The Role of Reputation when Awarding Contracts: Empirical Evidence from a Vendor Rating System

Francesco Decarolis^{*}, Riccardo Pacini^{**} and Giancarlo Spagnolo^{***}

Abstract

Reputational incentives are a powerful mechanism to improve suppliers performance, so strong to possibly start to influence suppliers behavior even before they are put in place. This paper presents a real experiment that provides empirical evidence on the effect of announcing the use of past performance information when awarding a public procurement contract. Suppliers react improving actual performance of the contract. Explicitly linking past (actual) performance to actual (future) award may help to solve the moral hazard problem in public procurement. While this is allowed by the US legislation, EU public procurement directives, which place competition above all, forbid to use past performance information at the awarding stage. However, this experiment suggests that the gains from avoiding suppliers' moral hazard when executing the contract may be higher than those from enforcing competition always and everywhere.

Keywords: public procurement, award criteria, past performance, reputation, vendor rating.

PRELIMINARY AND INCOMPLETE

^{*} Boston University, <u>fdc@bu.edu</u>.

^{**} Agenzia del Demanio, riccardo.pacini01@agenziademanio.it.

^{***} SITE-Stockholm School of Economics, DEF "Tor Vergata" and CEPR, <u>Giancarlo.Spagnolo@hhs.se</u>.

1. Introduction

This paper presents a real experiment on the role of long-term relationships (relational contracts) and reputational mechanisms in public procurement, taking into account crucial forces, like supplier competition, entry, buyer's discretion and the regulatory framework. Public procurement has become a large part of the world economy, amounting alone to over 15% of GDP in most advanced countries. For a number of different reasons, from poor/costly contract enforcement to the complexity of many goods and services, court-enforced contracts are often not sufficient to achieve an effective governance of public procurement exchange. Moreover, it may happen that even public procurement contracts are often not enforced. For example, some years ago there was an in depth inquiry in how public buyers manage the procurement contract auctioned off by Consip, the Italian Public Procurement Agency. A specialized audit firm collected information on the execution of a sample of these procurement contracts between 2005 and 2008 for a total of 4457 audits. The audits recorded whether the contractor violated contractual terms (technical and quality characteristics of the goods/services, timing of delivery and installation, accounting standards, after-sale support) and whether a penalty has been enforced in case of violation of one of the terms of contract for which a penalty is required. Descriptive statistics in Table 1 indicate that the percentage of contracts in which an infringement has been detected is relatively high (about 36%), of which 53% are identified as major non-conformities. However, the enforcement of penalties, the main contractual remedy, is dramatically low, only 3.4% of the major nonconformities.

	No.	Penalties	% of penalties
Non-conformities, of which	1614	63	3,90%
Major	848	29	3,42%
Minor	137	4	2,92%
Other nonconformities (not clearly identified)	629	30	4,77%
Conformity	2843		
Total	4457	63	1,41%

|--|

Corruption could of course be one of the reasons why contracts are not enforced in public procurement, the civil servant in charge may be bribed to accommodate lower performance without exercising remedies. We don't believe this to be the main explanation for these data, however, because in that case we would expect the civil servant in charge to hide the low performance level, not to record it in the books. Lack of enforcement of contractual remedies after low performance seems common in other countries where corruption is less of a problem than in Italy. Analogous anecdotal evidence on non-application of deduction exists for large procurement of complex services in the UK (HM Treasury, 2006) and for elderly care procurement contracts in Sweden, where in one third of the sample we analysed credible remedies are not even specified in the contract.

Since public procurement exchanges are rarely occasional, reputational forces may be exploited to improve on what formal contracting allows to achieve. Moreover, public procurement - besides

sharing the governance problems of private procurement - it also has to solve the major problem of public governance: how to keep public buyers accountable in the absence of market pressures and with the many layers of agency shielding them from tax payers' control. The interaction between this regulation and the governance of quality in procurement transactions is all but trivial.

The debate on public procurement regulation is particularly intense in Europe at the moment (see the EU GREEN PAPER 2011), where the revision of the 2004 Directives 17 and 18, which coordinate public procurement in all EU countries, has just been completed and the new directives (2014 Directives 23, 24 and 25) have come into force. However, there is a lively debate also in the US, in particular on how much discretion should be left to public buyers in the attempt to reduce transaction costs (see e.g. Yukins 2008) and on whether the use of reputational indicators based on past performance encouraged by the Federal Acquisition Regulations reduce the ability of new contractors to enter the market.¹

In private procurement reputational considerations are very important, whether they are informal and subjective or formalized in a feedback mechanism/Vendor Rating system (e.g. Bannerjee and Duflo, 2000). There are several reasons why complementing explicit contracts with reputational mechanisms based on ex post evaluations of contractor performance may improve the governance of procurement transactions. These are linked to both the inability of explicit contracts to describe (or the courts system to verify) important aspects of the procurement transactions at reasonable cost, but also to the high costs of enforcing explicit contracts through litigation. Many important quality aspects of supplied goods and services, particularly of more complex and valuable ones, are either very difficult to appropriately specify in an explicit contract in a practical and cost effective way or they are impossible to observe or to properly evaluate ex post for a third party that could enforce the contract, like a court or an arbitrator. Even when a certain qualitative dimension or supplier choice could be specified contractually and verified by the court, the cost of enforcing the contractual remedies through litigation and the effect that this may have on the continuing of buyer supplier relationship may often prevent an effective contractual governance.² (Macauley offer a very nice discussion of the latter costs).

