Department of Economics and Finance.

Chair of Advanced Microeconomics.

Scoring auctions in public procurement: Endogenous vs Exogenous quality provision.

Supervisor

Gian Luigi Albano

Co-supervisor

Giovanni Ponti

Candidate

Federico Cesarini

ID number**: 660701**

Academic Year 2014/2015

Scoring auctions in public procurement:

Endogenous vs Exogenous quality provision.

Abstract

This paper reports evidences of two first score experimental procurements with scoring auction awarding criterion; one in which subjects need to endogenously determine both the price and the technical attributes of their tender, and the other in which they are only free to bid on the price dimension, while the quality is them exogenously assigned before bidding. The experimental design of both the games consists of two different treatments of interest: one in which quality matters more for the selection of the winning subject and the other in which the price bid instead is the determinant factor. Margins extracted over production costs are here analyzed and compared, from a behavioral perspective, across experiments and against equilibrium margins provided by the theoretical models. Main findings are the following. Higher margins were obtained in exogenously quality environments but they here even resulted more heavily influenced by tender's characteristics choosed by the buyer. This directly depends on the selection each mechanism is able to carry out. When compared against optimal margins both the experiments provide the same results. Subjects almost always obtained lower margins when competition is on the quality dimension, while always higher when competition transfers on the price dimension.

Table of contents

1. Introduction 4
   1.1 Structure of the paper 7
2. Background 8
   2.1 Motivation 11
   2.2 Some useful remarks 12
3. Experimental design 14
   3.1 Sessions 14
   3.2 Matching groups 15
   3.3 Experimental procurement auction forms 15
      3.3.1 Procurement game with exogenous quality provision 16
      3.3.2 Procurement game with endogenous quality provision 17
   3.4 Monetary payments 18
4. Experimental results 19
   4.1 Empirical quality choices and deviations from optimality 19
   4.2 Markup distribution across experiments 21
   4.3 How does markup relate with each player's type? 23
      4.3.1 Experimental game with endogenous quality provision 24
      4.3.2 Experimental game with exogenous quality provision 25
      4.3.3 Some remarks 26
   4.4 Which is the markup across sessions? 27
   4.5 Regression analysis 29
      4.5.1 Experimental game with endogenous quality provision 31
      4.5.2 Experimental game with exogenous quality provision 33
      4.5.3 Margins' comparison 36
   4.6 Deviations from optimality 39
      4.6.1 Experimental game with endogenous quality provision 39
      4.6.2 Experimental game with exogenous quality provision 41
5. Conclusions 43

References 45

Appendix A 47

Appendix B 52

SECTION 1

1. Introduction

A huge amount of resources is allocated every year by the mean of public procurement.

The latter, which is commonly defined as the process of purchasing by Governments, state-owned entrerprises or local entitites, of goods, services and work, accounts for a large amount of the public expenditures, ranging from 18-20% in countries like Portugal and Greece, to 35-40% as in Japan or Korea. An increasing number of sectors, such as waste management, construction, energy, telecommunications, transport, social and sanitary protection, and the provision of financial and education services, are indeed characterized by having public authorities acting as the standard buyers. Whilst OECD countries register almost 10-16% of their GDP spent in public procurement, the so called developing countries manage to do even better, reaching the threshold of 20-22%.Thus, it not only means that government procurement produces a substantial effect on the aggregated demand of any domestic economy, but also that it accounts for a sizable portion of the taxpayers' money. For this reason it's a process which needs particular attention by policymakers and researchers, who are continuosly asked to look for optimal and efficient mechanisms to be able to safeguard the public interest and maximize the social welfare.

Although for years the literature of auction theory described a procurement contract as a game of incomplete information *based on price competition only*, leaving every tender to comply with some minimal quality requisites, technical specifications of the goods or services provided are becoming more and more important for the final decision of the buyers. In the contract for national defense acquisitions or for the allocation of public works for instance, the price to be paid is just one of the multiple attributes to be considered. In other words, we are progressively passing from the better known "*Lowest Price*" evaluation criterion (LP), according to which only the price bid makes a difference in the selection of the winning tender, to the "*Most Economically Advantageous Tender*" evaluation criterion (MEAT), which asks the buyer to assess and rank the presented bids in relation to multiple attributes, such as technical characteristics, delivery terms, after-sale services, sustainable aspects, innovative and environmental characteristics, and so on. Towards this aim, the most commonly used procedures in real practices include: (i) menù auctions, (ii) scoring auctions, (iii) beauty constests and (iv) bargaining, according to which the buyer selects some potential sellers and negotiate with them on all possible dimensions to be evaluated. The technique we are focusing on in this paper is that of a scoring auction, which according to Milgrom (2004) is particularly interesting since, under some circumstances, may let procurers to obtain more valuable contracts at the same time promoting partecipation by more bidders and not lowering sellers' expected profits, with so increasing value for both parties. In his opinion, bidders must always prefer scoring auctions procedure rather than price-only auctions, since in thie latter case they do not manage to exploit their peculiar advantages and characteristics. The robustness of these findings was confirmed by Asker and Cantillon(2008), who showed how scoring auctions let both buyers and sellers to achieve higher expected payoffs. Other several theoretical works focused recently their attention on multidimensional scoring auctions. The latter, which is commonly considered a two-stage evaluation process consisting of a *technical evaluation* - the technical merits of the tenders - and of a *financial evaluation* - which relates to the price only - let the contracting agency to award the contract to the tender that reached the highest overall score, which is just defined as a weighted average of both technical and fincancial criteria, with weights well defined before bidding.

Albeit Thiel (1988), Cripps and Ireland (1994) firstly approached the problem of multidimensionality by assuming respectively (i) that procurers do not value any savings and choose on the basis of preset budget – publicly communicated to all bidders – and (ii) quality thresholds to be passed to enter the auction, the first work which really came close to find a solution for the design of multidimensional scoring auctions was that one realized by Che(1993). The author, who considered a two-dimensional model in which every bid consists of a pair of price and quality assessed according to a scoring rule converting each pair into a single number, discovered how first and second score auctions are able to implement an optimal and efficient mechanism when the scoring function is quasi-linear in price. In his simple model, each bidder differs from the others just for his marginal costs of improving quality, which, following Harsanyi(1967), is randomly drawn by the "Nature" from a well-defined ex-ante probability distribution, which is common knowledge among all bidders. In other words, each firm is assumed to bear a cost which is independent from all the others. The latter is one of the most criticised point of Che's findings, since in real procurement practices it's reasonable to assume that the costs of the partecipating firms must have some in common and so cannot be completely independent from each other. Branco(1997), for example, tried to extend Che's model by analyzing the impact of costs correlation on the design of multidimensional mechanisms. His findings were surpising, since none of the properties regarding the mechanisms studied by Che resulted to be valid in his setting. In particular, when costs are correlated, any mechanism based on a single stage auction - and so even that one discovered by Che – is not optimal anymore. Only two-stage mechanisms where in the first stage bids are evaluated according to a scoring rule and in the second the winner of the first round bargains with the contracting authority, seem to be able to reach optimality.

On the other hand, other works contested the mono dimensionality of each bidder's type of Che's model, since in reality firms can differ not only in the marginal costs, but even in their fixed costs. Asker and Cantillon (2010), on the basis of several previous works on the same topic, extended the analysis of optimal procurement mechanisms to the more elaborate environment in which private information is multidimensional, although independent among bidders. The latter assumption, togheter with a scoring function quasi-linear in price, results here to be crucial to reduce the dimensionality of the relevant private information to one – i.e. each bidder's "pseudotype"- which is necessary to characterize the equilibrium of this particular scoring auction. In addition, the authors managed to find, as already done by Che in his simplified environment, an extension of the famous revenue equivalence theorem[1], showing that buyers are indifferent between first score, second score and ascending or descending scoring auctions when suppliers are symmetric in their pseudotypes. Although it's really hard to characterize the optimal mechanism when private information is multidimensional, the authors managed to show that scoring auctions yield the same performance of the optimal one, obtained by numerical simulation. Nishimura (2012)[2] extended the results of Che (1993) to an environment in which private information is mono-dimensional but the quality is here represented by a vector of multiple attributes and found that an optimal mechanism is achievable even in this case, as well as the buyer describe scoring rules which must be additively separable in some or all quality attributes. All of these works relate to procurement contract where quasi-linear scoring rules - as for instance the weighted criteria - have been adopted. Hanazono et al (2011)[3] tried to remove this strong assumption by assuming "price-quality ratio" scoring rules (PQR) and M-dimensional private information for each supplier. According to the authors, the intense competition provided by this setting induces undercutting the provision of technical attributes, which turns out to reduce the price asked in equilibrium. None of the properties regarding optimality and revenue equivalence resulted here to be valid.

Unfortunately, although interesting and well structured, none of the above mentioned theoretical works manage to capture all the difficulties encountered by practitioners in real procurement practices and to provide them useful and easily implementable paths to follow. This is mainly addressed to the impossibility for partecipating firms to understand the excessive sophisticated optimal strategies discovered by theorists and to the procurers' difficulties in

[1] See Myerson(1981), Riley and Samuelson(1981)

[2] Nishimura (2012), "Optimal design of scoring auction with multidimensional quality"

[3] M. Hanazono, J. Nakabayashi, and M. Tsuruoka. "A theoretical analysis of scoring auctions with a general form of scoring rules" (2011).

having full knowledge of the environment they deal with. As a consequence, laboratory experiments are progressively becoming always more useful tools to interpret the effects that a mechanism or another could have. Bichler (2000)[4] and Chen-Ritzo et al (2005),[5]for instance, explored experimentally a multiattribute reverse auction in which price and two technical attributes of the goods to be sold have been considered. Despite the fact that the laboratory setting still remains a simpler and less sophisticated environment than real world, and even if lots of doubts regarding the generalization of laboratory results are recurring among practitioners, could sometimes reveal very useful to study in the lab a new procedure before introducing it in the field. The aim of this paper follows exactly this intuition and consists of providing empirical evidence regarding the bidding behaviour of subjects involved in a public procurement tender with scoring auctions procedure. Specifically, we compared two different experimental works - whereof our investigation is an extention –, one in which suppliers, endowed with iid quality before offering, bid only in the price dimension, and the other in which instead they need to endogenously determine both the quality and the price to be assigned to the good or service they want to sell throught the procurement auction. In particular, we are interested in understanding, form a behavioral perspective, how the subjects' behavior changes when they are asked to deal with more tasks with respect to the case in which they have to provide a bid in the price dimension only. Margins obtained by winning subjects are here compared.

1.1 Structure of the paper

The remainder of this paper is organized as follows. *Section 2* presents a detailed analysis of the two experimental works at the base of our investigation, highlighting their main research questions and findings. *Section 3* deals with the methodology used to reach our aim, and so reports an exhaustive description of the experimental design of both procuement games here considered. Every key parameter and tratment of the games is here well explained. Main results are instead reported in *Section 4,* by the mean of graphs and regressions. It's the section in which the main reflections are drwan. *Section 5* concludes and put the roots for possible improvements and open problems.

[4] M. Bichler, " an experimental analysis of multi-attribute auctions". Decision Support System, 29:249-268, (2000)

[5] "Better, Faster, Cheaper: An Experimental Analysis of a Multiattribute Reverse Auction Mechanism with Restricted Information Feedback" (2005), Ching-Hua Chen-Ritzo, Terry P. Harrison, Anthony M. Kwasnica, Douglas J. Thomas

SECTION 2

2. Background

The framework upon which my research is based, mainly consists of two different experimental papers on scoring auctions in public procurement – that one of Albano, Ponti et al (2015)[6] and the Master degree Thesis of M. Rosi (2015),[7] a LUISS Guido Carli student – whose data I exploited to better understand the main characteristics of the empirical behavior of the subjects involved across the two different settings. For this reason, before explaining the logic which brought me to compare the two, it is necessary to briefly describe their structures, – illustrated in detail in *Section 3* – their research questions and their main findings.

