Competition, Contract Design and Quality of Outsourced Services

Jean Beuve
University of Paris 1 Pantheon Sorbonne - Chaire EPPP
jean.beuve@univ-paris1.fr

Lisa Chever
Sorbonne Business School - Chaire EPPP
lisa.chever@gmail.com

Very First Draft

Abstract

This paper estimates the impact of contract design and monitoring on the delivered quality of services provided through outsourcing. If the literature on competitive tendering had shown that contracting out allows to reduce costs, there exist up to now few evidences concerning the fact that costs savings could be achieved through lower quality of the service (quality shading hypothesis). Mixed results obtained by the previous literature on this issue seems to suggest that adverse effects on quality are the consequence of poor application of outsourcing process rather than outsourcing per se. Unlike previous studies, which have relied on case studies or cross-sectional data and on subjective measures of service quality, a four-year panel data of 102 contracts is used to estimate a series of fixed-effects regression models. These panel estimates suggest that quality can effectively be improved by implementing better contract specification and management.

JEL Codes: D82, L15, L24

Keywords: Outsourcing, adverse selection, moral hazard, quality, contract design, contract monitoring.

Introduction

In 2009, a public authority in charge of public procurement contracts in the field of social housing was sanctioned by the administrative court of Paris for disqualifying a candidate because of a bad past experience with him. Arguing that this firm was guilty for providing low level of service quality in past cleaning contracts, the public authority decided to reject its candidacy at the pre-qualification stage of the competitive tendering process. Nevertheless, the administrative court, seized by the eliminated firm, has considered that the argument used to disqualify the claimant was unlawful and condemned the public authority to re-organize the call for tenders and to evaluate the candidacy of each operator, including the claimant. The lessons to be drawn from this example are twofold. First, it illustrates the existence of potential inefficiencies associated with legal process organizing competitive auctions which prevent public authority to use some kind of information like past experiences and reputation even in the case of bad performances (Calzolari and Spagnolo (2009)). Second, and most importantly for what we are interested in this paper, it highlights the difficulties to maintain quality when public services are outsourced to an external firm, even for public services that are generally considered as simple such as cleaning services.

From a theoretical point of view, the contractor's temptation to provide lower efforts than those expected by public authorities has been widely studied by the agency theory. As it is put forward by this framework, asymmetric information between public and private parties is a major issue in the contracting out process. Indeed when a public authority delegates a task or a service to a private operator, the latter can use his information advantage in order to maximize his own interest through opportunistic behavior at the expense of the former (Jensen and Meckling (1976)). The public authority is thus exposed to classical risks of adverse selection and moral hazard that should lead the private operator to raise costs and/or to provide low level of service quality. In order to safeguard his interest, the public authority can rely on behavior-based contract by investing in monitoring system, thereby constraining the contractor's opportunity to shirk (Fama and Jensen (1983)) and on outcome-based contract by structuring agent incentives such that the two parties' interests are aligned (Jensen and Meckling (1976)). Nevertheless, the efficient use of such contracts clearly depends on, at least, two conditions. On the one hand, the monitoring ability of the public authority and the costs associated with such monitoring are essentials to behavior-based contracts while, on the other hand, the clarity of outcome measures is absolutely essential to have recourse to outcome-based contract (Eisenhardt 1989). Difficulties associated with the specification of parties' behaviors and outcome measures often lead to consider that some activities are good candidates to contracting out while there are no interests to outsource others. More precisely, several contributions show that contracting out should only be used when transactions are simple, *i.e.* when monitoring ability, task observability and clarity of outcome measures are high. Those particular transactions

are supposed to give raise to complete contracts, which are efficiently tendered through competitive mechanisms (Bulow and Klemperer (1996)). Among those particular transactions considered as "easy to manage", cleaning and waste collection services are the most frequently put forward. As a consequence, awarding contracts for these two types of activities through competition should systematically lead to efficiency gains. This statement seems corroborated by numerous quantitative studies that have shown that outsourcing actually allows cost savings in cleaning and refuse collection sectors (Domberger, Jensen, and Stonecash (2002); Milne and Wright (2004)). However, there remain some doubts about the way these savings are realized. Some studies show they might be achieved at the expense of quality, meaning that adverse selection and/or moral hazard issues may persist. A decrease in price and quality might indeed reflect the strategy of a firm which deliberately puts an aggressive bid to maximize its chance winning the contract (adverse selection) and/or which exploits the insufficiency of incentives mechanisms to lower quality (moral hazard).

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*. Hence, whatever the transaction at stake, the way outsourcing is realized matters. It is indeed rather straightforward to understand that no transaction is *per se* a good candidate to outsourcing: the perfect contractibility is based on the observability of a task or a service; dimension that can be considered as, at least partially, endogenous. First, the observability depends on the precision of the contract specification: do the parties agree about the dimensions they are supposed to observe? Second, it also depends on the monitoring of the contract: does the principal follow the execution of the contract on a regular basis? Does it enable to extract information to make clear potential gaps between excepted and observed performances? Finally, a good candidate to outsourcing depends on the transaction but also on the costs incurred by the principal to make the transaction observable. In cases in which those costs are not incurred, competitive mechanisms might fail in solving asymmetries of information and lead to low quality of the delivered service. Indeed, an imprecise contract specification allows firms to put aggressive bids (adverse selection) while an insufficient contract monitoring does not provide incentives for firms to increase their efforts (moral hazard).

The contribution of this paper is precisely to investigate how the efficiency of the outsourcing process regarding the quality of the service can be improved by a better contract specification and monitoring. Our view is that better contract specification allows parties to agree on what they are supposed to observe, whereas better contract monitoring enable the public authority to give incentives to firms and to make clear any gap between what they should observe and what they actually observe. The best way to verify such propositions would be to test whether contractors' performances depend on a change in contract design and monitoring, which is made difficult by the availability of appropriate

data. Indeed, a main issue to properly investigate and disentangle the effect of competition, contract design and *ex post* monitoring on quality of services is to benefit from longitudinal data set which allows "before and after" comparisons. In practice, "before and after" comparisons of quality can very scarcely be made because of the lack of data on service quality prior to contracts being let (Domberger and Jensen (1997)).

In this paper, the data we collected enable us to circumvent that difficulty. We have access to all the contracts signed between a public authority and a set of private operators in the sector of house-cleaning over a four years period (from July 2008 to June 2012), i.e. 102 contracts. Two important features of our data needs to be pointed out.

First, around a decade ago, the public authority we study decided to implement a tool so as to systematically assess the level of service quality. This tool take the form of an evaluation grid, based on multiple criteria, which should be monthly completed for each contract. As we will see hereafter, whenever the scoring falls below a defined and contractualized threshold, penalties have to be paid by the private firm. The aim of such a tool was to obtain an objective scoring that allows comparisons between contracts. Indeed, as underlined by Jensen and Stonecash (2005), even if service quality may be identified in terms of performance characteristics, their assessment may require subjective judgment rather than mere accumulation of data. The authors use the example of cleaning services, recalling that the only way this 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)). Through the construction of a detailed evaluation grid, the public authority minimizes this issue of subjective judgment and the quality-scoring is thus identically constructed for all the cleaning contracts we study.

Second, in April 2010 and in response to the decision of the administrative court previously mentioned, the public authority decided to implement two major changes: all the contracts launched after this date include (1) more detailed specifications (in particular, about the way to reach a satisfying quality of service) and (2) a reinforced control and penalties regime. Hence, we observe an exogenous shock both on the contract design and on the contract monitoring.

Hence, we have a panel data which allows us to observe the evolution of quality over time (measured through an objective and stable process) before and after an exogenous shock, leaving us with a quasi-natural experiment. Moreover, since we know the identity of the supplier, the value of the penalties paid and the frequency of controls, our data enable to disentangle the impact of potential adverse selection effects from the impact of the contract design and monitoring.

Our results suggest that the exogenous shock resulted in a significant improvement of the delivered quality. Part of this beneficial effect come from the correction of adverse selection: after April 2010, the new selected suppliers outperform oldest ones in terms of quality-score. Thereafter, the quality improvement which is not due to a reduction of

adverse selection issues result both from the change in contract design *per se* and from the monitoring. Moreover, we find this monitoring is made beneficial only if it supported through a well-specified contract. Robustness checks also allow us to show that those quality improvements were made possible without costs modifications, meaning that private firms had leeway to improve the quality of cleaning services.