In private procurement past performance indicators affect the selection of suppliers and their behaviour because buyers can act upon past performance, refraining from selecting suppliers with a poor track record and favouring those with a good one. In public procurement buyers' this type of 'discretion' is typically limited. The need to prevent favouritism and corruption led lawmakers around the world to ensure that open and transparent auctions where bidders have equal treatment (even when they have a very different track record) are used as often as possible. Open competition is not only seen as an instrument to achieve efficiency and value for taxpayer money, but also to keep public buyers accountable by limiting their discretion in the allocation of public funds.³

¹ GAO-12-102R, October 18, 2011. The relationship between reputation and entry will be a central theme of this essay. ² Macaulay (1963) classic study discusses extensively the latter problem and report a purchasing manager saying: "One doesn't run to lawyers if he wants to stay in business because one must behave decently" (p. 61). On the often very high costs of contract enforcement see the discussion in Iossa and Spagnolo (2011) and references therein.

³ Another way by which lawmakers limit civil servants' discretion is constraining 'discretionary' payments, i.e. monetary transfers not based on observable but non-contractible tasks. Public buyers then tend to recover their discretion – for the good or for the bad - at the contract management/enforcement stage; see Iossa and Spagnolo (2011) for an analysis of discretional contract enforcement.

This attempt to reduce discretion led in many countries to a two stage contractor selection process where a qualification stage that excludes firms without the basic ability to supply is followed by an awarding stage in which only the bids are evaluated, with no reference to the characteristic of the bidder. This amounted (almost) to a ban on reputation, as exclusion from the bidding stage is justified only for extremely poor past performance.

That limiting discretion to ensure public buyers' accountability comes at the possibly large cost of not allowing to use reputational forces to complement incomplete procurement contracts was stress for example by Kelman (1990). A recent study by Bandiera, Pratt and Valletti (2009), exploiting the introduction of a central procurement agency in Italy as a policy experiment, showed that accountability gains may also be small. They show that semi autonomous public buyers (universities and health authorities), who are endowed with more discretion, are significantly more efficient and are not more corrupt than more rigidly regulated ones (central administration).

Kelman pushed for a deep reform of the US system when he was the head of public procurement during the Clinton administration. The reform pointed at reducing the rigidity of procurement procedures built in the Federal Acquisition Regulations to enable public buyers to use more flexible purchasing methods similar to private sector practices, among which that of placing a stronger weight on suppliers' past performance.⁴ Since the Federal Acquisitions Streamlining Act in 1994 US Federal Departments and Agencies are expected to record past contractors' performance evaluations and share them through common platforms for use in future contractor selection.

In the EU things developed rather differently, in the opposite direction for some member countries used to more flexible procurement regulation than prescribed by EU laws (see Gordon and Racca (2014)). The EU Procurement Directives that coordinate public procurement regulation in the various European states considerably limit the possibility use of past performance information in the process of selecting offers. This has been one of the features under broader attack during the recent consultation for the revision of the EU Directives.⁵ Curiously enough, current European regulation acknowledges the importance of reputation for some types of procurement. For example, the European Research Council (ERC) funds top researchers in Europe, selected through peer review, and the track record of the researchers is then the main awarding criterion. ERC funding is distributed almost only on reputation criteria in order to reach the best and the brightest. Other European instruments for the procurement of research, such as the FET-OPEN program, are instead based on a completely anonymous evaluation.⁶ Why these two instruments are managed in such opposite ways is not clear. This is not surprising: the relationship between reputational forces, competition, entry and supplied quality/innovation is not yet fully understood, in theory as in practice.

http://ec.europa.eu/internal_market/consultations/docs/2011/public_procurement/synthesis_document_en.pdf .

⁴ As in the case of independent central banks, maintaining accountability after an increase in public buyers' ex ante discretion (independence) requires more stringent ex post controls in terms of performance measurement and evaluation. A real of perceived lack of stronger ex post performance controls may be at the root of recent concerns that this process may have led to excessive discretion and poor accountability in US public procurement (e.g. Yukins 2008). ⁵ See the summary of the replies to the consultation at

⁶ Indeed, on the dedicated homepage of these programs one reads that: "The anonymity policy applied to short proposals has changed and is strictly applied. The part B of a short STREP proposal may not include the name of any organization involved in the consortium nor any other information that could identify an applicant. Furthermore, strictly no bibliographic references are permitted."

This paper suggests that reputational incentives may be very strong, able to greatly influence suppliers' behaviour already after a first generic announcement that past performance measures will be collected and used in the future for selection purposes.

2. The context of the experiment

The experiment relates to the introduction of a vendor rating system and the announcement on the use of the ratings to award new contracts by one of the largest public multi-utility companies listed on the Italian exchange ("the Firm"). The Firm operates in the sale and distribution of energy, water services and public lighting. In 2010 the Firm had a turnover of 3.6 billion of euro and produced 15.651 GWh of electricity, placing it as the sixth largest operator in Italy. In order to maintain an orderly functioning of its power grid, each year the Firm outsources works worth over 300 million euro. Since the Firm is controlled by a public administration, it has to apply the Italian Code of Public Contracts when selecting contractors and awarding contracts.⁷ Being a multi-utilities company, the Firm falls in the "special sectors" which enjoy some flexibility in applying the Code. The Firm then employs a system of suppliers qualification to pre-select vendors. Starting from the second semester of 2007, it introduced a system of vendor rating for its qualified suppliers, with the plan to use its ratings at the awarding stage of the procurement process. The idea was to include vendors past performance regarding quality and security of works performed within the awarding criteria. The plan to introduce such a mechanism was announced to contractors, gradually disclosing details on its functioning and timing, along five main announcement events. This gives the possibility to study the reaction of vendors to the system introduction announcement. Before analyzing and empirically test it, we describe the legal framework and constraints in the use of past performance information in Italian public procurement tender and discuss the design of the vendor rating system.