Both the projects deal with an experimental game in which the partecipating subjects are randomly divided in groups of five and engage in a reverse auction by choosing a tailor-made tender with the aim of having their contract awarded, but they asymptotically diverge in the way partecipants formulate their precise bid. Specifically, in the first experiment subjects are endowed with a randomly drawn i.i.d. "quality"[8] – privatelly announced to every subject before bidding - and have just to present a price bid to enter the tender; in the second instead, - after endowed with a parameter representing the marginal cost of a given increase in quality (θ in the experiment) - they need to select both the price and the peculiar technical attributes – summarized for simplicity in the "quality" variable - of the goods or services they want to sell within the procurement auction. As a result, this translates into a significant difference between suppliers' cost functions in the two different procurement auction games. While in fact, in the experiment in which the quality has been made exogenous, subjects with the same quality – i.e. the same type – must bear the same cost of producing such a quality, in the other game instead, even if several subjects decided to choose the same quality, they could exhibit different costs too, depending on whether they even share the same type (θ), or not.[9]

Nevertheless, albeit in a different manner, subjects are assigned in both cases a final score, which is just a function of the quality – picked by the players in one case and randomly assigned to them in the other - and the price bid. Undoubtedly the quality and the price bid induce

---

[6] " Absolute vs. Relative Scoring in experimental procurements" (2015), G. Albano, G. Ponti, M. Sparro, R. Di Paolo and A. Paolillo.

[7] " Procurement Auctions with endogenous investment: a theoretical and empirical investigation" (2015), M. Rosi, LUISS Guido Carli, Rome.

[8] The authors made the individual quality to vary every round, so to be able to draw a complete bidding function per scoring rule for every partecipant. The approach followed, was that one of Grimm et al (2009). Following the approach discovered by Harsanyi, – to transform a problem of incomplete information into another one of imperfect information - the quality is just the player's type, drawn with a common distribution by the Nature.

[9] The cost function depends on the quality only, in the first case, while both on the quality and on (θ) in the other.

respectively a positive and a negative effect on the final score and the tender with the highest overall score results with having the contract awarded. Moreover, either projects tried to evaluate the bidding behavior of the subjects, by proposing them two different scoring rules during the whole experiment, in order to understand their attitudes in the two different frames. The first one is characterized by making the overall score to be more influenced by the technical one, – i.e. the quality choice or an increasing function of it - whilst the other gives an higher weight to the financial score, with so enhancing competition by forcing the sellers to lower their price bid to win the auction. Whereas the above distinction is the only one really relevant for the second project, which even considers two different specifications of the technical function, – one more concave than the other - the first experimental game is characterized by exhibiting a two between-subjects treatments, referring another important distinction of the scoring rules : the well known *Absolute* vs *Relative* scoring rules[10]. The latters, which in some ways link the final outcomes of all players making the score of one tender dependent on some or all the remainings, are used by the experimenters to calibrate the absolute scores, which instead assign to each tender and for every possible dimension to be assessed, a final score which is completely unrelated to that of all the other presented tenders. The difference above, obviously deals only with the financial score, the technical score being just equal to the quality bid, which is given for the entire experiment. For the purpose of my analysis would be interesting to consider only the Absolute Scoring Rules and, in the specific, just the linear one - i.e. a scoring funcion which is linear in the price variable - in order to make possible the comparison between the projects.[11]

In the light of what I explained, it cannot be sursping that the theoretical frameworks of the two experimental papers are really different from each other, although they share some common basic characteristics. In particular, the principal theoretical background upon which the second experiment is based, is a well known paper by Che (1993),[12] - a milestone for scoring auction literature – which establishes a strategic behavior for the subjects involved in a procurement auction when both price and quality matter. The unique Nash Equilibrium of the Bayesian-Game is characterized by an optimal quality bid – whose choice is independent from the choice of the score according to the author – and an optimal price bid, which consists of two different parts: (i) the cost the seller has to bear when producing the optimal quality level and (ii) the margin – i.e. the mark-up – he manages to obtain over the cost incurred, after the trade has been made. Even if the author solves the above mentioned problem for three different auction schemes – the first

[10] In this direction, the authors seem to follow-up to a recent published papers by Chang et al. (2013) - “ An investigation of the Average Bid Mechanism for Procurement Auctions, mimeo”- in which the authors made a comparison between Lowest Bid mechanisms and Average Bid mechanisms, for the case of common value reverse auctions only.

[11] The first project in fact is characterized by a financial score which is linear in the price bid.

[12] “Design Competition through Multi-Dimensional Auctions”, Che(1993).

score, second score and the second preferred offer – the paper I'm focusing on, lives off the backs of the first score procurement auctions framework only.

As far as the first project instead is concerned, and taking in consideration only the case in which the absolute linear scoring rules have been selected, the theoretical bakground was discovered by the authors, who follow the approach used by Che[13] to find the Bayesian Nash Equilibrium of the Game (BNE). In particular, they showed that is possibile to rank the partecipants according to their "maximum–achievable-score[14]" for every quality level (q) rather then according to the quality itself, and so to reduce the problem to a classical mono-dimensional first price auction (FPA). As a consequence, the BNE happens to be characterized by an optimal price bid which is a function of the financial score, the quality level (q) and the parameter (γ) representing the weight assigned to the technical and the financial score in the final overall outcome.[15]

The comparison between the theoretical optimal behavior of the players and its empirical counterpart, is at the heart of both the two procurement empirical investigations here considered. More in the specific, the authors' aim is in both cases that one of understanding if the theoretical predictions are in line enough to what the subjects really do in practice, in order to learn how and why the not-fully-rational subjects' strategic choices differ from that of the fully rational economic agents of the theoretical world. Furthermore, some considerations about the effects of the parameter expressing the weight of the technical and the financial score - and so about the choice of the overall scoring rule adopted, which because of procurers' poor commitment power ends up with revealing their true preferences - on the price bid, in one case, and on both the price and the quality bid, in the other, have been made. Both papers even focus on analysing the impact that players' type[16] can have on some interesting variables, such as the real chance of victory, the aggressiveness of their bids[17], their profits and so on , in order to understand if the most efficient subjects[18] managed to be rewarded by the mechanism design choosed by the contracting agencies for every possibile specification of the parameters involved. Thus, since one

[13] Both the projects deal only with risk-neutral bidders and transform the original problem into another one to facilitate the analytical construction of the symmetric best responses of the players and so the BNE.

[14] The maximum score a bidder with type α can obtain is computed just imposing that the price bid must be equal to the cost incurred, so that the final profit in case of winning the tender is reduced to zero.

[15] Indeed, the authors did not use the price bid, but the discount the seller makes on the base price – set here equal to one, with no loss of generality. Moreover the fincial score computed at the equilibrium, which is a function of the score picked for every player's type, results to be the standard optimal bid strategy of a FPA. See "Optimal Auctions", J. Riley, W. Samuelson (1981) for a better understanding.

[16] In one project consists of the personal marginal cost of increasing quality (θ_i), while in the other is just the quality level assigned to the subjects (q_i)

[17] Indeed, only the paper by Albano et al (2015) take in consideration the aggressiveness of the price bid and its evolution towards the experiment, since the other project has been made static, in the sense that subjects cannot see their outcomes till the end of the experiment and so cannot learn which is the right combination of price and quality which could let them to have the contract awarded.

[18] i.e. the subjects endowed with the highest quality level or with the lowest possible marginal cost.

of the main objectives pursued by the buyer is that of implementing, a priori, with the mechanism choosed, an efficient allocation of resources, the authors tried in both cases to understand which the most efficient session happens to be.

What they discovered in the two different settings is in some way in line to what previous investigations on similar topics carried out, albeit, at least for the project in which the quality has been endogenized, some specific findings appear to be counterintuitive and not in accordance with the theoretical predictions. Indeed, whereas the experiment realized by Albano et al (2015) exhibits empirical price bids which follow, at least in their shape, the theoretical ones, the other experiment's results are characterized by a quality bid sufficiently close to what the theory pronosticates, but a price bid which instead is completely unrelated to its optimal counterpart, even in the shape.[19]In addition, both the projects showed how the strong competition between the players force them to lower their price bid when the final overall score is more influenced by the financial one, and how in the latter case was difficult to select a winner of the auction which results even to be the most efficient. In particular, the experiment at the root of the Master degree thesis showed (i) a 10% higher level of efficiency when the competition shifts on quality rather than when is focused on price, and (ii) that less efficient subjects were less able to achieve positive profits from the trade when the first setting occurs;[20] the other experimental paper instead, after pointed out that the choice of an absolute or relative scoring rule does not significantly provide neither a greater efficiency nor meaningful changes in the strategic behavior of the players,[21] highlighted how the quality – i.e. the player's type – results to have always a significant impact on the efficient winner, expecially when it has an higher weight than the price in the final score - in which case only a very high quality can let the more efficient subjects to win.

2.1 Motivation

On the basis of all the above, we can easily argue that, although with some important differences, the two experimental projects go in the direction of improving some interesting aspects of the scoring auction literature. In some sense, the paper carried out by Albano et al refers to a less

---

[19] The author addresses this problem mainly to the monetary incentives given to the subjects, which were too low to ensure they will choose the right strategy.

[20] When quality matters more than the price, the data suggest that subjects endowed with the smallest cost parameter took advantages of that by making higher porfits, which obviously decrease as the cost parameter grows.

[21] The results are not surprising, since follow in some way what was already discovered by Chang et al (2013): i.e. subjects seem not to be sensitive to the different specifications of the rebate bid – average and maximum rebate bid.

sofisticated procurement auction scheme, since firms are here left just to elaborate a price bid, leaving out every possible consideration about the quality choice, which, as well as the technology level achieved by R&D investments and needed to obtain such quality, is them randomly assigned by the "Nature" and so exogenously treated. Thus, the authors simplified the real procurement practice to make the subjects involved to better formulate an optimal price bid, without beeing diverted by the choice of the quality. On the contrary, in the second experiment, each firm, albeit endowed with an exogenous technology level which cannot be decided through the game the authors proposed,[22]and as in real practices happens, needs to endogenously determine the quality to be attributed to the good or service to be sold within the procurement auction. Thus, could be interesting, from a behavioural perspective, to compare the two, both from the point of view of the buyer and the seller, to understand if their choices could be affected by this peculiar feature. In particular, would be useful to comprehend if subjects behave in the same way when they have more tasks to deal with and so if they manage to be as strategically optimal as in the case in which they are only left to decide which price to apply to a good of a given quality. Although the theoretical findings discovered by Che (1993) state that the optimal choice of quality should be completely independent of that of the score, – which means that at least theoretically the outcomes of the two experiments should not diverge – the aim of our investigation is that of analysing if some empirical results as the price bid, the markup, the overall score, the cost savings obtained by the buyer, and so on, exhibit substantial differences across the two different settings.

2.2 Some useful remarks

The two experimental projects presented in *Section 2.1*, were conducted separately and independently. Thus, they were not made up to facilitate the comparison between them.

Indeed, as will be described in detail in the following Section, although they share the common framework of a standard first score procurement auction, they are based on two experimental games which differ in their structure and for the specification of some relevant parameters involved. Therefore, could be here useful to explain their main discrepancies and to clarify, where necessary, the way in which we tried to make this comparison possible.