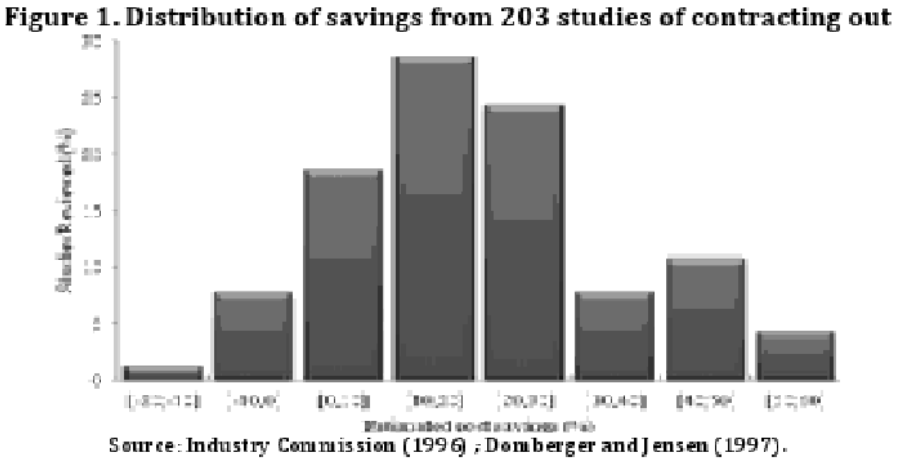
The rest of the paper stands as follow. Section 2 presents the empirical and theoretical related literature and puts forward the propositions we aim at test. Section 3 then gives some details about the institutional and the contractual framework we focus on. The data and the variables used in our empirical strategy are described in Section 4. The results from our estimations are provided in Section 5. Finally, Section 6 discusses those results and potential recommendations for outsourcing of public services.

1.Related literature

1.1.Efficient outsourcing

Because they deal with services of general interest and public funds, the way outsourcing influence the costs of public services is in the focus of academic and practitioner interests. Over the last decades, outsourcing activities to external providers has become a fairly common practice for government. Public procurement accounts for an estimated 15 percent of the world GDP (Lewis and Bajari (2011)). The established purpose of outsourcing decision is to reach better efficiency through costs savings. Substitution of capital for labor, more efficient work practices, more work intensity, economies of scale and greater innovation spurred by competition figure among the main reasons put forward to explain better efficiency. Up to now, there are numerous quantitative studies that have examined whether outsourcing allow to achieve cost reductions. Most of these studies rely on cross-sectional approach in order to compare cots of public management on the one hand and costs of private operator when the contracts are awarded through a competitive tendering on the other hand. In their overwhelming majority, those studies conclude that outsourcing achieves reduction in government expenditures. For instance, in the refuse collection industry, Domberger, Meadowcroft, and Thompson (1986) demonstrate that local authorities that had decided to outsource refuse collection benefit from cost saving of 22% on average. Still in the refuse collection industry, similar results are also reached by other and most recent studies. Szymanski (1993) and Dijkgraaf and Gradus (2004) respectively estimate the average cost reduction obtained through outsourcing at 20% and 15-20%. Identical evidences can also be found in sectors like road maintenance (Blom-Hansen (2003)), vehicle and warehousing maintenance (Domberger, Jensen, and Stonecash (2002)), cleaning and housekeeping (Domberger et al. 1995, Milne and Wright 2002) and even prisons (Cabral

Figure 1: Distribution of cost savings



et al. 2010). Overall, a meta-analysis conducted by the Australian Industry Commission (1996) on 203 different international studies on the effect of contracting on cost reveal that the most frequently reported magnitudes of cost savings lies in the ranges of 10-20 and 20-30 percent (see Figure 1).

Nevertheless, studies which states cost savings associated with contracting out has been subject to criticism. The first one concerns the issue of hold-up and the persistence of savings. Indeed, most of the previously mentioned studies are based on cross-sectional analysis while contractual relationships are, by definition, dynamic. As underlined by Jensen and Stonecash (2005), simple snapshot comparisons of expenditure are not sufficient to conclude whether outsourcing results in long-term improvements, notably because they cannot account for changes in price over time. This issue is all the more problematic than contracts are incomplete and prone to be renegotiate through time. Renegotiation associated with the occurrence of unforeseen events can be the occasion of opportunistic behavior, i.e. the private operator can use his bargaining power[1] to increase prices over time and, consequently, to erode the savings from outsourcing (hold-up). However, most recent

[1]The public authority always has the option of going back to the market or to public provision if it believes that its private partner is behaving opportunistically. However, the switching costs it entail might lead the public authority more willing to incur small losses associated with contract renegotiation rather than go back to the market of the in-house provision.

studies circumvent that criticism by using panel data analysis. Results indicate that unless its total magnitude appears lower than in previous studies, cost savings from outsourcing persist when panel approach is used (Domberger, Jensen, and Stonecash (2002); Milne and Wright (2004)). The second most common criticism addressed to those studies is not to take quality into account. We discuss this issue in which we are focusing on in this paper in the two following sub-sections.

2.2.The quality-shading hypothesis

The argument that cost savings are achieved, at least partially, through a decline in service quality (quality-shading hypothesis) is frequently put forward. This argument appears relevant when we also remind that contracts are (most of the time) incomplete and subject to moral hazard issues. Nowadays, although the effect of outsourcing on quality is of fundamental importance for the efficient organization of public services contracting out, the issue has attracted few empirical studies. The first and essential reason is because the management of service quality is more difficult than controlling the quality of manufactured products. Indeed services are intangible, i.e. they go out of existence at the moment they are created (Marshall 1947), and consequently cannot neither easily be sampled for testing nor measured along standard dimensions.

Thus it is difficult to subject non-contractible and subjectively perceived quality. For such a simple reason, empirical studies which examine this quality issue exist in far fewer numbers than those on costs savings. In addition to being less numerous, those empirical studies on quality provide mixed evidences on the effect of contracting on the quality of service provision. Some studies suggest that service quality may have improved with contracting out. For instance, Domberger et al. (1995), in their analysis of cleaning contracts in the Sydney area, used regression analysis to assess the effect of contracting on performance and find no significant effect of contracting on quality in the majority of the cases and positive effect in the best case (they estimate a 35 per cent improvement of performance in the case of special schools). Similar results of no change or improvement of quality are also put forward by Domberger (1993) and Farago, Hall, and S. (1994) in their respective studies using data on various services from New South Wales public sector and Savas (1997) in his study of waste collection in the City of Minneapolis (US). On the other hand, other studies reach the opposite conclusion of reduced quality following contracting out. For instance, the Evatt Research Center (1990) reported that many respondents to its survey of local councils in Australia were of the opinion that outsourcing resulted in deterioration in the quality of service. By using data from local governments and health authority in the United Kingdom, Hartley and Huby (1986) and Ascher (1987) also find that contracting out, whereas it allows to save costs, is associated with reduced quality of the procured service. With the same kind of data from local authorities and for various

services in the UK,McMaster (1995) reports that of 199 contracts examined, the standard of services delivered was considered by senior management of the tendering authorities to be unchanged in 113, to have deteriorated in 47 and to have improved in 39. In the end, there is no consensus on the impact of outsourcing on delivered quality. Moreover, all the previously mentioned studies are subject to criticisms for at least two methodological reasons. First, the research does not make a distinction between reductions in quality deriving from inappropriate specifications and quality problems arising from the failure to meet the standard specified. Second, all of those subjective measures of quality (such as the opinions of those who filled in questionnaires) rather than attempting to look at objective measures, such as contract failures and penalties. Aside those methodological issues, the vast majority of those studies share a same interpretation of the mixed results concerning the impact of outsourcing on quality which relies on contract specification and monitoring. Quality problems come from poor contract specification and improvements in quality can occur as a result of better specification of the service and better contract monitoring. For instance, Domberger, Meadowcroft, and Thompson (1986) argue that penalty clauses for non-performance should be sufficient to ensure the required standard of service. Nevertheless, there is no hard empirical proof that an improvement in contract design and monitoring can improve service quality. This is precisely how this paper contributes to the literature by filling this gap.

2.3. Theoretical explanations and propositions

As emphasized in the introduction, the agency theory offers a particularly relevant framework to explain why contracting out a public service may adversely impact its delivered. Because the agent may not share the principal's goals, and because the agent is more familiar with the details of the task, he may have both motive and opportunity to behave in ways that maximize his own interest at the expense of the principal's. First, the inability to entirely observe the characteristics of the private operator can lead to select a less efficient operator. This phenomenon of adverse selection can take the form of an "aggressive bidding" where the candidate voluntarily underestimates the costs of the service anticipating that it will be possible to compensate those low cost by providing low quality. It can also take the form of a "winner's curse" where winner of the bid was too much optimistic and then turns unable to keep its promises and adjust by providing low quality. The second information asymmetry issue, moral hazard, comes from the inability of the public authority to observe the actions of the operator. The latter is thus able to "shirk" by delivering a lower quality of the service. Nevertheless, there exist means at the disposal of the principal to discover the agent's private information about characteristics and behavior. Indeed, the principal can incur agency costs in order to constrain the agent's opportunity to shirk by increasing the monitoring of the agent's actions (behavior-based contract) and by including

incentives in the contract (outcome-based contract). Thus, public authorities which decide to outsource a public service have to invest on contract specification and monitoring if they want to minimize the problem of quality-shading.