2.1 Legal limits to consider reputation when awarding a contract

The possibility to introduce reputational elements for the selection of contractors and the award of contracts in public procurement has received wide attention in both academic and jurisprudential studies. This is of particular significance in Europe, where "contracting authorities shall treat economic operators equally and non-discriminatorily and shall act in a transparent way"⁸ and competition is of primary importance.⁹ Indeed, the use of reputational indicators presents a trade-off between the need to reduce the adverse selection and the moral hazard problems and the enforcement of the competition principle.

⁷ The Code is the law that has implemented the European Union public procurement directives 17/2004 and 18/2004.

⁸ Art. 2 of DIRECTIVE 2004/18/EC OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL of 31 March 2004 on the coordination of procedures for the award of public works contracts, public supply contracts and public service contracts.

⁹ "Contracts should be awarded on the basis of objective criteria which ensure compliance with the principles of transparency, non-discrimination and equal treatment and which guarantee that tenders are assessed in conditions of effective competition" Recital n.46 of the Directive 18/2004. "Non-discriminatory criteria should be indicated which the contracting authorities may use when selecting Competitors and the means which economic operators may use to prove they have satisfied those criteria." Recital n. 39 dir 2004/18/EC.

Concerning the adverse selection, the public procurement legislation for special sectors (according to Directive 17/2004/CE) is less stringent, since it allows public buyers to institute their own qualification system or, in general, to select potential candidates to be awarded, on the basis of their technical and professional skills, discretionally chosen by the contracting authorities. The unique limit in the choice of such criteria is the objectivity: "...Contracting entities which select candidates for restricted or negotiated procedures shall do so according to objective rules and criteria which they have established and which are available to interested economic operators...".¹⁰ To this regard the use of reputation indicators may be fully exploited if built on a system based on measurable parameters, that is verifiable by third parties and agreed by the qualified contractors.

The problem arises in the awarding phase. Since the EU gives special prominence to the free and fair competition principle, the use of reputation as an award criteria in public procurement can constitute an unfair advantage for the incumbents and a not proportionate disadvantage for new entrants: a potential supplier with no past experience cannot enjoy any reputational premium with respect to preexistent competitors. This may reduce entry and competition and violate the general principle of equal treatment. In the phase of awarding contracts, the most economically advantageous tender¹¹ is the criteria which allow to consider other criteria than the price. To this regard, the European Court of Justice clearly stated that the awarding authorities, when evaluating quality with the most economically advantageous tender should consider the object of the tender and not the bidder's characteristics.¹² Also the Italian Public Procurement Authority reaffirmed the same principle.¹³ However, the Firm was experiencing poor performances from its qualified contractors and the inefficacy of penalties to enforce contract provisions on quality.¹⁴ This drove the Firm to introduce reputational criteria when awarding contracts, through exploiting the higher flexibility given by the Code to the contracting authorities belonging to "the special sectors" and awarding contracts under the EU thresholds.

2.2 Designing the vendor rating system

The Firm designed its vendor rating system for the procurement of works in the electricity sector. The system considers a set of 136 parameters linked to the stringent quality and safety regulation of this industry, according to which contractors performances are evaluated and an overall reputation index is calculated. These parameters were collected by a team of (rotating) auditors in a number on site visits. The score given to each parameters is equal to 1 if the value is "regular", to 0 if the value is "irregular" or "n/a" if not possible to be inspected. The set of 136 parameters is divided into two macro-types, Safety (51) and Quality (83), further sub-grouped according to 12 Safety and Quality dimensions (7 for Safety and 5 for Quality) as follow:

¹⁰ Art. 54 comma 2, Dir.17/2004/CE.

¹¹ For awarding criteria specification, see art. 53 Dir.2004/18/EC and art. 55 Dir. 2004/17/EC.

¹² See judgments in Causes C-488/01 or C-31/87.

¹³ Resolution n. 30 of 06/02/2007.

¹⁴ Some data and experiences show that penalties are not effective because they are not even applied: a study conducted for Consip, the Italian public procurement agency, on a sample procurement contracts on goods and services, demonstrated that penalties were applied just in the 3.7% of the eligible cases (Albano, Dini, Spagnolo [2008]).

Table 2

Туре	Type Dimension	
	Equipment and machinery	5
Safety	Documentation	9
	Works execution	8
	Personnel	4
	Works site regularity	10
	Works site safety	10
	Works site controls	5
	Works on joints	19
Quality	Customer relationship mgnt	3
	Air works	25
	Underground works	25
	Works on transformer station	13

Before computing a unique reputation aggregated index, each parameter was associated with a relative weight, ranging from 2 to 10. The reputation index (RI) is then calculated as a flat weighted average mean across a predefined time span, according to the following formula:

$$\mathrm{RI} = \frac{\sum_{i=1}^{m} \sum_{j=1}^{n} v_{ij} p_j}{\sum_{j=1}^{n} p_j} \tag{1}$$

with v_{ij} indicating the score obtained in each of the *n* parameter over all the *m* audits considered, p_j the weight attached. Hence, the reputation index can range from 0 to 1 and be calculated for the overall experiment period, for specific periods, group of contractors, single contractors, Safety and Quality dimensions or mixed criteria.