As already explained, the two experimental games differ firstly in the signals the subjects recieve by the Nature. In one case it corresponds to the quality level, whilst in the other it is just a parameter standing for a personal component of the variable production costs, (θ). Consequently,

[22] i.e. none of the players can decide, during the game, how to invest in R&D to obtain a more sophisticated technology than the competitors. They can just decide the quality of the good or service they want to sell through the procurement auction.

the two settings should obviously differ in terms of the costs subjects have to bear when producing a certain good or service of a given quality. Indeed, in the first game the cost function results to be driven by the quality - exogenously assigned to them - only, while in the other, it's a function of the quality and the cost parameter with which each player is endowed. As a result, whereas subjects with the same quality level should always share the same production costs, in the first setting, this will not happen in the second one, unless partecipants share the same cost parameter (θ) too. Thus, would be useless to compare directly suppliers' profit across the games, being better to analyse the percentage margin they were able to obtain, which has no dimension. Moreover, the two procurement auction schemes slightly differ even in terms of the overall score the auctioneer have to choose to rank all players' bids. In both cases, it results to be a function of the technical score, the financial score and the parameter expressing the weight assigned to them (γ). Starting from the financial score, it results to be a linear function of the rebate the sucjects choosed on a base price, in one game, while a linear function of the price they asked for a good of a given quality level, in the other. Therefore, their comparison is an easy task, after we transform the price variable into the rebate one, or viceversa, according to the following relation : $rebate = (1 - price)$.[23] In addition, the two procurement games are based on a different characterization of the technical score. While in the game conducted by Albano, Ponti et al, it is just a linear function of the quality assigned to the subjects at the beginning of each period and remains the same for the whole game, in the other one, it can assume two different forms – one more concave than the other – according to which treatment has been considered. Thus, we can consider the latters just as two independent observations to be compared with that of the other experiment. Finally, even the weights assigned to the technical and financial score are different across the two experimental settings. The parameter $\gamma \in G = \{\frac{1}{3}, \frac{2}{3}\}$ in the second game, while $\gamma \in \{\frac{1}{4}, \frac{3}{4}\}$ in the other one. Therefore, for the purpose of our investigation, we need just to consider when $\gamma > 0.5$ and $\gamma < 0.5$, in order to estabilish if the final overall score is more influenced by the technical or the financial score.

[23] Recall that the base price is set equal to 1, with no loss of generality, for the entire experiment.

SECTION 3

3. Experimental design

In this section, we discuss in detail about the experimental structure of both the two different projects. As concerns the experimental game in which the quality provision has been made exogenous, instead of presenting the whole structure carried out by the authors, we consider the first treatment only, which consists just of the setting where the Absolute Linear Financial Score (ALS) has been adopted. Recall in fact that the other experimental game exhibit a linear financial score, so for the purpose of my analysis, it's completely worthless to consider all three remainings scoring rules contemplated by the authors. In this way, and following considerations made in *Section 2.1*, we can make possible the comparison between the two experiments.

3.1 Sessions

Both the two experimental games were fully computerized and programmed using the well-know software Z-tree,[24]and consist of four different sessions. Whilst the game in which the quality has been made exogenous was carried o at the Centro di Economia Sperimentale a Roma Est (CESARE), at Luiss University in Rome, the game in which the quality bid has to be determined by the subjects was conducted at the University of Alicante, Spain.

As concerns the first one, a total of 90 undergraduate students of Luiss University were engaged using the ORSEE recruiting system (Greiner, 2004); 80 students, both graduate and undergraduate, were instead recruited for the second experimental game. No particular bias towards any possible department of both Universities has been committed and all sessions were in both cases "gender balanced", in the sense that the subjects pool was approximately made up of the same number of males and females. Before the games started, as usual for Experimental Economics, the instructions – which include illustrative examples and tests of understanding[25] -, even delivered to every subject in a paper format, were read aloud by an experimenter,[26] in order to avoid that the subjects could not have been confident enough with the rules of the game, with so making their choices meaningless. As commonly happens in experimental studies, partecipants are here provided with examples of bid evaluation and profit calculations, with the aim of making

[24] Fischbacher (2007)

[25] See for a better comprehension, "Experimental Economics", D. Davis, C. Holt (1993), Princeton University Press

[26] In some cases, would be better that the instructions were read out by someone who is not particularly involved in the experiment, in order to be as objective as possible so to avoid the common Demand effect problem.

them to uderstand how the procurement game works. Following Kagel et al (1987), Battalio, Kogut and Meyer (1990), some dry runs with no money at stake were granted.

What really makes a difference between the two experimental games was the structure of the four above mentioned sessions. Each one is made up of two treatments - which refer to a difference in the concavity of the technical score - every one consisting of 11 different periods, in the experiment conducted at Luiss University; in the other game instead, each session is composed of 4 distinctive treatments, each consisting of 11 periods too, but representing the several scoring rules adopted. Again, we consider the first treatment only. All the discrepancies across the sessions of both the experimental projects will be discussed in detail in the remainder of this section.

At the end of every session of the two games, subjects were firstly asked to answer to some personal questions - which could help to understand if some characteristics of the players made them to make a choice or another - and, immediately after, were individually and privatelly paid according to a base fee plus their possible monetary winnings - explained in detail in Section 3.4 - obtained through the procurement auction game.

3.2 Matching groups

Subjects involved in the two experimental procurement games – 80 in one project and 90 in the other - were randomly divided in different groups, each of which represents an isolated "world" from the remainings. Thus, different groups never interact with each other towards the whole session. Specifically, as far as the experiment conducted at University of Alicante is considered, subjects are divided, each session, in 4 random groups of 5, with a fixed matching for the entire experiment – i.e. every group of five remains the same till the end of the game. As regards the other experiment instead, subjects were divided in 5 arbitrary groups of 5 in the first and the third session, while in 4 groups of 5 in the second and the fourth one. Moreover, as for the other experimental game, each group, after assembled, remains the same for the whole session. Therefore, every distinct group results to be an independent observation of the treatment considered.

3.3 Experimental procurement auction forms

In what follows, an accurate description of both the two experimental game structures has been presented. To better understand the main characteristics which typify one experiment from the other and to be sure that their main differences are transparent and clear enough, would be better to consider them separately.

3.3.1 Procurement game with exogenous quality provision

As already understood, the aim of this experimental game is that of simulating the standard first score procurement auction framework with adverse selection, when the quality is provided exogenously, while the price, required to produce such a quality, has to be determined by the partecipating subjects. In the specific, at the beginning of every period, each bidder $i \in I = \{1, \dots, n\}$, - with n=5 in the experiment, since every subject competes only with the other 4 subjects of his group and never interacts with all the others – after randomly sorted in a group of five, is endowed with a quality level, $q_i \in [0,1]$, of the good or service he pretends to sell through the procurement auction. The experimenters made the latter to assume only the 11 discrete values between 0 and 1.[27] The cost the subjects have to bear in order to produce such a quality, will be determined according to the following function:

$$C(q_i) = \frac{1}{4} + \frac{3}{4}(q_i)^2$$

As we can see, the total cost consists of (i) a fixed cost, completely unrelated to the quality, and (ii) a variable cost, which increases monotonically as the quality grows.

After observing that, each bidder i needs to select his own strategy, which is made up of a bidding function $\beta_i: [0,1] \to [0,1]$, mapping from the quality set to a discount on a predetermined base price, here set equal to 1 with no loss of generality.

At this point, the final outcome of each subject will be determined. The latter consist of a function – i.e. the scoring rule – which depends on the quality level, the rebate bid choosen by the player and the parameter γ standing for the weight associated to the financial and the technical score in the overall outcome, as shown below.

$$S_i(q, \beta, \gamma) = (1 - \gamma)t(q_i) + \gamma\sigma_i(\beta)$$

The final outcome is thus a weighted function of the technical score, which in the experiment is just equal to the quality level - t(q) $= q_i$ -, and a financial score, - $\sigma_i: \beta \to [0,1]$- which is the score associated with i's rebate. More in the specific, the Absolute Financial Linear score (ALS) has been used – i.e. $\sigma_i(\beta) = \beta_i$ -, which means that the score increases linearly in the rebate variable. The weight assigned to the former or the latter can here assume only two different values – i.e. $\gamma \in G = \{\frac{1}{3}, \frac{2}{3}\}$. In particular, it is equal to $\frac{1}{3}$ in the first and the third session, while equal to $\frac{2}{3}$ in the second and the fourth session.

[27] i.e. $q_i = \frac{k}{10}, where\ k = 0,1, \dots 10.$

After that each player i was assigned an overall score, he will receive a payoff which acts in accordance with the following rule:

$$\pi_i(\alpha,\beta,\gamma) = \begin{cases} \dfrac{1 - c(\alpha_i) - \beta_i}{n^*} & if\ s_i(.) = max_j(s_j(.)) \\ 0\ otherwise. \end{cases}$$

In other words, if player i wins the auction, he will be rewarded with a payoff corresponding to its profit obtained through the trade, while if he does not he will recieve a payoff of zero.[28]

3.3.2 Procurement game with endogenous quality provision

This experimental game was designed to mimic the framework of a first score procurement auction with adverse selection and moral hazard. The mechanism designer could in fact not only select a less efficient supplier – i.e. a firm with high marginal costs – but is even influenced by the action the subject carries out after selection – i.e. the effort to produce an high quality good or service. Differently from the other experiment, at the beginning of each of the eleven periods, the partecipating subjects, after sorted in groups of five, were randomly and individually assigned an exogenous parameter, $\theta \in \Theta = \{0.25, 0.3, 0.35, 0.4, \ldots, 0.75\}$, which affects the cost of increasing quality, they have to bear when producing goods or services they want to sell. The total costs incurred by type θ when producing a given quality level writes:

$$C(q,\theta) = \frac{1}{4} + \frac{3}{4}\theta q^2$$

After observing the parameter with which they were endowed, players must fristly select a quality bid, and, only after the total cost of producing it is computed, a price bid, which is constrained to be not lower than the total cost borne.[29] At this point, the software is able to rank every possible subjects according to the following scoring function, which exhibit just a slight change from that of the other experimental game:

[28] Notice that in case of a tie, the "breaking rule" estabilishes that the positive payoff obtained, must be split between the players who achieved the highest overall score.

[29] The bidding profiles of the partecipating subjects happen to be characterized by (i)quality bids, $\mathrm{q} = (\mathrm{q_i}) \in \mathrm{Q} = [0,1]^\mathrm{n}$, and (ii) price bids, $\mathrm{p} = (\mathrm{p_i}) \in \mathrm{P} = [0,1]^\mathrm{n}$.

$$S_i(q,p,\gamma,n) = (1-\gamma)t(q,n) - \gamma f(p)$$

As we can easily notice, the above scoring rule is just a function of the technical score, which in this case can assume two different forms depending on the parameter $n \in N = \{4, 8\}$, which makes the technical function more or less concave, - i.e. $t(q,n) = \sqrt[n]{q}$ – the financial score, which is just a linear function of the price bid – i.e. $f(p) = p$ – and the parameter γ representing the weight assigned to the financial score only. Differently from the other experiment, here $\gamma \in G = \{\frac{1}{4}, \frac{3}{4}\}$. As in the other instead, it is left free to vary only across sessions, while remains always constant within each session. Indeed, it's equal to $\frac{1}{4}$ in the first and the third session, while equal to $\frac{3}{4}$ in the second and fourth one. The parameter n, on the contrary, varies within each session depending on which phase is considered. Specifically, in the first and second session, $n = 4$ for the first phase while $n = 8$ for the second one; In the third and the fourth session instead, $n = 8$ for the first phase whilst $n = 4$ for the second one.

Again, at the end of each period every subject recieve a payoff which is equal to the profit he manages to obtain after the trade has been made if he wins the auction, and zero otherwise. In this case too if more than one player manages to obtain the highest final score, the payoff must be split between all players who won.

3.4 Monetary payments

Both the experimental games have the same rewarding method. At the end of every session, subjects are individually paid according to the outcome they manage to obtain in one randomly selected period. Because of this, subjects have the incentive to do their best all the periods they are asked to play; even when they are endowed with the highest θ- in the project with the endogenous quality provision - or with the lowest quality level - in the game with exogenous quality. In the first experiment they recieve 10 € just to partecipate to the game plus a variable amount which corresponds to the monetary payoff they were able to obtain in the selected period. In the other game instead, the partecipation fee was lower and equal to 5 €, while applies the same method for the variable monetary winnings.