Indeed, the public authority can reduce both the motive and the opportunity of the operator to behave opportunistically (i.e. to provide low quality) by investing in monitoring of operator's performances and/or providing incentives in the contract. Obviously, the upstream prerequisite is to precisely describe the outcome that has to be reached by the operator and the ways performance will be evaluate and monitor by the public authority. Thus, this leads us to our two first propositions:

Proposition 1. *Adverse effect on quality due to moral hazard can be reduced by increasing the monitoring of the contract.*

Proposition 2. *Adverse effect on quality due to moral hazard can be reduced by increasing the level of incentives of the contract.*

Furthermore, an increase in contractual precisions can also lead to diminish the severity of adverse selection issue. Indeed, even if performances are observed ex post, the contract specification may indicate ex ante how to achieve these performances. Strengthening the contract specification thus allow the public authority to precise how to achieve the expected performance and ease the measure of the adequacy of the resources that the operator commits itself to use (which is precisely the object of the "technical memorandum" that each bidders must provide during the competitive auction process) and theoretical resources that they should mobilize. Adverse selection is then much easier to detect and, consequently, to avoid. Moreover, as soon as the contract entails incentive mechanisms, it acts as a discriminating device which allows the public authority to attract more efficient operators, i.e. operators confident in their ability to fulfill contractual obligations and to deliver the expected level of quality. Such a statement is, for instance, highlighted by Lazear (2000) in his empirical study of performance pay schemes. In this work, Lazear shows that the introduction of this new remuneration schemes had allowed the firm to attract more productive workers. We thus have the following and last proposition:

Proposition 3. *Adverse effect on quality due to adverse selection can be reduced by increasing both monitoring and incentives of the contract.*

2.4. Alternative theoretical explanations

In this paper, we decided to focus on agency theory. Nevertheless, other theoretical framework could have also been mobilized to explain adverse effect on quality.

First, the transaction cost theory also emphasizes that incomplete contract can open rooms for opportunistic behavior of parties that can take the form of shading on quality. But transaction cost analysis is essentially concerned with when and why a business activity is vertically integrated and it is less useful for understanding relations between contractors

who have already decided to remain independent. Nevertheless, our propositions set out in the previous subsection can be related to the concept of credible commitments (Williamson (1983)). By increasing the level of contractual details and incentives, parties more credibly commit to fulfill their obligations, notably in terms of service quality.

Second, the incomplete contract theory also provides explanations to the phenomenon of quality shading. For instance, the seminal work of Hart, Shleifer, and Vishny (1997) states that private operator has stronger incentives both to reduce costs and to improve quality than the public sector. Nevertheless, they also highlight that the cost-reduction incentive may overwhelm the quality-improvement incentive if the quality is difficult to measure. Such difficulty leads to necessarily incomplete contracts which open rooms for opportunistic behavior. Indeed, when a contract is incomplete, the operator could opportunistically exploit contract ambiguities to its advantage and to the detriment of the public authority. As mentioned previously in this paper, the public authority must incur transaction cost in both contract design and monitoring in order to minimize such risk. More precisely, contract should establish performance measure, introduce explicit inspection processes and provide for penalties in the event of non-compliance with performance obligation. Obviously, a precondition for the proper functioning of such mechanisms is the ability to measure performances of the external provider. While it appears quite easy in terms of costs, the previously mentioned work of Hart, Shleifer, and Vishny (1997) raises the sticking point of contracting difficulties for outsourcing public services, which is the difficulty to measure quality. This second theoretical approach does not fit with what we observe since we are focusing on house cleaning contracts which correspond to a sector where quality is quite easily measurable.

2.Institutional and contractual framework

European Directives stipulate that service contracts, beyond the EU-thresholds (around 200 keuros), have to be tendered through an open call for tenders. While this obligation increases transparency and thus, limits abuses in discretion with public funds, it still appears as being insufficient to systematically obtain the best value for money. The rules established by European Directives make the emphasis on the awarding process. As pointed in the literature review, the *ex ante* competition does not prevent firms from behaving opportunistically. This section is dedicated to the analysis of the institutional framework and its limits.

2.1.Institutional framework

European Directives are transposed into national law through the French Public Procurement Code. As soon as service contracts reach the EU-thresholds, both regulations constrain the French public buyers (and thus, the European public buyers) to use the traditional open call for tenders. While a derogation is possible in specific cases, it is still the most used mechanism: according to a EU-report[2], 68% of service contracts awarded between 2006 and 2010 followed an open procedure. This the procurement mechanism used to award the contracts we study in the empirical section.

This procurement process is made of different steps. First, the buyer defines its needs and it launches a publicity. Second, firms send documents related to their candidature (the employee number, their references, their competences, ...) and their offer. Third, the buyer analyzes the candidatures. Whenever they are satisfying, the associated offers are also analyzed. Finally, the winner is the "most economically advantageous offer".

This selection process is deliberately rigid so as to respect the principles of the Code: transparency, equal treatment of candidates, freedom of access to public contracts. In theory, this rigidity is supposed to ensure the efficiency of competitive mechanisms : any firm can submit a bid which is evaluated according objective and transparent criteria; ignoring the number and the offers of rivals incite bidders to reveal their private information so as to maximize their chance to win the contract. Nevertheless, this statement is true only if, in particular, quality is observable and verifiable by a third party.

Otherwise, firms can decrease the delivered quality, so as to ensure winning the contract (adverse selection) or so as to increase their profits during the execution of the contract (moral hazard). Finally, open auctions focus on the awarding process, which ensures, under rarely gathered conditions, an efficient contract's execution.

2.2.Contracting on quality: from moral hazard to perfect contractibility

Contracting on quality is an arduous task. As previously mentioned, while the concept of quality is universally understood, applying it rigorously turns out to be devilishly difficult (Domberger and Jensen (1997)). Obviously, those difficulties vary from one sector to another. For instance, Brown and Potorsky (2005) send a survey to public managers about the transaction cost dimension of a variety of basic local government services. More precisely, they ask manager to determine the "ease of measurement" in 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

[2]Impact of the Effectiveness of EU Public Procurement Legislation, p.12, Part 1, European Commission (2011)

accurately represent the quantity and the quality of the service" and 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 achieve". 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 operation of mental health program (score > 4). According to such a typology, contracts established to outsource cleaning services might be rather complete and quality-shading might be rather scarce on this sector. Indeed, contract is the formal mechanism through which the expected level of quality is specified. A well-developed contract which clearly specified the service outcome is essential to ensure the level of service meets the expectations of the buyer. Contract details is thus the first element to care about to avoid quality-shading. The second one is the contract monitoring. In addition to precisely define the quality standards, the buyer must also foresee contractual responses to monitor contractor performance. Indeed, without an effective monitoring regime, there is no way of knowing whether the service provider is meeting the expectations specified in the formal contract. Contract specification and monitoring are thus the two fundamental elements which allow to outsource a service without suffering from an unplanned reduction of quality.

In our case of house-cleaning contracts, those elements take two forms. On the one hand, 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 to deadlines to provide contractual documents such as service manuals and periodical activities' reports and to answer to buyer's requirements). On the other hand, contract monitoring requires two elements. First, it is necessary to specify in the contract how the quality will be evaluated and by whom. In our case, it takes the form of an appendix to the contract which precisely described the quality control process. 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) are organized once per month and the contract holder is notified 48 hours in advance. During those controls, the representative of the buyer completes an evaluation grid (called "quality control sheet") which lead to a final mark out of 100. 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 are multiplied by two if the mark obtained during the second control is still less than 80. Second, contract monitoring also needs to foresee coercive means at the disposal of the public buyer to enforce contractual specifications and, in case, to punish any breach of contractual obligations. This is precisely the goal of the previously mentioned penalties. Indeed, in our case, the public buyer is contractually authorized to impose penalties when the cleaning companies fail to fulfill their obligations (in terms of quality of the service as

well as in terms of mandatory documents supply). Aside penalty clauses, contracts also contain 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 body we study in this paper is well protected against moral hazard issue resulting in low quality of the service. This is not what we observe in our data concerning contracts between the public buyer and several firms specialized in cleaning services. Users' complaints, breach and early termination of contracts persist despite the use of open auction and the definition of quality standards in the contract. To tackle those problems, the public buyer has decided to modify its formal contracts in the two different directions previously mentioned. From the specification point of view, the new versions of the General Condition and of the Guide of Special Techniques Specifications are more precised concerning the obligations of the operator. For instance, it describe more technically the way cleaning must be performed and include a glossary of cleaning operations. From the monitoring point of view, the new form of the contractual arrangements increases the level of details provided to the cleaner about how the evaluation are organized and add the possibility of unplanned and not contradictory controls by public agents. Moreover, the new contract adds three new categories of penalty clauses including one about the mandatory documents that the cleaner must periodically provide to the public buyer. Aside this formal transformation, the public buyer also decides to be more rigorous in its collection of penalties which were used to be unpaid in the past. Then, without any changes neither in the tendering process nor in the quality indicator used, the public body aim at improving the general level of provided quality of the service. Consequently, it offers us a unique opportunity to test whether the quality of outsourced service can be upgraded by improving contract design (in terms of specification and monitoring), i.e. by trying to move from moral hazard to perfect contractibility.