As we said, the award criteria most suitable to include elements other than price is the most economically advantageous tender, for which the general scoring rule (S) is:

$$\mathbf{S} = \sum_{c=1}^{s} \mathbf{w}_{c} f_{c} \tag{2}$$

with f_c being the criteria to score the specific element of the bid, usually giving a score between 0 and 1, and w_c the weight to that element as defined in the tender document. Since usually $\sum_{c=1}^{s} W_c = 100$, S can range from 0 to 100. In this experiment, the scoring rule for each bidding contractor was announced to be:

$$S = w_d D + w_r R I \tag{3}$$

where w_d was the weight attributed to the discount offered, *D* the discount offered, w_r the weight attributed to the reputation index and RI the reputation index, with $w_d + w_r = 100$, to substitute the current lowest price award criteria. The bidding contractor obtaining the highest S would be the winner of the specific tender.

2.3 Timing of implementation

The Firm defined the parameters in April 2007 and started conducting the audits in October 2007. On the 20th December 2007, the Firm announced to all qualified contractors the rationale behind the audits, namely the introduction of the vendor rating system and the plan to use it in the awarding of contracts in the future. Then there were other four events, on the 4th April 2008, the 10th July 2008, the 21st October 2008 and the 16th January 2009, at which the Firm gave updates on the functioning of the vendor rating system and the results of the audits in terms of RI for each contractors and impact on awarding contracts if the modified awarding rule were in place.

3. Empirical Analysis

The analysis concerns all inspections carried since October 2007 to November 2009 and the reactions by vendors in their performance to the various announcements given since the introduction of the vendor rating.

There are two main datasets. The first one is a panel datset containg all the outcomes of the inspections that were performed during the life of 187 distinct contracts. These inspections occurred in the period between the 16th October 2007 and the 19th November 2009. They involved 45 different contractors, 1,952 works sites and they were carried out over the above mentioned 136 parameters that were checked for a total of 64,537 times throughout the sample period.

The second dataset is a corss section of auctions reporting information about the type of contract, its awarding date, the identity of the winner and its price, along with other contract-specific information. We generated this dataset by combining data given to us by the Firm and data concerning similar auctions held by other entities. The auctions in these data range between 2004 and 2010 and a few basic statistics for these data are reported below, separating auctions held before/after the announcemment of 12/20/2007 and auctions helb by the firm or by other entities.

3.1 Descriptive statistics

We begin our descriptive analysis from the panel dataset of inspections. The left hand side of Figure 1 shows the monthly distribution of the 64,537 observations throughout the sample period (see the green bars) and the progress of the reputation index (RI) calculated both on a monthly and a cumulative basis (the black and gray line respectively) on all the observations with respect to the announcements. The red line shows a significant positive trend in the monthly RI. The right hand side of Figure 1 shows the evolution of RI calculated on the observations relating the Safety and Quality dimensions separately. Figure 2 shows the same monthly distribution of the 64,537 observations in Figure 1 distinguishing the parameters regular (in green) from the irregular ones (in red), and the number of audits/inspections carried out each month. Figure 3 shows the distributions of parameters inspected grouped per the 12 dimensions. The parameters response related to the 7 Safety dimensions are the ones most inspected (55,050 times), while those ones related to the 5 Quality dimensions were inspected 9,487 times. This may have a relation with the stronger response from contractors to improve the performance on Safety as shown in Figure 1. Figure 4 shows the progress of the reputation index (RI) calculated on a monthly basis for each of the seven most audited Safety and Quality dimensions.

Figure 1



Legend: The black line shows the average score calculated on all parameters inspected in the month of reference. The grey line shows the cumulated average score calculated on all parameters inspected until month of reference. The red line is the trend calculated out of the black line. The green bars are the total number of parameters checked throughout the month of reference. The vertical dashed line identifies each announcement date.

Figure 2



Figure 3



Figure 4



Figure 5 shows the distributions of parameters inspected per the 45 contractors. The most inspected contractor was G that had 6,510 parameters inspected over 183 audits. Seven contractors received more than 100 audits and 19 contractors had more than 1,000 parameters inspected. Twelve contractors received less than 10 audits and 11 contractors had less than 200 parameters inspected. Maintaining the same order, Figure 6 gives the corresponding number of contracts awarded by each of the 45 contractors: 9 contractors were awarded 10 or more contracts (with AL as the most awarded contractor with 23 contracts), 10 contractors between 4 and 8 contracts, 15 contractors between 2 or 3 contracts and the remaining 11 contractors only 1 contract. Figure 7 plots the progress of the reputation index (RI), calculated on a monthly basis, for the 4 different grouping according to the number of contracts awarded, showing a common positive trend across the four groups. Figure 8 shows the monthly trend of the reputation index (RI) for each of the 9 contractors belonging to the first group of the most awarded contractors.

Figure 5



Figure 6



Figure 7



Figure 8



Finally, Figure 9 summarizes graphically what will be the most interesting aspect of the auction dataset. The figure shows in the left panel a scatter plot of the winning discounts for all of the auctions held by the Firm. One immediately notices that the distribution of the discuonts seems fairly similar during each subinterval in the period 2005 to 2010. However, since the very last months of 2010 there appears to be a concentration of low discounts, the right plot of the figure focuses on the auctions held after 2005 and before 2010. It also reports a regression line and its confidence interval. The figure confirms that discounts do not seem to significantly change around the announcement date.





3.2 Empirical Analysis: Quality Performance and the Announcements

The previous descriptive analysis offers a clear visual description of two main facts: i) quality inceases after the first announcement and ii) winning discounts did not become lower after the first announcement. This section intends to establish through a more rigorous empirical strategy that both these facts are indeed a feature of the experiment that we study. To achieve this goal, we first focus on the analysis of the inspections data and carry out two simple statistical tests: i) a series of t-test on the five announcements relating the upcoming introduction of the vendor rating at the awarding stage on the reputation score and auction discount time series and ii) a probit regression on the single parameters scores. In subsequent versions of this study we plan to evaluate the robustness of our results to different methods to identify structural breaks in time series data.

