>                               | -0.446<br>(0.326)                       | -0.494<br>(0.497)                        | -0.901*<br>(0.510)                       | -2.067***<br>(0.545)                     |
| <i>Lambda</i>   | -1.928**                                | -1.874**                                 | -1.741**                                 | -0.906                                   |

(continued)



Table 6: (continued)

|             | Model 9<br><i>Quality<sub>ijt</sub></i> | Model 10<br><i>Quality<sub>ijt</sub></i> | Model 11<br><i>Quality<sub>ijt</sub></i> | Model 12<br><i>Quality<sub>ijt</sub></i> |
|-------------|---|--|--|--|
|             | (0.775)                                 | (0.746)                                  | (0.816)                                  | (0.753)                                  |
| <i>Lot</i>  | yes                                     | Yes                                      | yes                                      | yes                                      |
| <i>Firm</i> | no                                      | Yes                                      | yes                                      | yes                                      |
| <i>N</i>    | 1 359                                   | 1 359                                    | 1 359                                    | 1 359                                    |

Robust standard errors in parentheses; \*\*\*  $p < 0.01$ , \*\*  $p < 0.05$ , \*  $p < 0.1$ , +  $p < 0.15$ .

specific category of fixed effects. Since the contracts we study were awarded between 2004 and 2011, we are likely to observe some inflation. To address this, we also include a price index of cleaning services,  $PriceIndex_t$ . The results of the estimates are presented in Table 7.

The number of offers has the expected impact and is in line with previous results from the literature suggesting that more competition attracts lower bids.

Table 7: Price (bids) and change in contract design.

|                                      | Model 13<br><i>Price<sub>ijt</sub></i> | Model 14<br><i>Price<sub>ijt</sub></i> | Model 15<br><i>Price<sub>ijt</sub></i> |
|--------------------------------------|--|--|--|
| <i>NewDesign<sub>jt</sub></i>        | 2.203<br>(2.898)                       | 5.579<br>(3.898)                       | 0.253<br>(3.459)                       |
| <i>NbOffers<sub>jt</sub></i>         | -0.706***<br>(0.181)                   | -0.469**<br>(0.198)                    | -0.733***<br>(0.190)                   |
| <i>NbAccommodations<sub>jt</sub></i> | -0.001<br>(0.001)                      | -0.001<br>(0.002)                      | -0.001<br>(0.001)                      |
| <i>PriceIndex<sub>t</sub></i>        | 0.155<br>(0.128)                       | -0.098<br>(0.169)                      | 0.226<br>(0.178)                       |
| <i>Constant<sub>ijt</sub></i>        | 1.196<br>(13.587)                      | 26.947<br>(20.161)                     | -5.142<br>(18.138)                     |
| <i>Lot</i>                           | No                                     | Yes                                    | No                                     |
| <i>Firm</i>                          | No                                     | No                                     | Yes                                    |
| <i>N</i>                             | 102                                    | 102                                    | 102                                    |
| <i>Adj- R<sup>2</sup></i>            | 0.34                                   | 0.49                                   | 0.38                                   |

Moreover, we find that the new design has no significant impact on the received prices, whatever the specification we consider. This highlights the previous existence of opportunistic behaviors from cleaning operators that actually have sufficient leeway to increase quality without raising bids.

## 5 Conclusions

The results we obtain in this study show that reducing contractual incompleteness enables to significantly reduce moral hazard issues through direct and indirect effects. Indeed, while more complete contracts are associated with lower moral hazard, they also enable to increase the incentive power of controls and penalty processes. In addition, we find that this improvement does not necessarily result in a significant increase in prices. It validates the idea that asymmetries of information result in opportunistic behaviors that can be diminished by providing good incentives. The solution we provide is appropriate for standard transactions. However, for single-use contracts, bearing the costs of reducing contractual incompleteness might be irrelevant. Consequently, this paper also has an important message concerning the way that outsourcing of public services is organized in the European Union. As illustrated by the previously mentioned decision of the administrative court of Paris, European rules on public procurement do not allow for taking past experiences and reputation into account. Whereas this obligation increases transparency and thus, limits abuses in discretion with public funds, it still appears to be insufficient to systematically obtain the best value for money. The drawbacks come from the fact that those rules put the emphasis only on the awarding process, which ensures, under rarely obtained conditions, an efficient contract execution. In the end, when awarding custom-made contracts, public managers still have to find a way to address the issue of contractual incompleteness and contract enforcement.

## References

- Amaral, M., S. Saussier, and A. Yvrande-Billon. 2013. "Expected Number of Bidders and Winning Bids: Evidence from the London Bus Tendering Model," *47 Journal of Transport, Economics and Policy* 17–34.
- Angrist, J.D., and J.-S. Pischke. 2008. *Mostly Harmless Econometrics: An Empiricist's Companion*. Princeton: Princeton University Press.
- Ascher, K.. 1987. *The Politics of Privatization: Contracting Out Public Services*. London: Macmillan Education.
- Australian Industry Commission. 1996. *Competitive Tendering and Contracting by Public Sector Agencies*. Melbourne: Australian Government Publishing Service.
- Bajari, P., and S. Tadelis. 2001. "Incentives versus Transaction Costs: A Theory of Procurement Contracts," *32 RAND Journal of Economics* 387–407.
- Blom-Hansen, J.. 2003. "Is Private Delivery of Public Services Really Cheaper? Evidence from Public Road Maintenance in Denmark," *115 Public Choice* 419–438.
- Brown, T.L., and M. Potosky. 2005. "Transaction Costs and Contracting: The Practitioner Perspective," *28 Public Performance and Management Review* 326–351.