SECTION 4

4. Experimental results

In this Section we report the main conclusions of our experimental work. In particular, before comparing the two different projects from the partecipants' point of view, it is worth exploring wheter the theoretical predictions regarding the quality bid characterized by Che(1993) – obviously only in the case in which it's endogenously determined by the players - are compatible with the empirical counterparts. The author, as already mentioned, showed how the optimal strategy in the quality dimension results to be completely unrelated to that of the score, since just obtained by maximizing the consumers' surplus with respect to the quality and imposing a price which exactly covers the cost incurred. If empirical quality bids happen to follow this intuition, it means that the outcomes of the two experiments, although suffering the relevant discrepancies in terms of their structures, should be at least consistent.

4.1 Empirical quality choices and deviations from optimality

With the help of the following graphs we put in comparison empirical quality choices and their equilibrium counterparts, obviously dealing only with the experimental game charcterized by an endogenous quality provision. Each figure plots optimal and observed quality bid against the cost paramenter affecting the marginal cost of increasing quality – i.e each player's "type", with which suppliers are endowed at the beginning of each round before bidding.

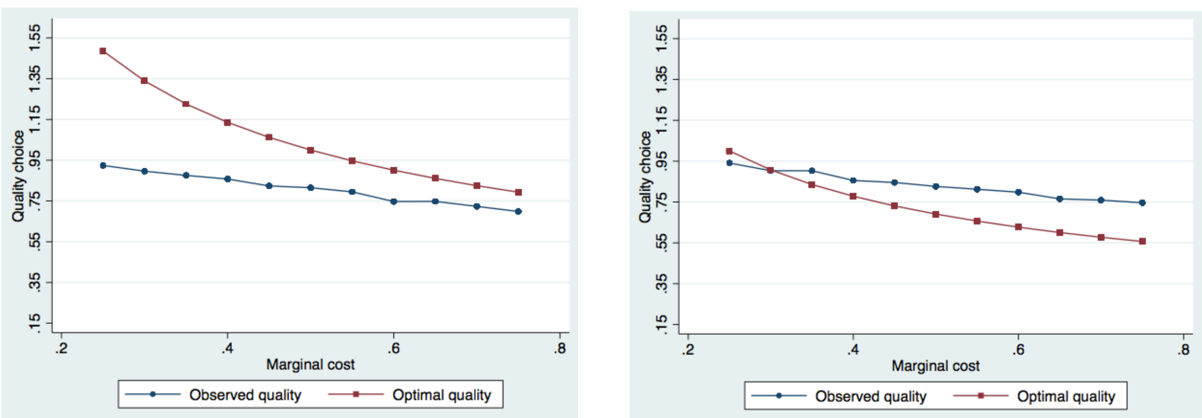


Figure 1: Optimal vs Observed quality choice. Gamma low.

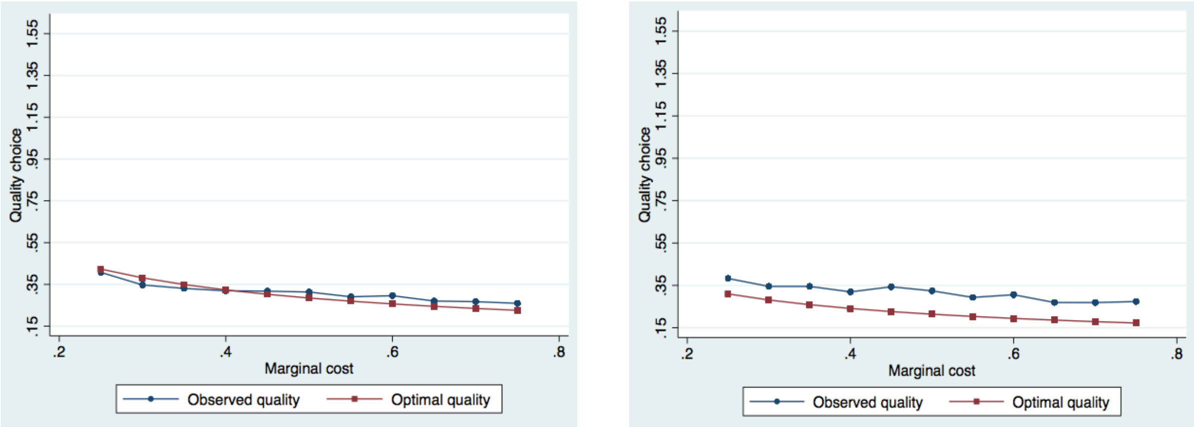


Figure 2: *Optimal vs Observed quality choice. Gamma high.*

Figure 1 reports average quality bid of suppliers with respect to that ones pronosticated by the Che's model, when the final overall score is more influenced by the technical one (i.e. $\gamma = \frac{1}{4}$). The left panel differs from the right one just in the specification of the technical score, which exhibits higher concavity in the latter panel. *Figure 2* instead, while representing the same comparison of the others, deal with a scoring function in which the financial score, which in the experiment is linear in price, has been assigned an higher weight (i.e. $\gamma = \frac{3}{4}$). In this case too, the right panel is characterized by having a more concave technical score than the left one. As we can see from the graphs, albeit subjects under or overbid most of the time with respect to their optimal strategies, they manage to follow the theoretical predictions in all the four settings here considered. More in the specific, the quality offered is always decreasing in suppliers' type – i.e. their marginal cost of increasing a given level of quality. This is in accordance with the theory and basic intuitions, since it's quite obvious that less efficient suppliers – i.e. firms whit higher marginal costs – are less able to offer very sophisticated goods or services unless by asking a very huge price to cover their costs. Even the distance between the two plots seems not to be large, expecially when price matters more than quality in the final overall score. Moreover, it becomes immediately evident the huge downward shift in the quality bid - under both two specifications of the technical score contempleted - when competition transfers on price, i.e when the financial score has an higher weight in the final overall outcome.

In the latter case in fact, an higher level of quality not only does not increase the probability of winning the tender, but can even lower it by forcing suppliers to increase the price asked, so to avoid to make very low profits.

4.2 Markup distribution across experiments

In this subsection we analyze the markup each winning subject obtains over the cost incurred when a given level of quality is produced, either when it's arbitrarly choosed by partecipants or when randomly assigned them by the "Nature". In each round of the games, at least one winner has been selected. For a better understanding we decided to report first the markup distribution for aggregate data – i.e. without differentiating across treatments.

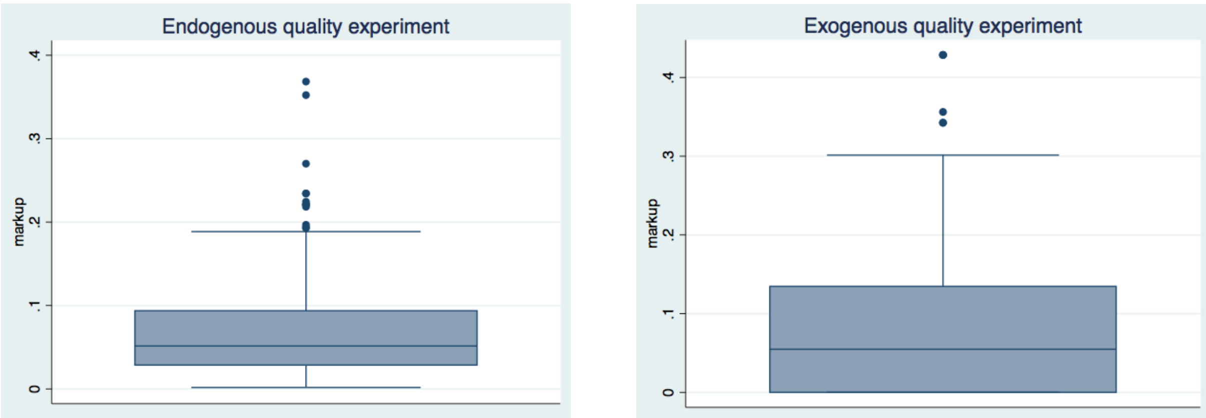


Figure 3: Markup distribution. Aggregate data

Box plots above give us a basic idea of the distribution of the markup each firm was able to obtain in the two different experiments, when dealing with aggregate data only. As we can easily notice, half the markups of the endogenous quality experiment are between 0.25 and 0.1 at the same time being really concentrated around the median; in the other experiment instead, not only the interquartile range is bigger than the other case – it starts from 0 and ends above 0.12 - but the distribution exhibits an upper adjacent striclty higher too. This result is surprising since seems that subjects when quality is exogenous wants to win just to win – i.e. even knowing their profit from the trade is zero. This cognitive bias could possibily be explained by the simple fact that in a private university, more than in a public one, competition among subjects is so harsh to force them to desire to always prevail over the others. Another possible interpretation could be that subjects may have suffered higher costs in the exogenous quality environment rather than in the other. The latter obviously reduce price choice possibilities with so lowering the margin extracted. All in all, the procurement auction in which the quality bid is exogenously treated seems to make subjects to extract higher margins than the other experiment. However recall that we are analyzing aggregate data, so could be now useful to understand the robustness of this result for each of the different settings provided by the two experimental games – i.e. the treatments of interest. In particular we analyze first the case in which the quality matters more than price in the overall scoring rule adopted – i.e. gamma is low – and then the other case – i.e. gamma high. For the experiment in which the quality is endogenous, one more distinction has been here considered. Recall in fact that, differently from the other procurement game, the technical score assumes here two different forms, one more concave than the other, and so we cannot avoid to see if this different specification affects the behavior of partecipating subjects.

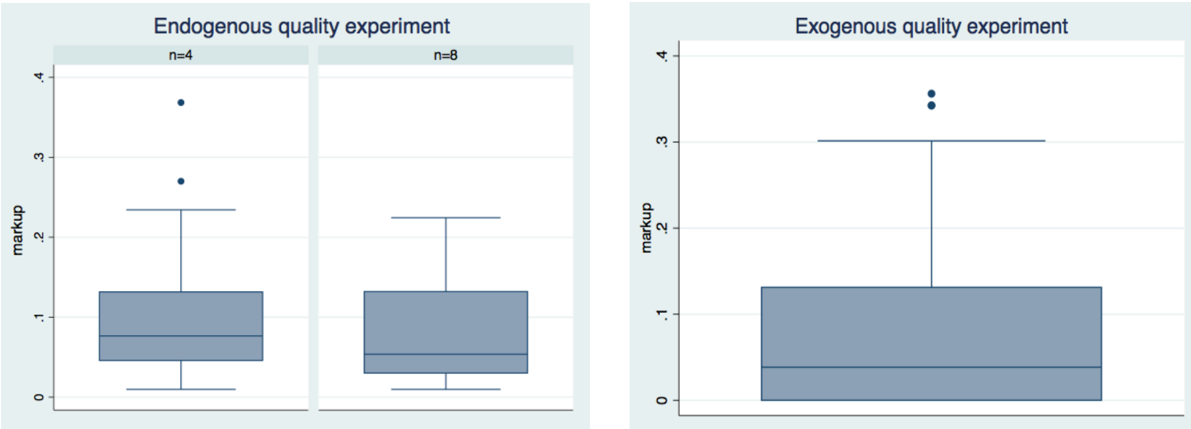


Figure 4: Markup distribution when quality matters more than price in buyer's preferences(γ is low)