The first of the two tests is presented in Table 3. This table shows the t-test results for the five announcements carried out over 25 time series relating to the reputation for safety, for quality, for the 12 specific dimensions, for the 9 most awarded and audited contractors and for the auction discounts, for a total of 120 t-test (in other 5 cases it was not possible to run the test for lack of observations). Each test is run between the group of all parameters inspected before the specific announcement and the group of all parameters inspected after the specific announcement. In 106 cases the reputation before the announcement resulted significantly lower than after, while only in 3 cases significantly higher. In the remaining 11 cases the reputation was not significantly different.

Table	3
-------	---

T-test							
		Announcements					
	1st 20 dec 07	2nd 4 apr 08	3rd 10 jul 08	4th 21 oct 08	5th 16 jan 09		
Overall	+	+	+	+	+		
Main Dimensions							
Safety (S)	+	+	+	+	+		
Quality (Q)	+	+	+	+	+		
Specific Safety and Quality Dimensions							
Q: Works on joints	-	0	+	+	+		
Q: Customer relationship management	-	-	+	+	+		
Q: Air works	n.a.	+	+	0	0		
Q: Underground works	+	+	+	+	+		
Q: Transformer station works	0	0	0	0	0		
S: Personnel	+	+	+	+	+		
S: Works site regularity	+	+	+	+	+		
S: Works site safety	+	+	+	+	+		
S: Works site controls	n.a.	n.a.	n.a.	0	0		
S: Equipment and machinery	+	+	+	+	+		
S: Documentation	+	+	+	+	+		
S: Works execution	+	+	+	+	+		
Most awarded and audited firms							
AL	+	0	+	+	+		
D	+	+	+	+	+		
G	+	+	+	+	+		
L	n.a.	+	+	+	+		
Μ	+	+	+	+	+		
U	+	+	+	+	+		
			1		1		
Auction discount	+	+	+	+	+		

Legend:

+/- = score before the announcement is significantly (5%) lower/higher than after

0 = score not significantly different

n.a. = not available

Each test is run between the group of all parameters inspected before the specific announcement and the group of all parameters inspected after the specific announcement, relating to the category reported in each row.

The second type of test is a probit regression on the single parameter scores. Table 4 reports the probit regression results for the following equation:

 $Pr(RegularParameter) = \Phi[\alpha^*weight + \beta^*announc_{(1to5)} + \gamma^*dim_{(1to11)} + \delta^*group_{(1to4)}]$ (4)

where Φ is the normal cdf, *RegularParameter* is the binary score, 0 or 1, taken by the single parameter audited, *weight* is the weight associated with the parameter, *announc*_(1to5) are five dummies which take 0 or 1 if the parameter is audited before or after the specific announcement of reference, $dim_{(1to12)}$ are twelve dummies which take 0 or 1 depending on the parameter belonging to one of the twelve specific safety or quality dimension, and $group_{(1to3)}$ are three dummies which take 0 or 1 depending on the parameter referring to one of the four grouping of firms according to the number of contracts awarded (see Figure 7).

Table 4

Probit regression				Numb	er of obs	=	64537
Log pseudolikelihood = -35158.621					<pre>ch12(21) > chi2</pre>	=	0.0000
parametro	Coef.	Robust Std. Err.	z	P> z	[95%	Conf.	Interval]
weight announc1 announc2 announc3 announc4 announc5 dim1 dim2 dim3 dim4 dim5 dim6 dim7 dim8 dim9 dim10 dim11 dim12 group1 group2 group3	.0085191 .4449855 .1315237 .4341951 .218222 .059171 .3535574 7230065 2012378 .6826454 .5061796 1.217221 5762149 1.840798 7522982 9344627 538382 604407 0034356 0193238 0727709	.0031194 .0347929 .0215823 .0193842 .0223362 .018998 .0536464 .0493303 .0527846 .0755167 .2343647 .2937219 .0479084 .3615874 .0550745 .0500092 .0524455 .0868998 .0313599 .0328206 .0337456	2.73 12.79 6.09 22.40 9.77 3.11 6.59 -14.66 -3.81 9.04 2.16 4.14 -12.03 5.09 -13.66 -18.69 -10.27 -6.96 -0.11 -0.59 -2.16	$\begin{array}{c} 0.006\\ 0.000\\ 0.000\\ 0.000\\ 0.000\\ 0.002\\ 0.000\\ 0.$.0024 .3767 .0892 .3962 .1744 .0219 .2484 819 3046 .5346 .0468 .6415 6701 1.132 8602 -1.032 6411 7747 0648 083 083 1389	052 926 232 027 438 356 124 692 937 353 364 137 099 422 479 733 274 999 651 112	.0146329 .5131783 .1738242 .4721875 .2620002 .0964065 .4587024 6263209 0977819 .8306555 .965526 1.792905 4823161 2.549496 6443542 8364465 4355907 4340866 .0580288 .0450034 0066307

Legend:

dim1	= S, Equipment and machinery	aroun1	- 9 contractors awarded with 10 or more contracts
dim2	= S, Documentation	groupi	
dim3	= S, Works execution	aroun?	- 10 contractors awarded between 4 and 8 contracts
dim9	= S, Personnel	groupz	
dim10	= S, Works site regularity	aroun?	- 15 contractors awarded 2 or 2 contracts
dim11	= S, Works site safety	groups	- 15 contractors awarded 2 or 5 contracts
dim12	= S, Works site controls		
dim4	= Q, Works on joints	announc1	= 20/12/2007
dim5	= Q, Customer relationship mgnt	announc2	= 04/04/2008
dim6	= Q, Air works	announc3	= 10/07/2008
dim7	= Q, Underground works	announc4	= 21/10/2008
dim8	= Q, Transformer station works	announc5	= 16/01/2009

Overall, the probit regression confirms the positive and significant effect of all the five announcements, along with the significant presence of differences between the safety and quality dimensions and between the groups of firms. In the final version of the paper we will explore alternative methods to study structural breaks in the data to further confirm the large and positive quality change associated with the announcements.