3.Data

3.1.Scope of the database

Paris Habitat-OPH, the public buyer we study, organizes its cleaning activity by establishing a geographic allotment: the buildings located in a same area correspond to a given lot j whose characteristics remain rather[3] constant over time. This organization is a key-point to exploit the panel properties of our data. For instance, it allows to assess whether different types of contracts and/or suppliers generate different level of performances on a

[3]Since we can naturally not exclude that our buyer will build, buy or sell a building, the characteristics of a lot can still marginally evolve. Our data enable to control for this types of changes. This aspect is discussed when presenting the control variables.

same lot.

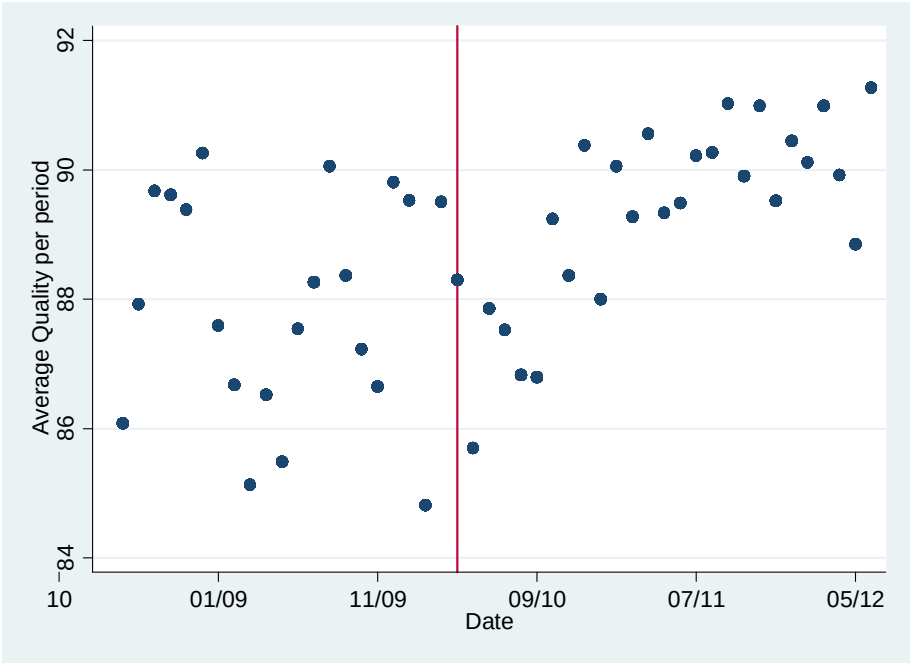
We have collected information about the quality which has been delivered on 49 lots between July 2008 and June 2012; 102 contracts[4] were in progress during this period. It leaves us we an unbalanced panel database of 49 lots which are followed over a period of 48 months. Among these 49 lots, 45 have been renewed one time over the period (which means that we observe a lot through two different contracts and sometimes through two different suppliers), whereas 4 have been renewed two times. We know that these contracts are shared out among 13 firms and they are managed by 6 different directions on the buyer's side (each direction is in charge of a geographical area). We also have information regarding the monthly level of implemented quality, the *ex post* monitoring, the tendering phase and the price of the contract. Summary statistics of our variables are presented in table 10, whereas table 12 provides descriptive statistics when distinguishing contracts before and after the change in contract design; we called panel A the 50 contracts launched before April 2010 and Panel B the 52 launched after this date.

3.2.Dependent variable

We are mainly interested in explaining the variable $Quality_{ijt}$, which measures the level of quality delivered by firm i on lot j at time t. Quality indicators are scored based on a scale of 0 to 100. The average level of quality is 88.91, which is around 9 points beyond the threshold implying penalties. 2 points distinguish panel A from panel B; therefore, this last seems a bit more performing. Figure 2 reports the quality evolution over time. The distribution seems rather uniform and we do not notice major changes neither between the two panels nor between the two periods. Nevertheless, figure 3 reports the average value of the variable at each period and shows that $Quality_{ijt}$ significantly increased after April 2010. Given that, after April 2010, most of the on-going contracts belongs to panel B, this observation strongly corroborates our intuition that the level of service increased after the change in contract design. However, since the variable $Observe_{ijt}$ captures that $Quality_{ijt}$ was not measured in 39% of all cases, we suspect a sample selection issue we will have to deal with later on because it can bias our interpretations of the figures.

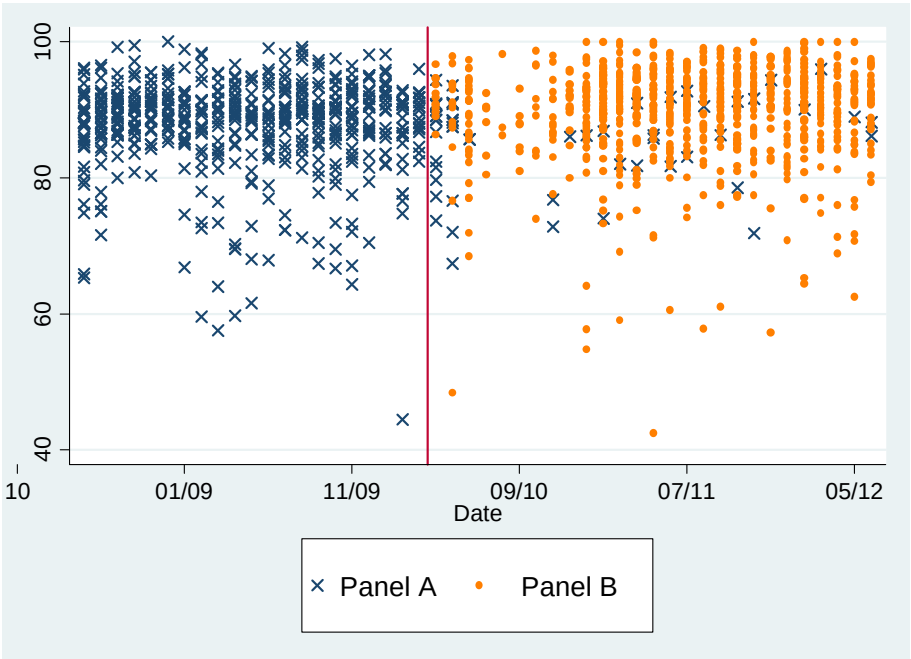
[4]We excluded from our analysis the contracts which are made of particular social clauses.

Figure 3: Mean value of the quality indicator, per period



The red line corresponds to t = April 2010

Figure 2: Quality evolution over time



The red line corresponds to t = April 2010

3.3.Independent variables

Contract design and *ex post* monitoring

We aim at assessing the impact of an exogenous change in the contract design and to assess whether our buyer's active behavior toward *ex post* monitoring is beneficial or not. Both aspects have been frequently pointed out as promising - while complicated/costly/unfeasible - solutions to address the issue of poor performances in public procurement.

We first constructed the variable New_Design_{jt}, which is a binary variable taking the value 1 whenever the contract running on lot j at time t belongs to Panel B. This variable captures the exogenous chock affecting all the contracts launched from April 2010: this shock resulted in specifying the contracts in more detail and in increasing the penalties regime associated to low quality. The reinforcement of the contractual obligations toward quality may attract more performing firms and/or increase firms' incentives toward quality. Therefore, New_Design_{jt} may positively impact our dependent variable $Quality_{it}$.

Then, we built two variable to assess the way the buyer puts the *ex post* monitoring into practice. First, the variable $Monitoring\_Freq_{ijt}$ 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 equals to 0.54. It is higher in Panel A than in Panel B, reflecting that our buyer decreased the frequency of the controls, while he simultaneously reinforced the contractual requirements and the penalties paid. One interpretation would be a substitution effect between the greater amounts of penalties (allowed through the reinforcement of contractual requirements) 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. However, we may reasonably expect that a firm managing a frequently monitored contract (*i.e.* having a higher $Monitoring\_Freq_{ijt-1}$) may feel more intensely under the scrutiny of the buyer. As a result, it may reach better current performances.

Second, the variable $Penalties_{ijt}$ corresponds to the overall value of penalties paid by firm i on lot j at time t. Around 731 additional euros per contract have been payed because of not reaching the contractual requirements. It is rather low when considering the contracts' size, which reflects that penalties are a least resort. 150 euros of penalties still distinguish Panel A from Panel B, which may capture our buyer's decision to reinforce the sanction mechanisms. This evolution appears even more clearly in table 1, which summarizes the value of penalties paid each year by the cleaning contracts' suppliers.