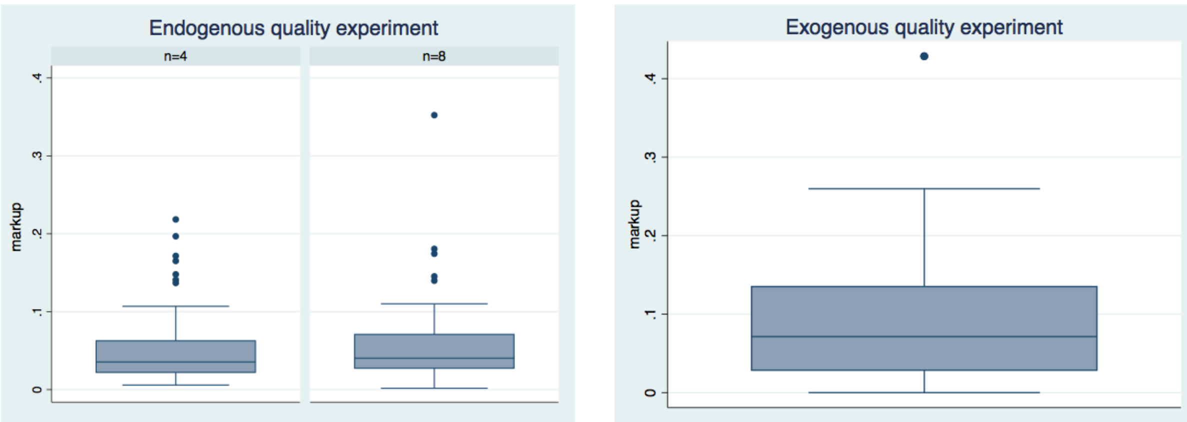


Figure 5: Markup distribution when price matters more than quality in buyer's preferences (γ is high)

At first glance, the two specifications of the technical function, here summarized in the variable "n", have no effect on the markup winning firms manage to extract over the costs borne. The difference results in fact not to be relevant, both in high γ settings and low γ ones. After performed the Mann-Whitney test we cannot reject the null hypothesis that the two markups are the same, since p-values are equal to 0.1536 (low γ setting) and 0.1364 (high γ setting).

On the other hand, the difference across experiments result here to be always significant at 1% level of confidence, regardless which setting has been considered. As the figure shows, it results particularly relevant when competition transfers on prices (i.e. when γ is high). In the latter case, while the endogenous quality experiment exhibit a huge downward shift with respect to the case in which quality matters more - half of the observations are here concentrated between 0.25 and 0.75 -, the other experiment instead, seems to go in the opposite direction. Although, as for the first procurement game, half of the observations are more concentrated around the median, when γ is high subjects managed to obtain higher margins than when γ is low. The cognitive bias above mentioned is not observable in the latter case. Subjects seem indeed not to have fully understood the instructions of the game.

Moreover, the huge difference across the experiments observed when dealing with aggregate data only, seems to be explained expecially in the treatment in which price matters more than quality in the overall score. While in fact, as *Figure 4* shows, there are not particularly differences when γ is low - except for the cognitive bias discovered -, when competition shifts on prices, subjects managed to extract higher margins in the exogenous environment rather then in the other.

4.3 How does markup relate with each player's type?

In order to enable the comparison between the two experiments, the markup is here defined as the difference between the selling price and the cost price, computed as a percentage of the latter. Recall in fact that the cost space is the only one subjects have in common across the two procurement games. Although the experiment in which the quality is endogenously treated exhibit marginal costs always lower than that provided by the other experiment, – mainly because of the presence of the cost parameter, θ, which is smaller than one – the cost functions were indeed constructed exactly in the same way. Therefore, for the purpose of our investigation, it is interesting to understand which results to be the experiment in which subjects extracted, for each value of the costs borne, higher margins.

Particularly appealing is the case of winners only. In this situation, because of the different cost structures could happen that winners are characterized by different costs according to which experimental game is considered. As already mentioned in *Section 3*, costs depend on the quality only, in the case in which the latter is exogenously treated, while both on quality and theta in the other case. Thus, before comparing the experiments, could be here useful to consider them separately and understand, for the case of winners only, how the markup relates to the quality level and the cost parameter θ,[30] so to be able to predict which could be the winners' overall production costs in both the games. Observing that, we will be able to capture the behavior of

[30] i.e. we want to understand which is the markup for each player's type so to be able to predict which will be the costs that let subjects to win the auction, across the experiments

the winners across the two different settings, controlling for what could be just explained by the differences in terms of structure.

4.3.1 Experimental game with endogenous quality provision

The following figure plots average margins against θ. Observed values are here analyzed for every specification of gamma - the weight of the financial score in the final overall score – and technical functions. For the sake of simplicity and to avoid that the notation used can be misleading, we defined T1 as the situation in which $\gamma = \frac{1}{4}$ and $n = 4$; T2 when $\gamma = \frac{1}{4}$ and $n = 8$; T3 when $\gamma = \frac{3}{4}$ and $n = 4$ and T4 when $\gamma = \frac{3}{4}$ and $n = 8$.

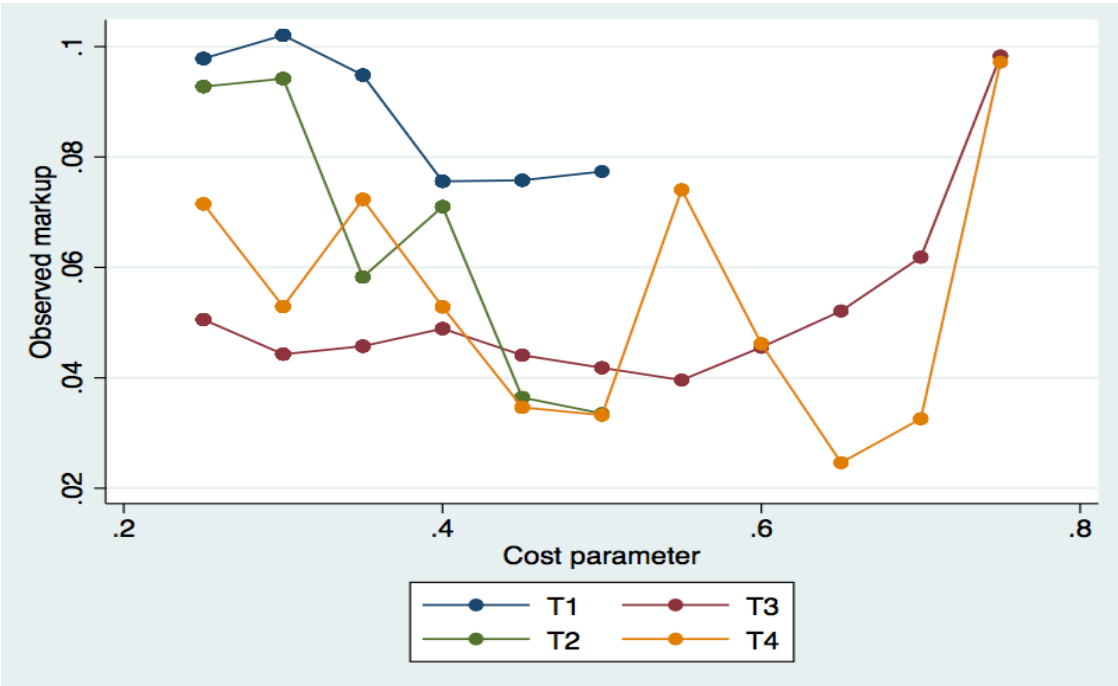


Figure 6: Average margins against cost parameter (θ). Sessions comparison.

As the figure shows, when quality matters more than price for the final decision of the buyer (T1, T2), the winners' average markup is always decreasing in the cost parameter. When θ is higher costs become higher as well, with so reducing the markup for every level of quality choice. The trend is the one expected, more efficient firms extracting higher margins, decreasing as the cost parameter gets bigger. Moreover, there is a threshold beyond which none of the partecipating subjects manage to ask a price higher than the cost incurred at the same time winning the auction. When the cost parameter is higher than 0.5 in fact, the graphs are interrupted since no firms manage to win. As soon as competition shifts on prices instead, we can immediatelly notice the downward shift of the subjects' average markup with respect to the previous case. In this setting, firms are forced to lower their price bid in order to win the auction, which immediately translates in lower margins. The trend, although still decreasing, is not monotonic anymore, meaning that less efficient subjects managed here to win more often with respect to before. In T3 or T4 for instance, even subjects with high cost parameters were able to extract huge margins when winning the auction. In some sense, when the weight assigned to the financial score in the overall score is higher than that of the technical score, all subjects, even the less efficient, have a chance of victory. Only efficient firms instead results to be able to stay in the market in the first above mentioned setting. Consequently, we expect that winning firms will exhibit lower costs when technical score has an higher weight than when fincial score is more relevant for the final decision of the buyer.

4.3.2 Experimental game with exogenous quality provision

The figure below plots the average markup this time against the quality level, which is here exogenously treated and randomly assigned to subjects before bidding. We used the quality since it's the only variable affecting here the cost of partecipating firms. As we did for the previous experiment, we denoted as T1 the environment in which the final score is more affected by the choice of quality and T2 that one in which the financial score has been assigned an higher weight.

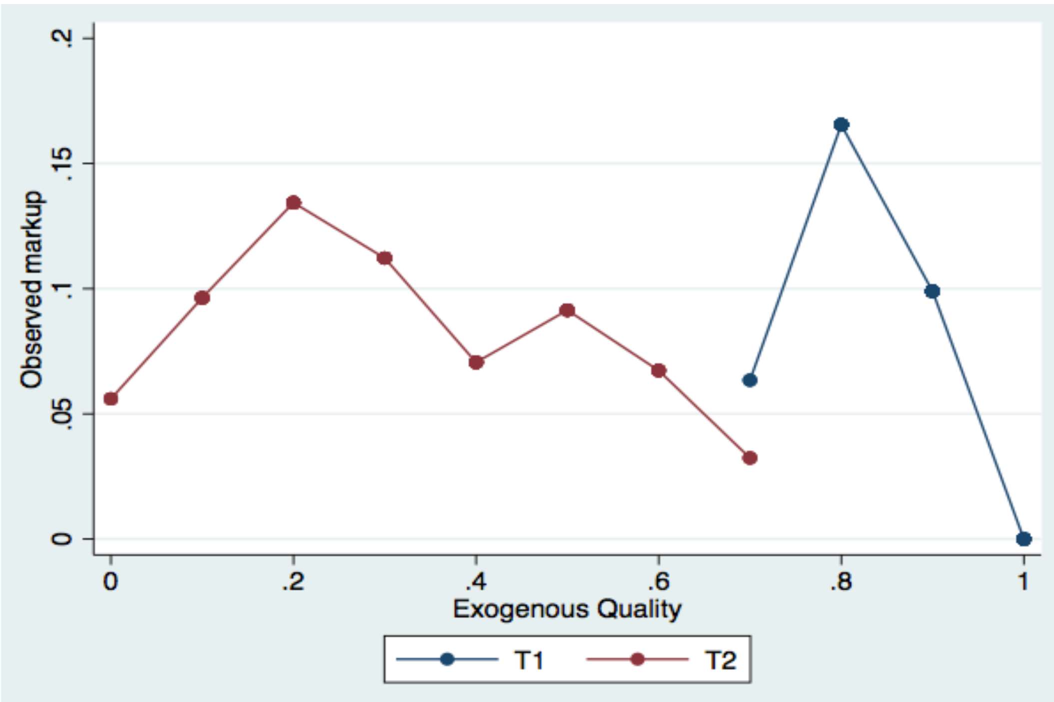


Figure 7: Average margins against cost parameter (γ). Sessions comparison.