3.3 Empirical Analysis: Winning Discounts and the Announcements

Although the previous section revealed that quality improved, the extent to which this is desirable for the firm depends on how its procurement cost changes. In this regard, it is essential to evaluate whether lack of any significant changes in winning discounts after the announcements shown in Figure 9 stands a more careful scrutiny. Our approach to study this problem exploits the time variation observed in the data to develop a difference-in-differences (DD) strategy. More precisely, we seek to estimate the following regression model:

W.Discount_{ist}= $a_s+b_t+cX_{ist}+\beta$ (Treatment)+ ε_{ist}

where the index *i* indicates the auction, *s* the entity awarding the contract and *t* the year. The coefficient of interest is β , the effect on the winning discount of a dummy variable (indicated as Treatment) equal to one for the contracts awarded by the Firm after its 12/20/2007 announcement, conditional on fixed effects for the entity awarding the contract (a_s) and time (b_t) and on other covariates (X). The dependent variables considered are the winning rebate, the number of bids and a dummy for whether the winner is from the same region of the PA.

		Statisti	cs for the	e Firm's Aucti			
_	Period: 2005-2007			Period: 2008-2010			
	Mean	SD	Ν	Mean	SD	Ν	
Discount	23.07	11.59	205	20.21	10.79	117	
Res. Price	796,632	910,199	205	709,835	702,661	117	
Duration	400.11	169.03	191	440.66	222.258	27	
Central Italy	100%			100%			
Munic. Firm	100%			100%			
	Stat	istics for the	Union of	the Control C	Froups 2 and 3		
_	Period: 2005-2007			Period: 2008-2010			
	Mean	SD	Ν	Mean	SD	Ν	
Discount	22.41	10.42	915	22.4	11.76	476	
Res. Price	773,185	990,148	915	812,699	1,094,818	476	
Duration	378.55	345.49	795	437.33	654.04	372	
Central Italy	89%			85%			
Munic. Firm	69%			64%			

Table 5

Table 5 reports summary statistics dividing the auction dataset into four subsets: auctions held before the announcement (in the period 2004-2007) or after the announcements (in the period 2008-2010); and held either by the Firm or by other entities. The top panel of Table 5 shows that the average winning discount in the Firm's auctions slightly declines, but in a non statistically significant way, in the period after the announcement. However, this could be driven, for instance, by some external factor like a change in input prices occurring at the same time of the announcement and confounding the interpretation of the lack of significant mean changes. The advantage of the DD strategy is that though the use of an appropriate control it can be established what is the causal effect of the announcement on the change of winning discounts for the Firm.

In an ideal dataset, we would observe that, given a group of similar auctions, a randomly chosen subset is run under the treatment, while the rest remains under the status quo. This allows to interpret the estimate of β as the causal effect of the introduction of the rule changes. Our dataset differs from this scenario because the only treated auctions are those held by the Firm after its first announcement. However, the DD method ensures that a causal interpretation of β is possible if we can find a control composed of auctions that would have expressed the same outcomes of the treated ones absent the treatment. Our solution to the choice of the control group consists in considering three different sets of control groups obtained by the combination of different sets of entities awarding the contracts. In particular, our first control group consists of auctions awarded by all the entities that are subject, like the Firm, to the public procurement code and that entail performing a job of the same type of those in the treated auctions. The second group is identical to the previous one, with the exception that only entities located in the four Italian central regions are used. This restriction likely makes the auctions in these group closer to that of the Firm that also only operates in this area and, hence, faces similar conditions. The third control group consists of all the entities that are municipal firms providing electricity-related services, like the Firm.

			Contro	1 2: PA	Cont	rol 3:
	Control 1	: All PA	Central	Regions	Municip	al Firms
Treatment	0.296	0.342	1.422	1.495	1.728	1.688
	(.835)	(.830)	(1.972)	(1.995)	(1.712)	(1.720)
Obs.	6,270	6,270	1,334	1,334	647	647
R ²	0.70	0.71	0.52	0.53	0.26	0.26

Table 6

Robust standard errors in parentheses

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

Table 6 reports the DD estimates obtained using the three control groups and two model specifications. The first two columns refer to control group 1, the next two to control group 2 and the latter two to control group 3. For each one of these pairs, the first results concern a model that in addition to the *Treatment* dummy includes exclusively fixed effects for the year and the identity of the entity awarding the contract. The other model, reported in the second column of each pair, adds to the previous covariates the auction reserve price, and dummy for the macro region where the entity is located (north, center, south/islands) and its type (regional government, provincial

government, municipality or else). Despite the three control groups are rather different both in terms of their size and in terms of their composition, they all deliver a quite consistent finding confirming the absence of any statistically significant change in the winning discount. Moreover, the magnitude of the coefficients is rather small suggesting that indeed we are estimating a zero effect and the lack of significance is not driven by noise in the data.