Table 1: Penalties per year

Year	2008	2009	2010	2011	2012 (until June)
Penalties (in Euros)	21 646.54	12 347.44	37 297.18	30 641.48	20 800

Since penalties account for small amounts and since reaching a fixed amount of penalties

lead to a breach of the contract, we expect that it is the accumulated value of penalties paid in the past (i.e. the variable $Penalties_{ijt-1}$) rather than the (few) amount payed every month that may have an incentive effect on the suppliers' decision to improve their current performances.

Finally, we are interested in assessing the impact of the change in contract design and monitoring on the delivered quality. These two changes occurred simultaneously in April 2010 and therefore, their effects may be indifferently captured through the variable New_Design_{jt}. Nevertheless, since we have information about the putting into practice of the contracts, *i.e.* about the diligence of the monitoring as well as the severity of the sanctions, we can disentangle the impact of the contract specification *per se* from the impact of the *ex post* monitoring. Moreover, by crossing the variable associated to the new design with those associated to the monitoring, we can assess whether the impact of the monitoring also depends on the formal contract itself. Additional details about the way to perform this type of analysis are discussed when presenting the empirical specification.

Heterogeneity across contracts

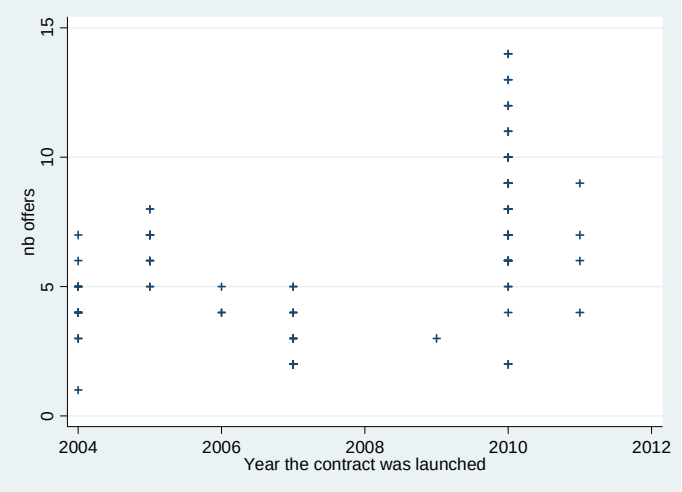
We can use additional controls to capture some heterogeneity across contracts: some dimensions (like the degree of competition, the price or the scope of the contract) can affect the level of delivered quality. The variables associated to these dimensions and their expected impacts on quality are described below. However, we suspect that the price and the degree of competition may have been affected by the change which occurred in April 2010, making them outcome variables (and not control variables). According to Angrist and Pischke (2008), outcome variables are bad controls, but *"timing matters"* : " *Variables measured before the variable of interest was determined are generally good controls. In particular, because these variables were determined before the variable of interest, they cannot themselves be outcomes in the causal nexus*". We use this statement to determine whether the variables associated to price and competition have to be included as controls or not.

First, $Nb\_Accommodations_{jt}$ captures the number of accommodations which are included on lot j at time t. We indeed aim at exploiting the panel nature of our data by following a same lot over time. To correctly perform it, the lots have to remain stable. Nevertheless, we cannot exclude that our buyer will build, buy or sell a building, affecting the characteristics of a given lot. The variable $Nb\_Accommodations_{jt}$ enables to account for the potential changes in the size of the lots. Since our buyer's strategy in favor of allotment is based on the reasoning that larger lots are more difficult to manage, we expect that an increase in the number of accommodations negatively impacts on the level of delivered quality.

Second, Nb_Offer_{jt}, captures 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, lowering their rents; the buyer should obtain a best value for

money. However, as pointed out in the literature review, more competition might also encourage aggressive bids at the expense of quality. Indeed, if quality is not perfectly verifiable and if performing firms can not be rewarded through positive discrimination during the future tendering processes, firms have not interest in taking care of unverifiable dimensions: it would be costly, leading firms to post higher bids and decreasing their winning probability. Therefore, the impact of competition on quality in not consensual and hard to make out.

Figure 4: Number of received offers and time the contract is awarded

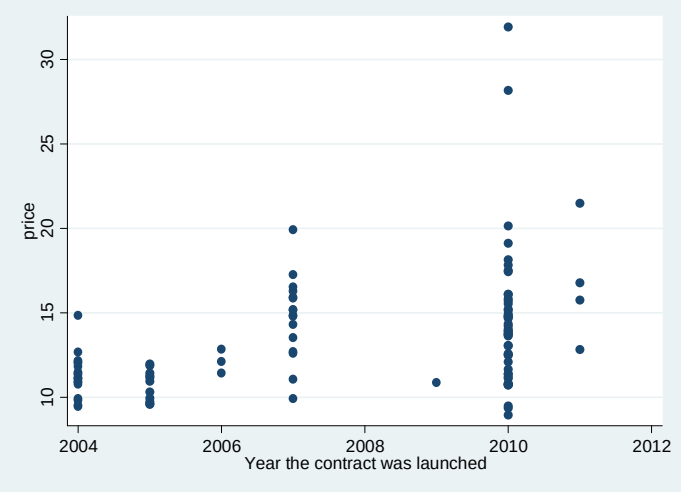


So as to clarify the sense of the causal relationships between the degree of competition, the change in the contract design and the delivered quality, figure 4 describes the relationship between the number of offers and the year the contract is awarded. All the contracts starting from 2010 include the change which occurred in April 2010. We can see that the number of offers tends to increase for contracts launched in 2010 and 2011; this relationship is corroborated by table 11, which presents the matrix of correlations between the different contracts' characteristics (see the correlation between the variables Nb_Offer_{jt} and New_Design_{jt}). Hence, the change in contract design occurred at the same time as an increase in the degree of competition. At first view, it is surprising that reinforcing the contract design and the penalties regime creates an increase in the number of potential suppliers. Nevertheless, practitioners argue that this change sent to firms the signal that the buyer was unsatisfied with its current main suppliers and aimed at finding new performing firms. This signal may have renewed the set of interested firms, creating a chock on the degree of competition. Finally, the change which occurred in April 2010 may not only act, as expected, on the *ex post* dimensions, but also on the *ex ante* process : in terms of quantity of suppliers, but also in terms of quality. As previously said, this increase in competition (in terms of quantity) is likely to generate costs saving at the expense of quality. However, if this increase in competition actually enabled to renew the set

of suppliers (19% of the contracts launched in 2010 are awarded to new firms) by attracting performing firms, we may also observe a decrease in adverse selection and thereafter, an improvement of the level of service. Thus, the impact of competition on quality is still hard to anticipate. In any event, the information resulting from figure 4 leads us to suspect a decrease in adverse selection. It also points out the difficulty to consider the number of offers as an independent variable. This last dimension will be discussed below.

We third built the variable $Price_{ijt}$, *i.e.* the bid of the winning firm i for lot j at time t divided by the number of accommodations. It is one way to control for the competitiveness of the winning offers. However, this way remains imperfect. Indeed, each contract is made of a two-part tariff with a fixed part F_k and a purchase orders part $\sum_m p_{km}.q_{km}$; the magnitude of these two dimensions is specific to contract k; the prices p_{km} of each task m=*1,...,M* are those contained in the winner's offer, whereas the performed quantities q_{km} are progressively defined to meet the buyer's needs. When advertising firms, the buyer simply sets the minimal amount of purchase orders, $Inf(\sum_j p^*_{km}.q^*_{km})$, and its maximal amount, $Sup(\sum_j p^*_{km}.q^*_{km})$, with p^*_{km}, corresponding to the estimated prices and q^*_{km}, the estimated quantities. However, the documents we had access to collect the data did not give us a synthetic and standard way to control for the competitiveness of the prices submitted for the purchase order part, constraining us to limit our information about prices to the fixed part. Hence, the variable $Price_{ijt}$ only controls for the degree of competitiveness related to the fixed part F_k of the contract. According to the literature, if adverse selection (or winner curse) is possible, more aggressive bids might tend to lower the delivered quality.

Figure 5: Winning price and time the contract is awarded



Once again, we suggest in table 5 to analyze how the variable $Price_{ijt}$ evolves depending on the year the contract is awarded. We can see that the price tends to increase from 2010; this relationship is also corroborated thanks to the matrix of correlations (see the correlation between the variables $Price_{ijt}$ and New_Design_{jt} in table 11). Hence,

unsurprisingly, the reinforcement of both the contract design and the penalties regime is associated to less competitive bids: firms rationally compensate the costs they incur to fulfill the increasing expectations of the buyer by increasing their posted prices. Despite the fact that the suspected trade-off between price and quality might naturally be matter of concern in this paper, considering $Price_{ijt}$ as an independent variable which affects the delivered quality may be an issue: it can be considered as an outcome variable, which has been significantly affected through the change in April 2010.