The results provided by the figure above are interesting and consistent with basic intuitions. In particular, trends are always decreasing in quality - since here an high quality level immediately translets into higher costs, which negatively enter our markup function - except for subjects endowed with very low quality type in T2. In the latter case in fact, subjects with type smaller than 0.2 exhibit an inverted trend in the margin extracted, meaning that even if endowed with higher costs, they were able to ask higher prices as well still winning the auction. Moreover, average margins significantly differ according to which criterion has been given an higher weight in the final scoring rule. Specifically, both the graphs are interrupted after or from a given threshold of quality. In the first treatment (T1), none of the subjects with type smaller than 0.7 managed to win the auction and so to obtain a positive payoff. This has a clear explanation. When quality matters more than price in buyer's preferences – here summarized in the overall scoring rule adopted – we expect that only subjects endowed with very high technical attributes manage to win the auction, exploiting their peculiar advantages. When price matters more than quality instead (T2), we can easily notice how subjects with type higher than 0.7 are completely out of the market. In this case in fact, even if more subjects resulted having the contract awarded, the harsh price competition arised forces partecipants to lower their price to win the auction. As a result, subjects enriched with quality higher than the above threshold are obliged to offer prices higher than their competitors to cover costs they have to bear, but this prevents them to win the tender. Different from the experiment in which the quality has been endogenized, winning subjects managed here to obtain almost the same margin either when financial or techincal score has an higher weight in the overall outcome. This obviously depends on the higher costs suffered by subjects in T1 with respect to that of the other treatment, which directly reduse the markup for every characterization of the price bid.

4.3.3 Some remarks

What we discovered in this subsection results to be particularly useful for the interpretation of some significative differences across the esperiments, which will be discussed in detail in the remainder of this paper. In particular, considering the case of winners only, the procurement games here considered happened to be not comparable when the technical score has an higher effect in the determination of the overall score. Indeed, in the experiment in which subjects are endowed with technical attributes (q), only subjects with high type parameter – which translates in high costs as well – manage to win the auction; in the other experiment instead, only partecipants with θ lower than 0.5 manage to be selected by the mechanism. This obviously means that in the latter setting winners are characterized by costs not particularly high. Therefore, regarding winners only, the two experimental games may be compared just for the case in which the financial score is more relevant for the selection of the winning firm; situation in which both the games can be studied for every level of the production costs they have in common.

4.4 Which is the markup across sessions?

Margins obtained by winning subjects in the two procurement games are here compared against production costs. We first report results for each experiment separately and then put them togheter, distinguishing for settings in which gamma is high or low.

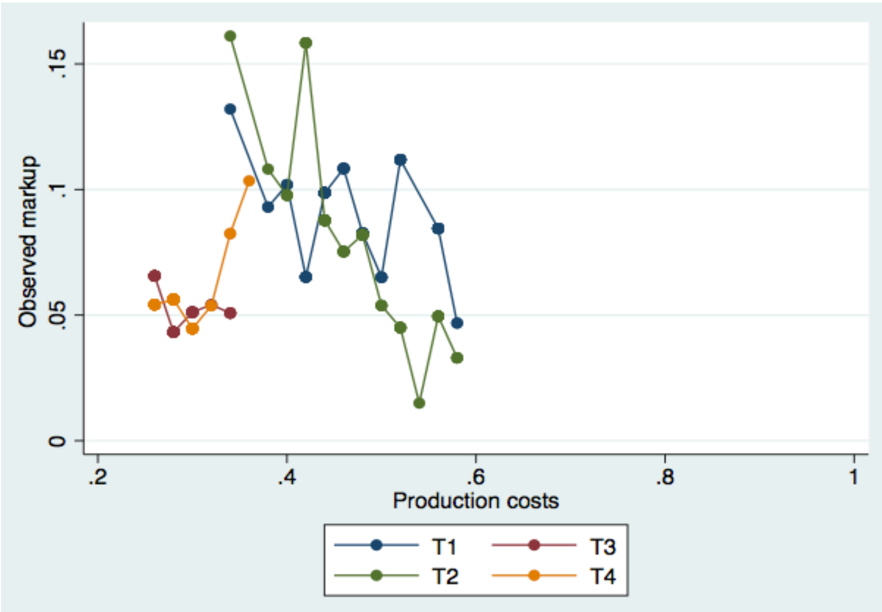


Figure 8: Observed markup against production costs. Endogenous quality experiment.

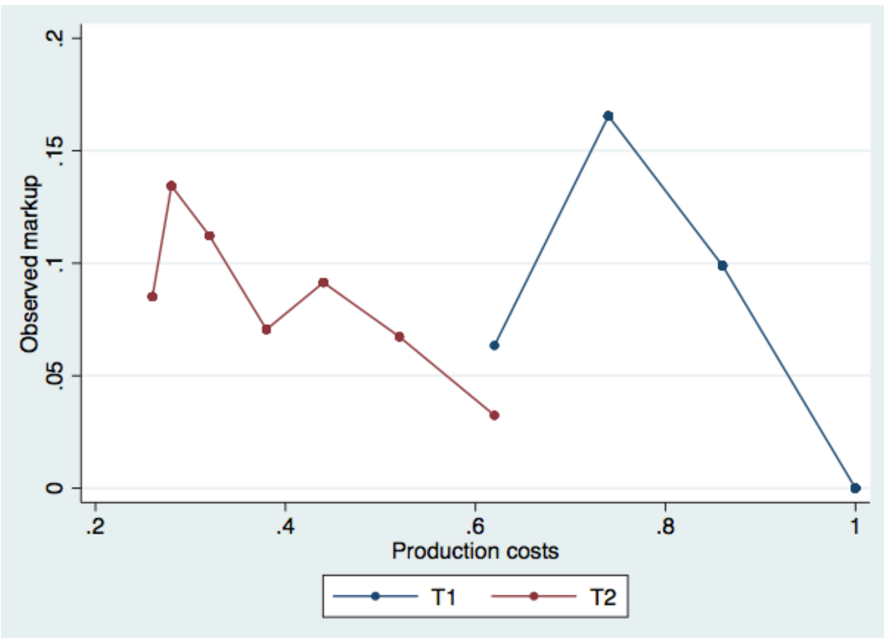


Figure 9: Observed markup against production costs. Exogenous quality experiment.

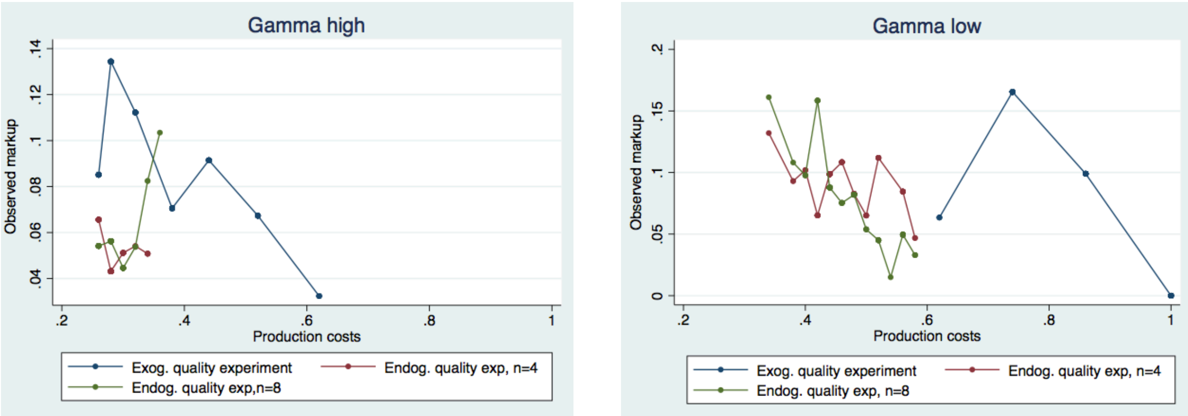


Figure 10: Observed markup against production costs. Experiments comparison by gamma.

Several considerations can be made. Both *Figure 8* and *Figure 9* suggest us how the different settings of the tender (i.e. different treatments) affect subjects' behavior and consequently their probability of winning the auction and the margin they are able to obtain over the production costs. Starting with the procurement game in which the quality is endogenously treated (*Figure 8*), average margins are always decreasing in production costs, but substantially differ according to which treatment of the game has been played. Specifically, when γ is high (T3 and T4), because of the strong price competition, only subjects suffering costs lower than 0.3625 manage to win the auction; when γ is low instead, the winners were the subjects characterized by costs higher than the previous case but still lower than a given threshold. This has a clear explanation. When quality matters more than price (T1 and T2), only subjects who offered high quality goods were selected by the mechanism. This directly translates in high costs as well, but still lower than 0.6 since from the previous section (*Figure 6*) we know that only subjects endowed with a cost parameter smaller than 0.5 won the tender (which obviously reduce production costs for every quality bid). Moreover, as already discussed in another Section of this paper, margins were drastically lower when price matter more than quality in the overall scoring rule.

Another interesting result is that the concavity of the technical score seems here not to influence the behaviour of the subjects. After performed the Mann-Withney test observed margins resulted to be never statistically different neither when gamma is high nor when it's low.[31]

As far as the other experiment instead is considered, as *Figure 9* shows, subjects result here to follow exactly the same intuition of the other experiment. Only subjects endowed with high quality attributes won the tender when gamma is low while only those with costs smaller than 0.6

[31] P-values were 0.1634 and 0.1370 respectively.

were selected when competition transfers on the price dimension. In some sense, in the latter setting quality looses importance in the determination of the winning firm. Recall that in this case the quality is the only variable affecting the costs of production, therefore it directly means that only subjects with high costs managed to win when γ is low, while only those suffering very high costs managed to win in the other setting. The latter finding becomes significantly relevant when we tried to compare the experiments differentiating for the parameter γ (*Figure 10*). Only when price is the determinant factor for the selection of the winning subject the markups extracted in the two procurement games are comparable for the same interval of the production costs. In particular, when every consideration about the quality choice is left out, subjects managed to obtain higher margins than when they are asked to determine the technical attributes of the goods to be sold. The difference is significant at 1% level of confidence running M-W test.
To better understand this result and so to be aware about the determinants of the markup in the two procurement games, the following subsection reports a more statistical approach.

4.5 Regression analysis

In this paragraph, margins obtained are analyzed more in detail. The effects of some independent variables on the markup extracted from the trade are deepened. Variables considered include: (i) tender characteristics, even interacted with other variables; (ii) production costs; (iii) some exogenous parameters; (iv) individual characteristics,[32]obtained by the answers of questionnaire provided at the end of each experiment. To reach this aim, we use three different econometric approaches which give us useful information about the robustness of results. Recall our dependent variable – the markup – is defined for the winners only, thus we cannot exploit the advantages of the Panel analysis simply since one subject may have won in one round and nevermore, which makes the Panel strongly unbalanced. We start the analysis by running the simple **OLS regression**, which is just our standard benchmark. However, some problems could arise when using OLS. The sample of observed margins is not random anymore, but it has been selected according to a specific rule – i.e. for each period, only subjects who reached the highest overal score of their group return with positive margins. In other words we recognize that some coefficients of OLS regression could suffer a distortion because of the selection of the sample. As a consequence we improve the previous test by running for the same variables the **Heckam's two step regression** in which the selection equation expresses the ex-post probability of

[32] Fall in this category all the exogenous individual characteristics such as age, gender, income, faculty of origin,…. and also the answers of the CRT test. The latters are here used to divide subjects in two groups, chategorized as impulsive and reflective, depending on the answers given.

winning[33]as explained by (i) the probability of winning the auction at equilibrium; (ii) some individual characteristics, which could influence the behavior of partecipants; (iii) tender's characteristics and their interactions with the others. In this way, we can analyze the sample of winners observed controlling for the ex-post probability that a each subject had to win. Results confirm what OLS already showed but each coefficient has been adjusted. The only thing we can do at this point to keep on improving our results is working on the selection equation of the Heckman regression. The latter in fact, estimates the effects of the independent variables on the probability of winning the auction, but treat each observation as an independent one, without understanding that each subject played for more than one rounds. So, we use **Two-stage least square regression** where at the first stage we control for the sample selection by regressing the ex-post probability of winning on the same variables used in Heckman, but considering the observations as panel data, so taking the adavantages that these models garantee.[34] The markup is year expressed as dependent on the same variables used in OLS and Heckman, and the prediction of the probability of winning obtained at the first stage.