In the future versions of this paper, we plan to extend this analysis to strengthen further our findings. For instance, one limitation of our approach is that since we will be considering cases involving changes exclusively one entity, the Firm, the effect that we estimate could be confounded by any other change happening to the Firm at the same time of the policy change. This type of problem often affects DD studies. However, since the number of entities in each one of our control groups is large, then it is possible to correct for it using methods proposed by Conley and Taber (2011). This approach is followed, for instance, in Decarolis (2014) in a closely related setting.

Finally, we plan to study more in depth the relationship between the discount offered in the tender procedure and the performance (measured by RI, the reputation index) resulting during the execution of the contract by the awarded contractor. To this purpose we computed RI over all the parameters audited for each contract and we calculated the correlation with the winning discount. From Figure 10, which plots the discount/reputation combinations, it appears that there is no correlation between discounts and quality/safety of works. The correlation is equal to 0.098 and not statistically significant, while using a simple linear regression the R^2 is less than 1% (see Figure 10). Apparently, this means that improvements in quality and safety has come as a free lunch to the Firm. Analyzing through which mechanism this has been possible will be our next goal.



Figure 10

Legend: Each blue dot identify one contract, to which they are associated the auction discount offered by the winning contractor (on the x axis) and the score calculated on all parameters inspected throughout the

same contract life (on the y axis). The red line is the linear regression line calculated out of the 120 auction discount / reputational score combinations, where the reputational score is the dependent variable and the auction discount is the independent variable. The auction regression coefficient, 0.1855, is not statistically significant (p value = 0.29).

4. Discussion and conclusions

The required performance from contractors could in principle be governed contractually, but contract enforcement is very slow and costly in Italy. Moreover, managers in charge of contract management found it difficult to exercise explicit contractual sanctions without worsening the prospects of long-term cooperation with suppliers. Hence other mechanisms are required.

This paper has studied the merits of using a vendor rating system data at the awarding stage as a mechanism to spur higher efforts from contractors when executing the contract on the basis of an experiment run by an Italian corporation listed on the Italian Exchanges. The results of the experiment has shown a strong significant upward trend in Safety and Quality performance after the firm has announced the future use of this reputation mechanism to award contracts and after all the subsequent announcements. Interestingly, this was also true for the auction discounts series. However, there was no correlation at all between auction discounts and Safety/Quality improvements, the latter apparently occurring at no costs.

The vendors' average score emerging in the first audit was equal to 0.29, while the last audit we analyzed presented a score equal to 0.81. The first time vendors heard that they would be evaluated for their "quality", they immediately improved their performance, causing a break in the series. To some extent this may recall the well-known *Hawthorne effect*.¹⁵ However, contrary to the Hawthorne effect, the improvement was not short-lived, even if we consider that the contractors could have stop to trust the Firm for the delayed implementation of the new awarding criteria and that it was easier for contractors to improve their score when the starting point was lower than later when the marginal cost to improve became higher. Indeed, some Safety and Quality parameters compliance requires very low investment costs: in particular for the Safety type, it happens that the regularity just calls for a greater level of care in running the works. For instance, the requirement for road signal in the vicinity of the work site (a parameter of the Safety type) is quite costless, and the same happens with other similar parameters. This may also explain part of the lack of correlation between the Safety and Quality improvements and the discounts.

Our results confirm that reputation can represent an important mechanism to exploit for buyers who have to rely on a relatively stable set of contractors. However, the experiment results are different from those ones from the traditional models showing that a reputational equilibrium can be sustained if the level of price is greater than the competitive one and such to guarantee a level of future rents greater than the immediate profits obtained from a cheating behavior. Indeed the Safety and Quality constant improvements occurred over the 2-year sample period were not correlated at

¹⁵ The Hawthorne effect is a form of reactivity whereby subjects improve or modify an aspect of their behavior being experimentally measured simply in response to the fact that they know they are being studied, not in response to any particular experimental manipulation.

all with the prices offered at the awarding stage. These results confirm the view in Bandiera, Barankay and Rasul (2011) that experiments conducted within firms can have a powerful role to shed light on important aspects of firm behavior.

Furthermore, although several different mechanisms might explain why the increased quality and safety achieved was not reflected into higher prices, it is interesting to mention that the explanation offered by the management of the Firm is that most of the gains came from costless improvements in management practices within contractors. Thanks to new data on management practices collected in the last ten years through World Management Survey¹⁶, there has been an increased attention to the role of management in explaining productivity differences (Bloom et Al. (2014)). In this respect our results are interesting because they provide a clear example of how (the announcemnt of) new procurement rules can trigger an improvement in management.

Finally, once the merits of this kind of reputation mechanism to improve contractors' are proven, many aspects remain open and give room for future researches: how to discipline the rating for new entrants, to structure the weights in the awarding criteria, and to opportunely choose the optimal "memory" of the indicator (i.e. how going backward for the calculation of RI).

¹⁶ See: <u>http://worldmanagementsurvey.org</u>

References

Akerlof G.A., The market for lemons: quality uncertainty and the market mechanism, (1970), Quarterly Journal of Economics, 84

Allen F., Reputation and product quality, (1984), RAND Journal of Economics, 15

Atkeson A, Hellwig C. Ordonez G., Optimal Regulation in the Presence of the reputation Concerns, (2010), Mimeo.

Bandiera, O., I. Barankay, and I. Rasul, (2011). "Field Experiments with Firms." *Journal of Economic Perspectives*, 25(3): 63-82.