Finally, we have to decide about whether introducing Nb_Offer_{jt} and $Price_{ijt}$ as control variables. Following Angrist and Pischke (2008), we focus on the timing of the decisions to make that choice. Since Nb_Offers_{jt} and $Price_{ijt}$ are fixed before the time the quality is delivered, we may decide to use them as control variables. However, this reasoning ceases to be true for the variable $Price_{ijt}$ if firms decide about the level of quality they will deliver in the same time as they propose their price. Hence, while price and quality are strategic variables made by a same agent for unobservable but probably related reasons, the number of offers is fixed before the delivering of quality and it results from decisions which do not depend on the winner's decision. As a consequence, we decide to include the variable Nb_Offers_{jt} but to exclude the variable $Price_{ijt}$ when estimating the determinants of quality; we prefer to separately explore its determinants in a second step of our analysis.

3.4.Empirical specification

We are first interested in understanding which dimensions improve suppliers' performances regarding quality. This is made possible because the buyer we study built a quality indicator, called $Quality_{ijt}$, measuring the level of service reached by firm i on lot j at time t. Therefore, we first estimate the following model:

$$Quality_{ijt} = \beta_1 + \beta_2.New\_Design_{jt} + \beta_3.Y_{jt} + w_j + \epsilon_{ijt} \quad (1)$$

where Y_{jt} is a vector of variables capturing the characteristics of the lot j at time t, New_Design_{jt} is our first main variable of interest capturing the change in contract design; we expect this variable has a positive impact on quality. To control for unobservable biases due to the nature of the lots, we also add w_j, lot fixed effects.

This first model is a simple before-after estimation to assess the impact of the switching from panel A to panel B. The following models aim at breaking down the global effect of the switching into different elements: when switching from panel A to panel B, which part of the observed effects comes from the *ex post* monitoring (proposition 1)? From the change

in the structure of incentives (proposition 2)? From the reduction of adverse selection (proposition 3)?

Hence, in a second step, we estimate the following model with firm fixed effects z_i:

$$Quality_{ijt} = \beta_1 + \beta_2.New\_Design_{jt} + \beta_3.Y_{jt} + w_j + z_i + \epsilon_{ijt} \quad (2)$$

Firm fixed effects capture the impact of adverse selection on quality. Hence, this model permits to disentangle the impact of adverse selection from the impact of moral hazard. For instance, if the coefficient associated to the variable New_Design_{jt} decreases when switching from equation (1) to equation (2), it means that (at least) part of the increase in quality comes from the selection of more efficient firms; this would validate our third proposition. Hence, in equation (2), the variable New_Design_{jt} captures changes in the level of quality due to a decrease (or an increase) in moral hazard issues. According to our propositions 2 and 3, moral hazard might be solved through two channels : a change in the structure of incentives and a change in the putting into practice of the contract. To disentangle both aspects, we run a third model, which includes the variables X_{ijt} related to the *ex post* monitoring but also the interaction terms between the demeaned value of the variables related to the *ex post* monitoring ($\widehat{X}_{ijt}$=$X_{ijt} - \bar{X}_{ijt}$) and the variable New_Design_{jt}[5]:

$$Quality_{ijt} = \beta_1 + \beta_2.New\_Design_{jt} + \beta_3.Y_{jt} + \beta_4.X_{ijt} + \beta_5.(New\_Design_{jt} * \widehat{X}_{ijt}) + w_j + z_i + \epsilon_{ijt} \quad (3)$$

This specification allows to distinguish the impact of the change in the structure of incentives (through the variable New_Design_{jt}), a change in the monitoring of the contract (through the variables X_{ijt}) and a change in the monitoring of the contracts which belong to panel B (through the interaction terms). The coefficient associated to the variable New_Design_{jt} is related to our proposition 2, whereas the coefficients associated to X_{ijt} and the interaction terms are associated to our proposition 1. Our first proposition indeed postulates that the monitoring of the contract might increase firms' incentives to deliver high standard of quality. However, we also suspect that these incentives might be even more efficient is they are supported through a well-specified contract. Hence, we expect that the coefficients associated to the interaction terms should be higher than those associated to X_{ijt}.

[5]See pages 68-69 of chapter 4 in Wooldridge (2001) for the explanation regarding the construction of the interaction term

Moreover, to account for the missing quality indicators and potential selection biases they generate, we run additional regressions. The estimation strategy is presented in section 4.2. We also explore the impact of the decision of the administrative court of Paris to make sure that our change in contract design is exogenous. Results are presented in section 4.3.

Given that we have decided to consider the prices as an outcome variable, we also explore the impact of the change in contract design on this dimension. Therefore, we estimate the following equation:

$$Price_{ijt} = \beta_1 + \beta_2.New\_Design_{jt} + \beta_3.Y_{jt} + w_j + z_i + \epsilon_{ijt} \quad (4)$$

This estimation allows 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 the 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. Therefore, we challenge this issue by testing various specifications: each one includes a different type of fixed effects. This strategy still allows to test whether results are stable accross specifications.

4. Results

4.1.The determinants of quality

Results are presented in table 2.[6] We aim at assessing the impact of the contract's structure and its putting into practice on the delivered quality. So as to exploit the panel nature of our data, all the models include fixed effects by lot. Moreover, we also systematically add some controls to account for heterogeneity across contracts. Model 1 only includes the variable related to the change in contract's structure. In Model 2, firms fixed effects are added to account for potential adverse selection's effect on quality. In Model 3, we cross the variables capturing the change in contract design with the demeaned value of

[6] Since we have lagged variables in model 3, the number of observations is equal to 1359 and not 1382 : regarding 23 observations, we observe the level of quality when the contract started; however the variables related to the *ex post* monitoring are coded as missing values for this first period (the contract just started; hence, by definition, we have no information about past monitoring). As a consequence, we decided to drop these 23 observations from each model to replicated our estimates on comparable samples.

Table 2: How to implement quality?

	Model 1	Model 2	Model 3
	$Quality_{ijt}$	$Quality_{ijt}$	$Quality_{ijt}$
New_Design_{jt}	2.526***	2.006***	1.433*
	(0.605)	(0.721)	(0.738)
$Monitoring\_Freq_{ijt-1}$			-4.283***
			(1.494)
$New\_Design_{jt} * \hat{Monitoring\_Freq}_{ijt-1}$			8.601***
			(1.813)
$Penalties_{ijt-1}$			-0.000+
			(0.000)
$New\_Design_{jt} * \hat{Penalties}_{ijt-1}$			0.001***
			(0.000)
Nb_Offers_{jt}	0.063	0.018	0.062
	(0.159)	(0.196)	(0.196)
$Nb\_Accommodations_{jt}$	-0.002***	-0.001	-0.002+
	(0.001)	(0.001)	(0.001)
$Constant_{ijt}$	91.332***	85.040***	88.308***
	(1.505)	(2.288)	(2.696)
Lot	×	×	×
Firm		×	×
N	1359	1359	1359
$Adj - R2$	0.21	0.24	0.25

Significance levels: +0.15, * 0.10, ** 0.05, *** 0.01; Robust standard errors in parentheses.

the variables related the putting into practice of the *ex post* monitoring (we construct $\hat{Monitoring\_Freq}_{ijt-1}$ and $\hat{Penalties}_{ijt-1}$): we want to test whether the impact of *ex post* monitoring is different in Panel A and Panel B, *i.e.* whether the enforcement of the contract depends on the contract design itself.

Whatever the specification, we find that the change in contract design has a significant and positive impact on the delivered quality. Moreover, since the coefficient associated to the contract design decreases when switching from Model 1 to Model 2, we find that part of the improvement comes from the limitation of adverse selection; it validate our proposition 3. Finally, Model 3 shows that non-precisely specified contracts makes *ex post* monitoring harmful: the *ex post* monitoring is efficient only if it is supported through a well-specified contract. Moreover, since the variable associated to the change in contract design remains positive and significant when controlling both for adverse selection and contract enforcement, we conclude that the change in the structure of incentives has *per se* an impact on the delivered quality, corroborating proposition 2.

4.2. First robustness check: sample selection issue

We do not observe $Quality_{ijt}$ in 39% of the cases. The reasons we may invoke to justify this sample selection are multiple, since the decision to make a control is highly decentralized, left to the caretaker's discretion, whose motive are hard to make out. Therefore, the way this sample selection affects our estimates is difficult to anticipate.