- Cabral, S., and S. Saussier. 2013. "Organizing Prisons Through Public-Private Partnerships: A Cross-Country Investigation," 10 *Brazilian Administration Review* 100–120.
- Coviello, D., and M. Mario Mariniello. 2012. "Publicity Requirements in Public Procurement: Evidence from a Regression Discontinuity Design," *Working Paper*.
- Coviello, D., L. Moretti, G. Spagnolo, and P. Valbonesi. 2013. "Court Efficiency and Procurement Performance," *Working Paper*.
- Demsetz, H.. 1968. "Why Regulate Utilities?," 11 *Journal of Law and Economics* 55–66.
- Dijkgraaf, E., and R. Gradus. 2004. "Cost Savings of Contracting Out Refuse Collection," Technical Report, EconWPA.
- Domberger, S., C. Hall, and M. Jeffries. 1995. *Competitive Tendering and Contracting in Commonwealth Government Agencies: The 1995 Survey Findings*. Australia: CTC Research Team, Graduate School of Business, University of Sydney.
- Domberger, S., C. Hall, and V. Skinner. 2004. *Contracting of Services in the Western Australian Public Sector: The 1995 Survey Findings*. Australia: CTC Research Team, Graduate School of Business, University of Sydney.
- Domberger, S., and P. Jensen. 1997. "Gaining from Outsourcing," 4 *New Economy* 159–163.
- Domberger, S., P. Jensen, and R. Stonecash. 2002. "Examining the Magnitude and Sources of Cost Savings Associated with Outsourcing," 26 *Public Performance and Management Review* 148–168.
- Domberger, S., S. Meadowcroft, and D.J. Thompson. 1986. "Competitive Tendering and Efficiency: The Case of Refuse Collection," 7 *Fiscal Studies* 69–87.
- Dye, R.A.. 1985. "Costly Contract Contingencies," 26 *International Economic Review* 233–250.
- European Commission. 2004. Directive 2004/18/EC of the European Parliament and of the Council.
- European Commission. 2011. *Impact and Effectiveness of EU Public Procurement legislation*, Evaluation report, part 1.
- Evatt Research Centre. 1990. *Breach of Contract: Privatization and the Management of Australian Local Government*. Sydney: Pluto Press.
- Girth, A.M.. 2012. "A Closer Look at Contract Accountability: Exploring the Determinants of Sanctions for Unsatisfactory Contract Performance," 24 *Journal of Public Administration Research and Theory* 317–348.
- Grossman, S.J., and O. Hart. 1986. "The Cost and Benefits of Ownership: A Theory of Vertical and Lateral Integration," 94 *Journal of Political Economy* 691–719.
- Hartley, K., and M. Huby. 1986. "Contracting Out Policy: Theory and Evidence," in J. Kay, C. Mayer and D. Thompson, eds. *Privatization and Regulation: The UK Experience*. UK: Oxford University Press.
- Heckman, J.J.. 1979. "Sample Selection Bias as a Specification Error," 47 *Econometrica* 153–162.
- Holmström, B.. 1979. "Moral Hazard and Observability," 10 *The Bell Journal of Economics* 74–91.
- Hong, H., and M. Shum. 2002. "Increasing Competition and the Winner's Curse: Evidence from Procurement," 69 *Review of Economic Studies* 871–898.
- Jensen, M., and W. Meckling. 1976. "Theory of the Firm: Managerial Behavior, Agency Costs and Ownership Structure," 3 *Journal of Financial Economics* 305–360.
- Jensen, P.H., and R.E. Stonecash. 2005. "Incentives and the Efficiency of Public Sector Outsourcing Contracts," 19 *Journal of Economic Surveys* 767–787.
- Kézdi, G. 2004. "Robust Standard Error Estimation in Fixed-Effects Panel Models," *Hungarian Statistical Review*, Special number 9.
- Kvaløy, O., and T. Olsen. 2009. "Endogenous Verifiability and Relational Contracting," 99 *American Economic Review* 2193–2208.

- Kvaløy, O., and T.E. Olsen. 2016. "Incentive Provision When Contracting Is Costly," 83 *Economica* 741–767.
- Milne, R.G., and R.E. Wright. 2004. "Competition and Costs: Evidence from Competitive Tendering in the Scottish National Health Service," 51 *Scottish Journal of Political Economy* 1–23.
- Savas, E.S.. 1977. "An Empirical Study of Competition in Municipal Service Delivery," 37 *Public Administration Review* 717–724.
- Spagnolo, G.. 2012. "Reputation, Competition, and Entry in Procurement," 30 *International Journal of Industrial Organization* 291–296.
- Townsend, R.M.. 1979. "Optimal Contracts and Competitive Markets with Costly State Verification," 21 *Journal of Economic Theory* 265–293.
- Wooldridge, J.M.. 2001. *Econometric Analysis of Cross Section and Panel Data, Volume 1*. MA: The MIT Press.