Bar-Isaac, H., Imperfect Commitment and reputation Commitment, (2005), Economics letters, 89

Bar-Isaac H., Tadelis S., Seller reputation, (2008), Foundations and Trends in microeconomics, 4

Bloom, N., R. Lemos, R. Sadun, D. Scur, J. Van Reenen (2014). The New Empirical Economics of Management, CEPR Discussion Paper No. 10013.

Board S., Vehn M., Reputation for quality, (2010), mimeo

Calzolari G., Spagnolo G., Relational contracts and competitive screening, (2009), CEPR discussion paper

Chou S.Y., Shen C.Y., Chang Y.H., Vendor selection in a modified re-buy situation using a strategy aligned fuzzy approach, (2007), Int. J. Production Research, 45

Conley, T. G., C. R. Taber, (2011). "Inference with Difference in Differences with a Small Number of Policy Changes." The Review of Economics and Statistics, 93(1): 113–125.

Cook W., Johnston A., Evaluating suppliers of complex systems: a multiple criteria approach, (1992), J. Operational Research Society, 43

Dellarocas C., Dini F., Spagnolo G., Designing reputation mechanisms, (2007), in N. Dimitri, G. Piga and G. Spagnolo: Handbook of procurement, Cambridge University Press

Dellarocas C., Reputation mechanisms, (2005), Handbook on Information Systems and Economics. Elsevier Publishing

Detoni A., Nassimbeni G., A method for the evaluation of suppliers' co-design effort, (2001), Int. J. Production Economics, 72

Doni N., The importance of reputation in awarding public contracts, (2006), Annals of Public and Cooperative Economics, 77

Ekmekci M., Sustainable reputation with rating systems, (2008), mimeo NYU

Ely J., Fudenberg D., Levine D.K., When is reputation bad?, (2008), Games and Economic Behavior, 63

Ely J., Valimaki J., Bad reputation, (2003), Quarterly Journal of Economics, 118

Fishman A., Rob R., Is bigger better? Investing in reputation, (2002), J. Political Economy, 113

Gordon S., (2008) Supplier Evaluation and Performance Excellence: A Guide to Meaningful Metrics and Successful Results, J.Ross Publishing

Gordon D., Racca G., (2014), Integrity Challenges in the EU and U.S. Procurement Systems, Ius Publicum, 1.

Heikkila J., From supply to demand chain management: efficiency and customer satisfaction, (2002), J. Operations Management, 20

Holmstrom B., Managerial incentive problems: a Dynamic persective, (1999), The Review of Economic Studies, 66

Horner J., Reputation and competition, (2002), American Economic Review, 92

Iossa E., Ray P., Building reputation for contract renewal: implications for performance dynamics and contract duration, (2009), Discussion paper

Kahraman C., Ruan D., Dogan I., Fuzzy group decision-making for facility location selection, (2003) Information Sciences, 157

Kangas A., Kangas J., Pykalainen J., Outranking methods as tools in strategic natural resources planning, (2001), Silva Fennica, 35

Kaplan R.S., Norton D., The balanced scorecard - measures that drive performance, (1992), Harvard Business Review

Kleijnen J.P.C., Smits M.T., Performance matrix in supply chain management, (2003), J. Operational Research Society, 54

Klein B., Leffler K., The role of market forces in assuring contractual performance, (1981), J. Political Economy, 89

Kranton R., Competition and the incentive to produce high quality, (2003), Economica, 70

Land C.K., Lovell C.A., Thore S., Chance-constrained data envelopment analysis, (1993), Managerial and Decision Economics, 14

Liu F.F., Hai H.L., The voting Analytic hierarchy process method for selecting supplier, (2005), Int. J. Production Economics, 97

Liu Q., Skrzypacz A., Limited records and reputation, (2009), Discussion paper

Mailath G.J., Samuelson L., Who wants a good reputation?, (2001), Review of Economic Studies, 68

Marchi G., Lenti L., La valutazione nei processi di piano - strumenti di trasformazione urbana, (2003), Angeli

Roodhooft F., Konings J., Vendor selection and evaluation: an activity based costing approach, (1996), European Journal of Operational research, 96

Saaty T.L., Decision making with the analytic hierarchy process, (2008), Int. J. Services Sciences, 1

Schmitz J., Platts K.W., Supply logistics performance measurement: indications from a study in the automotive industry, (2004), Int. J. Production Economics, 89

Shapiro C., Premiums for high quality products as returns to reputations, (1983), Quarterly Journal of Economics, 98

Stiglitz, J., (1987), Competition and the Number of Firms in a Market: Are Duopolies more competitive than atomistic markets, Journal of Political Economy, 95

Tadelis S. (1999)What's in a Name? Reputation as a Tradeable Asset, American Economic Review June 1999, 89.

Talluri S., Narasimhaen R., Nair A., Vendor performance with supply risk: a chance-constrained DEA approach, (2006), Int. J. Production Economics, 100

Trent R.J., Monczka R.M., Purchasing and supply management: trends and changes throughout the 1990s,(1998), Int. J. Purchasing and materials

Vanin P., Competition, reputation and compliance, (2009)

Wang J., Li H., Developing a decision model for supplier selection, (2006), Chinese Journal of Management Science

Weber C.A., Current J.R., Benton W.C., Vendor selection criteria and methods, (1991), E. J. Operational Research, 50

Weber C.A., Desai A., Determination of paths to vendor market efficiency using parallel coordinates representation: a negotiation tool for buyers,(1996), E.J. Operational Research, 90

Wu D., Olson T.L., A comparison of stochastic dominance and stochastic DEA for vendor evaluation, (2007), Int. J. Production Research

Yahya S., Kingsman B., Vendor rating for an entrepreneur development program: a case study using analytic hierarchy process method, (1999), J. Operational Research Society, 50