To deal with it, we can still use a two-step heckman method (Heckman (1979)). Provided that we achieve to explain why quality is observed or not, it enables to extrapolate the missing quality indicators as if they would have been observed. Thereafter, the model indicates whether the bias due to sample selection is severe and it accounts for the bias effect both on the dependent and on the independent variables. To do so, the model is made of two steps: the first step corresponds to the selection equation, modeling, in our particular case, the probability of being observed; the second steps corresponds to the corrected outcome equation: it explains the level of quality, given it is observed. However, to be overidentified, 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 $Observe_{ijt}$ (it refers to the *instrument relevance* issue), but not with any unobservable that could influence the variable $Quality_{ijt}$ (it refers to the *exclusion restriction* issue). Therefore, each instrument must respect the following conditions:

$$\begin{cases} Corr(Observe_{ijt}, instrument) & \neq 0 \\ Corr(\epsilon_{ijt}, instrument) & = 0 \end{cases}$$

We suspect $Monitoring\_Freq\_Others_{ijt-1}$ being a valid instrument. It measures the rate of control at *(t-1)* on a sub-sample of observations. This sub-sample is made of all the observations related to the period *(t-1)* with the exception of both the observations related to the firm i and the observations related to the direction managing lot j. We detail below why this variable may be a valid instrument.

Instrument relevance. To overcome the lack of clarity regarding caretaker's motives to perform controls, we assume that people having both the same job and the same employer may observe and influence each others, through a spillover effect. Thus, on average, they may share the same motives over the short term. It enables us to suspect that one caretaker's diligence to carry out *ex post* monitoring depends on the observation of the others caretaker's diligence; this may be captured through their control rate at *(t-1)*.

Exclusion restriction. Since a valid instrument must be uncorrelated with the residual of the second step, we build our instrument on a sub-sample of observations so as to exclude those likely to be correlated with the unexplained performances of firm i on lot j at time t. Thus, the sub-sample excludes the observations related to the firm i and those related to the direction managing lot j. Indeed, a firm is likely to observe the past controls frequency of its territorial direction and/or the past controls frequency of its other ongoing

contracts: this information may influence its behavior; for instance, a firm may be more prone to increase the quality of service if it observes that the caretakers tend to increase their controls. On the contrary, this firms should not be aware, at least in the short term, of the caretakers' propensity to perform controls in the other directions regarding the other firms: this is true assuming that firms do not communicate among each other on a highly regular basis. As a consequence, this private information of the caretaker in charge of managing lot j at time t should not influence a firm's incentives toward quality. Moreover, since the firm's rating is based on rather objective criteria, it should also not influence the firm's rating, satisfying the exogeneity condition.

Results of the two-step Heckman estimates are presented in table 4. Whatever the specification, we can see that our instrument is significant at the level of 1%, satisfying the relevance condition. Thereafter, 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. Despite this sample selection, our main findings remain extremely stables. There is one notable change in Model 6 : the coefficient associated to the variable New_Design_{jt} ceases to be significant. Nevertheless, since *Lambda* is no more significant, we suspect that we have enough control variables in this specification to overcome the selection bias issue. Therefore, Model 3 might be more accurate than Model 6.

4.3. Second robustness check: testing for the exogeneity of the change in contract design

Since the change in contract design occurred after the decision of the administrative court of Paris (, we want to make sure that the change in contract design can actually be considered as an exogenous shock. To do so, we replicate our estimations by including the variable $Decision_t$. Results are presented in table 6. Whatever the specification, we can see that the decision has no impact neither on the buyer's propensity to observe the quality, nor on the level of delivered quality: the agents have not adapted their behavior to this decision. This result confirms what we observe on figure 3: the average level of quality clearly increased after the change in contract design, not after the decision of the court. Moreover our results still remain stable.

4.4. The determinants of price

To complement our analysis, we explore the determinants of prices. Results are presented in table 8. We first run models 10, 11 and 12, which include variables related to the heterogeneity within and across lots. The three specifications include different fixed effects to check the robustness of the results. Since we have few observations, we cannot include

all the fixed effects within one model without dramatically reducing the degrees of freedom.

These first models indicate a price increase due to the change in contract design. The number of offers has the expected impact and is line with previous results from the literature that more competition attracts lower bids.

However, the contracts we study have been awarded between 2004 and 2011. Hence, the increase in price we observe might be due to a general increase in price in the sector, which is collinear with other changes over time (including the change in contract design we study). Therefore, to control for inflation, we collected a price index of cleaning services on the website of the French National Institute of Statistics (the "INSEE"). We call it $Price\_Index_t$ and replicate our estimations with this variable. We find significantly different results: the new design has no more any significant impact on the received prices, whatever the specification we consider.

5. Conclusion

According to our literature review, outsourcing simple transactions might result in cost savings. However, their remain some doubts about the way these savings are realized. Indeed, previous studies found mixed results regarding the impact of outsourcing on quality, suggesting that outsourcing does not systematically result in better efficiency: to obtain the best value for money, the way outsourcing is realized matters too, even for simple transactions like cleaning services. The data we collected enable to test propositions related to the impact of some dimensions of the outsourcing process on the delivered quality. It appears that improving the specification of the contract, the incentives structure and the monitoring of the contract enable to increase the level of service through two channels: it solves both adverse selection and moral hazard issues. We also find some evidence that *ex post* monitoring can be harmful when it is not supported by a well-defined formal contract : one interpretation would be that, without having the possibility to implement a relational contract[7] and without having a clear formal contract, trying to carefully put into practice a contract does not make punishment mechanisms credible; it just deteriorates the relationships between the buyer and the suppliers. In addition, we find that increasing the risk allocated to suppliers does not systematically result in a significant increase in prices; it validates the idea that asymmetries of information result in rent-seeking behaviors that can be diminished by providing the good incentives. Finally, our results corroborates propositions from the literature that there exists no good candidate to outsourcing *per se*: obtaining a better value for money in outsourcing processes might go through a good formal contract, with an incentive structure and an application of this incentive structure.

[7] European rules in public procurement do not allow to take reputation into account

References

Angrist, J. D., and J.-S. Pischke (2008): *Mostly Harmless Econometrics: An Empiricist's Companion*. Princeton University Press.

Ascher, K. (1987): *The Politics of Privatisation: Contracting Out Public Services*. MacMillan Education, London.

Blom-Hansen, J. (2003): "Is Private Delivery of Public Services Really Cheaper? Evidence from Public Road Maintenance in Denmark," *Public Choice*, 115(3-4), 419–38.

Bulow, J., and P. Klemperer (1996): "Auctions versus Negotiations," *American Economic Review*, 86(1), 180–94.

Calzolari, G., and G. Spagnolo (2009): "Relational Contracts and Competitive Screening," CEPR Discussion Papers 7434, C.E.P.R. Discussion Papers.

Dijkgraaf, E., and R. Gradus (2004): "Cost savings of contracting out refuse collection," Public Economics 0409002, EconWPA.

Domberger, S. (1993): "Privatisation: What Does The British Experience Reveal?," *Economic Papers*, 12(2), 58–68.

Domberger, S., and P. Jensen (1997): "Gaining from outsourcing," *New Economy*, 4(3), 159–163.

Domberger, S., P. Jensen, and R. Stonecash (2002): "Examining the magnitude and sources of cost savings associated with outsourcing," *Public Performance and Management Review*, 26(2), 148–168.

Domberger, S., S. Meadowcroft, and D. J. Thompson (1986): "Competitive tendering and efficiency: the case of refuse collection," *Fiscal Studies*, 7(4), 69–87.

Fama, E. F., and M. C. Jensen (1983): "Separation of Ownership and Control," *Journal of Law and Economics*, 26(2), 301–25.

Farago, S., C. Hall, and D. S. (1994): "Contracting of Services in the Western Australian Public Sector," *Graduate School of Business, University of Sydney*.

Hart, O., A. Shleifer, and R. W. Vishny (1997): "The proper scope of government: Theory and an application to prisons," *Quarterly Journal of Economics*, 112(4), 1127–1161.

Hartley, K., and M. Huby (1986): "Contracting Out Policy: Theory and Evidence, in Kay, J., Mayer, C. and Thompson, D. (eds.)," *Privatisation and Regulation: The UK Experience*.

Heckman, J. J. (1979): "Sample selection bias as a specification error," *Econometrica*, 47(1), 153–162.

Jensen, M. C., and W. H. Meckling (1976): "Theory of the firm: Managerial behavior, agency costs and ownership structure," *Journal of Financial Economics*, 3(4), 305–360.

Lazear, E. P. (2000): "Performance Pay and Productivity," *American Economic Review*, 90(5), 1346–1361.

Lewis, G., and P. Bajari (2011): "Procurement Contracting With Time Incentives: Theory and Evidence," *The Quarterly Journal of Economics*, 126(3), 1173–1211.

McMaster, R. (1995): "Competitive Tendering in UK Health and Local Authorities: What Happens to the Quality of Services?," *Scottish Journal of Political Economy*, 42(4), 409–27.

Milne, R. G., and R. E. Wright (2004): "Competition and Costs: Evidence from Competitive Tendering in the Scottish National Health Service," *Scottish Journal of Political Economy*, 51(1), 1–23.

Szymanski, S. (1993): "Cheap rubbish? Competitive tendering and contracting out in refuse collection, 1981-88," *Fiscal Studies*, 14(3), 109–30.