Again, because of the different structures of the two procurement games, we first analyze them separately, trying to replicate as well as possible the regressors used. Before starting the analysis a brief description of some variables used in the determination of the markup is reported.

(i) *gamma*, is a dummy variable taking value 1 if gamma is high

(ii) *n_dumy*, takes value 1 if the technical score is more concave in quality

(iii) *n_dummy*gamma*, represent the interaction between n_dummy and gamma

(iv) *Impulsive*, takes value 1 if subject is impulsive, i.e. if he answered impulsively to the crt

(v) *Gender, faculty, room size ratio, weekly budget, age, age^2, grade, schooling, schooling father*, is our group of control. It identifies some exogenous individual characteristics, which are here used to predict the markup extracted.[35]

(vi) *Cost*, expresses the production costs borne for each subject

(vii) *Cost*gamma*, gives us the effect of costs on the markup, when gamma is equal to 1. The effect should be added to that of cost to interpret the net effect.

[33] It's a dummy variable, equal to 1 if the subject won the auction, 0 otherwise.

[34] The command used is *xtprobit, STATA*.

[35] We tried to use whenever possible the same variables for both the experiments, except for some tender and individual characteristics, which differ across the projects.

4.5.1 Experimental game with endogenous quality provision

Before analyzing results on the markup of the different approaches followed, it is worthful to understand how we managed to control for the sample selection. We used the theoretical probability of winning as one of the independent variables. The latter depends here just on the optimal quality choice and the cost parameter with which each subject is endowed before bidding. Obviously, higher is θ lower results to be the probability of winning according to the theoretical model. We consequently extract for each group the subject with the smallest type and then construct a dummy variable, "*opt_prob_win*", taking value 1 if θ is equal to the minimum of the group for each period, and 0 otherwise.

The following table summarizes results for heckman's selection equation and the first stage of the two-stage least square estimation.

Variables	**Panel Probit (Winner)**	**Heckman selection (Winner)**
opt_prob_win	2.095458**	1.7458837***
gamma	1.344174***	1.3185608***
opt_prob_win*gamma	-.8364287***	-.75424061***
opt_quality	.9032671***	.7959735***
impulsive	-.0129471	.04941501
impulsive*gamma	-.5384204	-.56829813***
gender	-.2896501	-.19041893**
faculty	-.0431969	-.03585533
room size ratio	.1675796	.17849016***
weekly budget	.0064475**	.00471949***
age	-.4589053***	-.34884542***
age^2	.0063882***	.00486285***
grade	.1223087	.10295688**
constant	3.081042**	1.8876588**

Reg 1: Estimation of the probability of winning the auction. Panel Probit, Heckman.

No particular differences appear when comparing the two methods above. Almost every coefficient maintains its sign unchanged when passing to one regression to the other, except for *impulsive*, which although never significant, becomes consistent with our expectations using panel probit only – i.e. subjects categorized as impulsive seem to have lower probability to win the auction. *opt_prob_win* on the contrary, have always a positive effect on the realized ex-post probability, significant at 1% level of confidence. However, the magnitude of this effect changes substantially according to gamma. Specifically, when gamma is high – price matters more than quality in buyers' preferences – subjects endowed with the lowest θ have still an higher probability to win but lower than when gamma is low. This is in accordance with *Figure 6*, which reports how inefficient subjects managed to win more often in high gamma settings rather than in high gamma ones, where they are completely out of the market. Another interesting results comes from the *gamma*'s coefficient. As ***Reg1*** shows and consistently with previous findings, subjects extracted lower margins when competition shift on the price dimension. None of the individual characteristics instead results to be crucial for the determination of the winner, but the *age* of the subjects and their *weekly budget*. The net effect of age is indeed posititve, meaning that older subjects, maybe because of the more experiences they lived, seem to be able to win more often than the others.

It's now time to analyze the determinants of the margins obtained. OLS, Heckman and Two-Stage estimations are proposed.

Variables	**OLS (markup)**	**Heckman (markup)**	**Two-Stage least square (markup)**
gamma	-.11354072***	-.09885461***	-.09597521***
cost	-.2372933***	-.16718995**	-.17328076**
n_dummy	-.01425252	-.0117439	-.01185636
n_dummy*gamma	.02268865**	.02221214**	.02191753**
Impulsive	-.01277774	-.01403195	-.01326429
Impulsive*gamma	.02751632**	.03257496***	.03039091**
gender	.020375***	.02297525***	.02323486***
faculty	.00178669	.00154384	.00186981
room size ratio	.00912665***	.00830509***	.00890532***
weekly budget	-.00025573***	-.00030214***	-.00030778***
age	.0087482**	.01216662***	.01262969***
age^2	-.00012709**	-.00017382***	-.00018036***
grade	.00037438	-.00163463	-.00156008

winner_hat			.03146912***
constant	.06221399	.00840431	-.03008267

Reg 2:*Markup estimation. Ols, Heckman, Two-Stage least square.*

The first observation to be done is that the *selection bias* was not so relevant, since all the approaches return almost with the same results, at least in the sign. Specifically, *gamma* and *cost,* as expected, negatively affect the markup and are significant at 1% and 5% level of confidence respectively[36]. The concavity of the technical score and the impulsiveness of the subjects instead, although inducing a negative effect, are never significant except for the cases in which are interacted with gamma. When the technical score is more concave in quality and the competition is on the price dimension – γ is high – subjects seem indeed to be able to extract higher margins. The puzzling result is that the same thing happens for impulsive subjects. The latter has one possible explanation: if among the winners there is someone which is categorized as impulsive and so less rational, we probabily expect he will ask high prices which, still reducing his probability to win, increase his margin in case of victory. Both *gender* and *age* are significant. Females seem in this experiment to manage to obtain higher markup than men as well as older subjects do with respect to the others. The same is valid for *room size ratio.* On the other hand, *Weekly budget's* cofficient has always a negative effect on the margin extracted. One possible reason could be that subjects with lower budgets are more incentivized to get money from the esperiment and so put more efforts when taking their decisions. The variable *winner_hat* is computed for two-stage least square estimation only.[37]It's sign is positive and significant at 1% level of confidence, which means that an increase of the predicted probability of winning the auction provides an higher markup obtained after the trade has been made. The fact that it's significant, means that our first stage regression captured pretty well the effective probability.

4.5.2 Experimental game with exogenous quality provision

Again, we prefer to present first the results of the first stage regression and the heckman's selection one, in which our dependent variable is the ex-post probability of winning the auction – *winner* – and the independent variables are a set of individual characteristics we used as "controls", and some tender's attributes. As we did for the other experiment we even consider the probability to win provided by the model, i.e. the theoretical probability of winning. The

[36] Two-Stage least square

[37] OLS does not take into account the selection problem and Heckman computes itautomatically by default.

mechanism should be able, at least in equilibrium, to select the producer which results to be the most qualified - i.e. the subjects with the highest exogenous quality. As a consequence, after isolated the subject that for each group and for each period is characterized by the highest quality type (α), we construct a dummy variable – *opt_prob_win* – which takes value 1 if α is equal to the maximum, and 0 otherwise. We even consider *cost* as one of the possible predictors since in this game it's influenced only by the exogenous quality, i.e. the type assigned to each player before bidding. The following table summarizes results obtained.

Variables	**Panel Probit (Winner)**	**Heckman selection (Winner)**
opt_prob_win	2.657862***	2.2320389***
gamma	6.172001***	4.6713176***
opt_prob_win*gamma	-7.999103***	-7.5589781
cost	6.13599 ***	4.5763608***
cost*gamma	-8.501093***	-6.5408554***
impulsive	-.7794877*	-.65625936**
impulsive*gamma	.4824631	.51664345
gender	.1035718	.01229516
faculty	-.0487195	.0088467
schooling	.1178072	.24049192
schooling_father	3176208	.3234906***
room size ratio	.6641911	.58141901*
weekly budget	-.0008299	-.00019145
age	-.7183284	-.46467205
age^2	.0144117	.00865067
constant	1.225316	-.39806597

Reg 3: *Estimation of the probability of winning the auction. Panel Probit, Heckman*

In this case too, both panel probit and heckman's selection regression result almost with the same findings. Both *opt_prob_win* and *gamma* coefficients are consistent with our interpretation and previous findings. Specifically, subjects who the theoretical mechanism should select as

possible winners really have more probability to win the auction and the difference with the others is also significant at 1% level of confidence for both the approaches used. The latter effect drastically changes direction when interacted with gamma. When gamma is high, the effect of being the subject with the highest quality becomes negative, since higher quality here directly translates in higher costs as well which substantially limit subjects to lower their price bid, so excluding them from the list of possible winners. The same reason could be given for *cost*'s coefficient. Another interesting result is obtained by the analisys of *gamma*. When competition shifts on the price dimension - γ is high -, as already discussed in *paragraph 4.2* of this paper, more subjects managed to win than when quality matters more. In some sense we could say it's the most egalitarian setting, since more subjects are really able to win than the other case. Passing to the individual characteristics, only *impulsive* results significant and exhibits a negative coefficient, meaning that impulsive subjects have lower probability of winning the auction. On the contrary, all the other personal features seem not to be relevant in the selection of the winning firm.[38]
Let's now look more in detail which happen to be the margin's determinants in this experiment.

Variables	**OLS (markup)**	**Heckman (markup)**	**Two-Stage least square (markup)**
gamma	-.39580779***	-.50210984***	-.48494463***
cost	-.56964554***	-.72918275***	-.73174302***
cost*gamma	.38613824***	.65638723***	.64159387***
Impulsive	.00522562	.01449052	.01464965
Impulsive*gamma	-.03382748	-.03653994	-.03208645
gender	-.00555828	-.00426446	-.00691416
faculty	.0053239	.00098006	.00289137
schooling	-.01221717	-.01824639	-.01433848
schooling_father	-.00797943	-.01589763	-.01287982
room size ratio	-.04012161	-.04738213	-.0469022
weekly budget	.0000512	.00003289	.00004329
age	.04715849	.05859804	.06416455**
age^2	-.00099973	-.00117754	-.00130684**
winner_hat			.13617272***

[38] Only in heckman' selection the schooling of the father and the room size ratio have postive effects on the probability of winning. When passing to panel probit, none of the two remains significant, albeit the sign of the effect is confirmed.

constant	.08341278	.12331317	-.09411787

Reg 4:Markup estimation. Ols, Heckman, Two-Stage least square.

Again, no particular differences among the approaches used seem to be relevant. Exception made for the magnitude of the coefficients, all of them respect the sign and their statistical relevance across the approaches. Only the *age* of partecipants, althoug not significant with OLS or Heckman regressions, becomes significant when using two-stage estimations, exhibiting positive sign. For what concerns *gamma*, as expected, it shows a negative sign, significant at 1% level of confidence. Subjects managed to extract lower margins when competition is centred on the price dimension. The same effect is discoverd for the *Cost* variable, which induce a huge negative effect on the markup extracted and is significant at 1% level of confidence for all the tests. Only when interacted with gamma, the magnitude of this effect gets smaller. As *Figure 11* reports, when analyzed against production costs, average margins, although almost always downward sloping, are more flat when γ is high than when it is slow. In the latter case in fact, a small increase in the production costs causes a huge negative jump in the markup obtained. The effect instead is mitigated when competition shifts on prices. None of the individual characteristics here analyzed seem to be able to have a considerable impact on the observed markup. Differently from the other procurement game, the mechanism is here not able to make subjects to exploit their advantages in terms of personal skills.