Williamson, O. E. (1983): "Credible Commitments: Using Hostages to Support Exchange," *The American Economic Review*, 73(4), pp. 519–540.

Wooldridge, J. M. (2001): *Econometric Analysis of Cross Section and Panel Data*, vol. 1 of *MIT Press Books*. The MIT Press.

Appendix

Table 10: Descriptive statistics

Variable	Description	Mean	St. Dev.	Min	Max	N
$Quality_{ijt}$	Level of quality supplied by firm i on lot j at time t (from 0 to 100)	88.91	6.86	42.5	100	1382
$Observe_{ijt}$	Takes the value 1 if the $Quality_{ijt}$was measured, 0 otherwise	0.61	0.49	0	1	2248
$Penalties_{ijt-1}$ (§)	Overall value of penalties payed by firm i for lot j at time t-1 (in euros)	563.73	1 531.12	0	13 790	2195
$Monitoring\_Freq_{ijt-1}$(§)	Number of times the buyer controlled the quality supplied by firm i on lot j at time *t-1* / Maximal number of times it could have done it	0.54	0.28	0	1	2195
New_Design_{jt}	Takes the value 1 if the contract for lot j is awarded after April 2010, 0 otherwise	0.56	0.50	0	1	2248
$Decision_t$	Takes the value 1 after the decision of the administrative court of Paris, i.e. after May 2009, 0 otherwise	0.78	0.42	0	1	2248
$Price_{ijt}$	Winning bid of firm i at time t for the fixed part of lot j, divided by the 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
Nb_Offers_{jt}	Number of offers for lot j at time t	6.23	2.90	1	14	102
$Nb\_Accommodations_{jt}$	Number of accommodations for lot j at time t	1 846.167	593.76	544	3 066	102

(§) Variables built thanks to information related to periods from January 2008 through June 2008

Table 12: Sample comparisons

Variable	PANEL A	PANEL B
$Quality_{ijt}$	87.75	89.90
$Observe_{ijt}$	0.64	0.60
$Penalties_{ijt-1}$	495.64	620.59
$Monitoring\_Freq_{ijt-1}$	0.64	0.45
$Price_{ijt}$	12.26	14.41
Nb_Offers_{jt}	4.43	8.12
$Nb\_Accommodations_{jt}$	1 831.63	1 858.474

Table 4: Dealing with sample selection

	Model 4	Model 5	Model 6
	$Quality_{ijt}$	$Quality_{ijt}$	$Quality_{ijt}$
New_Design_{jt}	2.187***	1.655**	1.184+
	(0.611)	(0.688)	(0.734)
$Monitoring_{ijt-1}$			-5.192***
			(1.643)
$New\_Design_{jt} * \hat{Monitoring}_{ijt-1}$			9.238***
			(1.818)
$Penalties_{ijt-1}$			-0.000
			(0.000)
$New\_Design_{jt} * \hat{Penalties}_{ijt-1}$			0.001***
			(0.000)
Nb_Offers_{jt}	0.128	0.074	0.076
	(0.133)	(0.161)	(0.163)
$Nb\_Accommodations_{jt}$	-0.002***	-0.001	-0.002+
	(0.001)	(0.001)	(0.001)
$Constant_{ijt}$	88.576***	87.731***	91.324***
	(1.387)	(2.333)	(2.802)
	$Observe_{it}$	$Observe_{it}$	$Observe_{it}$
$Monitoring\_Others_{ijt-1}$	2.616***	2.681***	2.768***
	(0.162)	(0.165)	(0.180)
New_Design_{jt}	0.374***	0.460***	0.651***
	(0.110)	(0.124)	(0.130)
$Monitoring_{ijt-1}$			2.303***
			(0.277)
$New\_Design_{jt} * \hat{Monitoring}_{ijt-1}$			-2.550***
			(0.335)
$Penalties_{ijt-1}$			-0.000
			(0.000)
$New\_Design_{jt} * \hat{Penalties}_{ijt-1}$			-0.000
			(0.000)
Nb_Offers_{jt}	-0.088***	-0.093***	-0.107***
	(0.023)	(0.029)	(0.031)
$Nb\_Accommodations_{jt}$	0.000	0.000	0.000
	(0.000)	(0.000)	(0.000)
$Constant_{ijt}$	-0.439	-0.506	-2.049***
	(0.312)	(0.488)	(0.537)
Lambda	-1.923**	-1.873**	-0.861
	(0.775)	(0.746)	(0.753)
Lot	×	×	×
Firm		×	×
N	1359	1359	1359

Significance levels: +0.15, * 0.10, ** 0.05, *** 0.01; Robust standard errors in parentheses.

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

	Model 7	Model 8	Model 9
	$Quality_{ijt}$	$Quality_{ijt}$	$Quality_{ijt}$
$Decision_t$	0.310	0.297	0.591
	(0.499)	(0.497)	(0.499)
New_Design_{jt}	2.014***	1.497**	0.877
	(0.670)	(0.738)	(0.777)
$Monitoring_{ijt-1}$			-5.319***
			(1.645)
$New\_Design_{jt} * \hat{Monitoring}_{ijt-1}$			9.341***
			(1.820)
$Penalties_{ijt-1}$			-0.000+
			(0.000)
$New\_Design_{jt} * \hat{Penalties}_{ijt-1}$			0.001***
			(0.000)
Nb_Offers_{jt}	0.129	0.075	0.070
	(0.133)	(0.161)	(0.163)
$Nb\_Accommodations_{jt}$	-0.002***	-0.001	-0.002+
	(0.001)	(0.001)	(0.001)
$Constant_{ijt}$	88.272***	87.457***	90.936***
	(1.471)	(2.376)	(2.823)
	$Observe_{ijt}$	$Observe_{ijt}$	$Observe_{ijt}$
$Decision_t$	0.007	-0.011	0.019
	(0.091)	(0.091)	(0.097)
$Monitoring\_Others_{ijt-1}$	2.616***	2.681***	2.769***
	(0.162)	(0.165)	(0.180)
New_Design_{jt}	0.371***	0.466***	0.641***
	(0.120)	(0.134)	(0.139)
$Monitoring_{ijt-1}$			2.305***
			(0.277)
$New\_Design_{jt} * \hat{Monitoring}_{ijt-1}$			-2.552***
			(0.335)
$Penalties_{ijt-1}$			-0.000
			(0.000)
$New\_Design_{jt} * \hat{Penalties}_{ijt-1}$			-0.000
			(0.000)
Nb_Offers_{jt}	-0.088***	-0.093***	-0.107***
	(0.023)	(0.029)	(0.031)
$Nb\_Accommodations_{jt}$	0.000	0.000	0.000
	(0.000)	(0.000)	(0.000)
$Constant_{ijt}$	-0.446	-0.494	-2.067***
	(0.326)	(0.497)	(0.545)
$Lambda$	-1.928**	-1.874**	-0.906
	(0.775)	(0.746)	(0.753)
Lot	×	×	×
Firm		×	×
N	1359	1359	1359

Significance levels: +0.15, * 0.10, ** 0.05, *** 0.01; Robust standard errors in parentheses.

Table 8: Number of offers and Relative bid

	Model 10	Model 11	Model 12	Model 13	Model 14	Model 15
	$Price_{ijt}$	$Price_{ijt}$	$Price_{ijt}$	$Price_{ijt}$	$Price_{ijt}$	$Price_{ijt}$
New_Design_{jt}	5.082***	3.708***	4.407***	2.203	5.579	0.253
	(0.955)	(0.987)	(0.982)	(2.898)	(3.898)	(3.459)
Nb_Offers_{jt}	-0.756***	-0.430***	-0.808***	-0.706***	-0.469**	-0.733***
	(0.167)	(0.153)	(0.191)	(0.181)	(0.198)	(0.190)
$Nb\_Accommodations_{jt}$	-0.001	-0.001	-0.001	-0.001	-0.001	-0.001
	(0.001)	(0.002)	(0.001)	(0.001)	(0.002)	(0.001)
$Price\_Index_t$				0.155	-0.098	0.226
				(0.128)	(0.169)	(0.178)
$Constant_{ijt}$	17.421***	16.560***	18.408***	1.196	26.947	-5.142
	(1.927)	(3.343)	(2.310)	(13.587)	(20.161)	(18.138)
Lot		×			×	
Firm			×			×
N	102	102	102	102	102	102
$Adj-R2$	0.33	0.49	0.36	0.34	0.49	0.38

Table 11: Matrix of correlations

	New_Design_{jt}	Nb_Offers_{jt}	$Price_{ijt}$	$Nb\_Accommodations_{jt}$
New_Design_{jt}	1	-	-	-
Nb_Offers_{jt}	0.63	1	-	-
$Price_{ijt}$	0.32	-0.15	1	-
$Nb\_Accommodations_{jt}$	0.01	- 0.05	-0.13	1