4.5.3 Margins' comparison

After having understood which are the main drivers of the margin obtained by subjects in the two different games, it's interesting now to see which experimental setting let to extract higher markup and how tender's caracteristics are percieved across the experiments. For this purpose, we run a two-stage regression in which our dependent variable is the aggregate margin for both the games. For the first-stage regression we used predicted values obtained in *paragraphs 4.4.1*and *4.4.2,* when we dealt with each experiment separately. We then generated a dummy variable which discriminates among experiments and which was even interacted with all tender's characteristics the two procurement games have in common. Before discussing results, we report a short description of the variables we never met so far and the equation of the model used.

(i) *D_exp*, is the dummy taking value 1 if the endogenous quality experiment is considered, 0 otherwise

(ii) *D_exp*cost*, is the interaction between D_exp and costs. It gives us the effect of a raise in cost only for the endogenous quality game.

(iii) *D_exp*gamma,* is the interaction between gamma and the project. It's equal to 1 if gamma is high and we are in the endogenous quality environment.

(iv) *D_exp*cost*gamma,* is a dummy which discriminates not only among experiments, but even among gamma (high/low).

(v) *Individual_control,* is just a set of individual characteristics we used as control. They exactlu correspond to those used in previous analysis.

$$
\begin{aligned}
Markup = \beta_0 + \beta_1 D\_exp + \beta_2 cost + \beta_3 D\_exp\_cost + \beta_4 gamma + \beta_5 D\_exp\_gamma \\
+ \beta_6 cost\_gamma + \beta_7 D\_exp\_cost\_gamma + \beta_8 impulsive + \beta_9 impulsive\_gamma \\
+ \beta_{10} winner\_hat + \varphi\ individual\_control
\end{aligned}
$$

Independent variables	**(Aggregate markup)**
D_exp	-.4125995***
cost	-.5949209***
D_exp*cost	.3787424***
gamma	-.4224944***
D_exp*gamma	.3602156***
cost*gamma	.4433502***
D_exp*cost*gamma	-.4891308** (pvalue 1.6%)
impulsive	-.0036212
impulsive*gamma	.0068024
winner_hat	.0303443***
gender	.0154474***
faculty	.0017898
room size ratio	.0090109***
weekly budget	-.0001002*
age	.0064887*

age^2	-.0000998**
constant	.4685917***

Reg 6: Aggregate Markup estimation. Two-Stage least square approach.

As *Reg 6* shows, when subjects are asked to bid on both the price and quality dimension margins obtained from the trade are substantially lower than when they have to formulate an offer on the price dimension only. The coefficient of *D_exp* is indeed negative and significant at 1% level of confidence, as well as that of *Gamma.* As already noticed in the previous subsection, when gamma is high, subjects return on average with lower margins because of the strong price competition. What it's interesting here is that this effect is significantly mitigated in the endogenous quality experiment with respect to the other. When adding the effect of *D_exp*gamma* in fact, margins continue to be lower in high gamma settings but the difference tends to be substantially smaller than that observed in the other experiment. Another interesting result comes from the different marginal effects that a raise in production costs have on the margin extracted. The following table summarizes these effects discriminating across experiments and low/ high gamma settings.

Costs' marginal effect	**Gamma low**	**Gamma high**
Endogenous quality experiment	$\beta_2+\beta_3$ (- 0.2161)	$\beta_2+\beta_3+\beta_6+\beta_7$ (- 0.1516)
Exogenous quality experiment	β_2 (- 0.5949)	$\beta_2+\beta_6$ (- 0.2620)

Table1: Marginal effects of a rise in production costs on the markup extracted.

Table1 reports the importance of production costs in the determination of margins obtained for each specification of the parameter γ. As can easily be noticed, although markup are always negatively related with costs, their impact considerably changes according to which experimental game has been played. Starting from the game in which the quality is exogenously treated, the marginal effect of a raise in production costs is β_2 (- 0.5949) in low gamma settings, while $\beta_2+\beta_6$ (- 0.2620) when gamma is high. Thinking about the latter game's structure and the selection rule of the winning firm previously discussed, this result is normal and consistent with our expectations. When gamma is low, only subjects endowed with very high technical attributes – and so high costs as well - manage to win the auction. This obviously lower the action space of each subject and makes them more dependent on costs they are suffering.[39] What is interesting here is that this relation between gamma, production costs and markup is not confirmed in the experiment in which the qualidy bid is endogenously determined by partecipants. In the latter situation, the marginal effect of a rise in production costs is equal to β_2+β_3 (- 0.2161) when gamma is low, while β_2+β_3+β_6+β_7 (- 0.1516) when gamma is high, and these two effects result to be not significantly different from each other. In other words when subjects decide both the price and the technical attributes of the good or service they want to sell within the auction, they seem not to be particularly constrained by buyer's preferences. Albeit on average subjects obtained lower margins in high gamma setting, either when gamma is low or when it's high the marginal effect that a raise in costs have on the margin extracted is almost the same. This result obviously depends on the different selection of the winning firm across experiments, which works better in the exogenous quality environment rather than in the other, mainly because of the more degree of freedom of partecipants.

4.6 Deviations from optimality

In this Section deviations from equilibrium strategies are analyzed. Equilibrium margins are here computed using optimal price bid strategies and, for the experiment in which subjects bid even on the quality dimension, optimal costs. The latters, for every player's type, depend on the optimal quality bid only. For the other experiment instead, the players' type is the production costs – they depend on quality only, which here is exogenously assigned to subjects – and the optimal price bid is just computed according to the latter formula:[40]

$$optimal\ price\ bid = 1 - optimal\ rebate\ bid$$

For the sake of transparency, as common practice in this paper, we prefer to analyze each procurement game separately.

4.6.1 Experimental game with endogenous quality provision

As already mentioned in previous sections optimal strategies of this bayesian game are provided by Che (1993). Obviously only winning subjects are considered, since who did not win the auction returns with a payoff of zero. Below, we first report the case in which the final overall score is more affected by the quality choice (γ low) and then the other (γ high). Again, we differentiate for the parameter "n", expressing the concavity of the technical score.

[39] Recall in fact that subjects cannot formulate a price bid lower than costs they have to bear.

[40] See *Section 2* for more details.

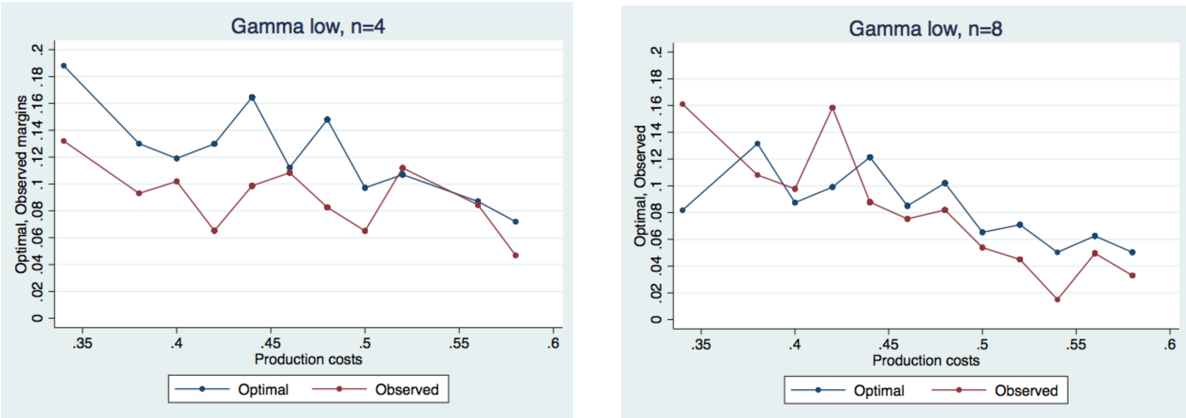


Figure 11: Optimal vs Observed margins. Endogenous quality provision. N= 4|8, γ low

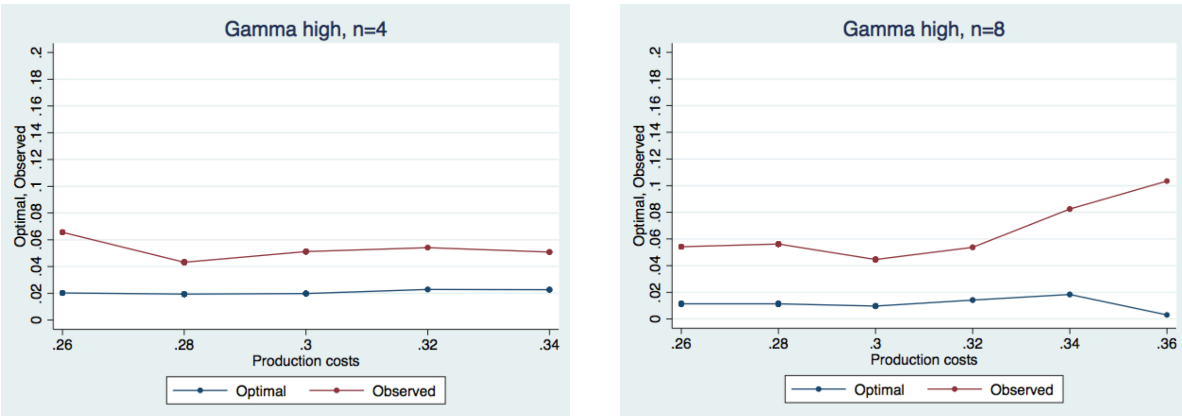


Figure 12: Optimal vs Observed margins. Endogenous quality provision. N= 4|8, γ high

Results summarized in *Figure 11* and *Figure 12* are particularly interesting expecially in light of M. Rosi's results (2015). One of the main findings of the author was that the difference between observed price bids and average costs – the markup in absolute value - was almost constant for every specification of the cost parameter (θ).[41] This obviously contradicts the theoretical predictions, which imply that the markup obtained must be decreasing with θ. Theoretically, less efficient subjects – that ones endowed with higher type – have less probability of winning the tender and so to extract positive margins.

When analyzed against production costs, deviations from equilibrium margins strongly depend on which setting of the game has been played. In particular, albeit in all of the treatments above

[41] The author considered average costs and average prices for all subjects. In this paper only the markup extracted by the winner is instead analyzed

observed margins follow the optimal ones at least in the shape, subjects seem to understand and replicate better the theoretical predictions when γ is low, rather then when it is high. As *Figures* shows, observed margins where almost always lower than that pronosticated by the model when γ is low, regardless of the different concavity of the technical score - here not statistically significant - while are sistematically higher when competition transfers on the price dimension. In the latter case, althoug subjects' margins are strictly lower than when γ is low – due to the harsh competition on the price dimension – the equilibrium ones result to be even lower. The distance between the two is in addition almost constant for each level of the production costs. One possible reason of this behavior could be addressed to one common cognitive bias. When γ is high, subjects percieve that even if they win the auction they probably return with a payoff close to zero, expecially if they suffer high production costs. Therefore, they prefer to bid higher prices than expected, which, albeit lowering the probability of winning, can garantee them acceptable payoffs in case of winning.

Only when the technical score is more concave – n=8 – few subjects with higher production costs managed to obtain higher margins as well. This clearly depends on the poor composition of other players' tenders, which let them to ask for high prices, still resulting with winning the auction.

4.6.2 Experimental game with exogenous quality provision

The equilibrium bid,[42]which in this experiment regards the rebate on a base price set equal to 1, are here provided by the G. Albano, G. Ponti et al (2015), who followed the approach used by Che to find the symmetric optimal strategy for each subjects. Specifically they provided an optimal rebate bid which now strongly depends on the specification of the parameter expressing the weight of the financial score in the final overall score. The following figures compare observed and equilibrium margins both when competition is on price or quality dimension.

[42] Recall in this case the quality is completely exogenous, so the only bid subjects need to determine is that on the rebate dimension.

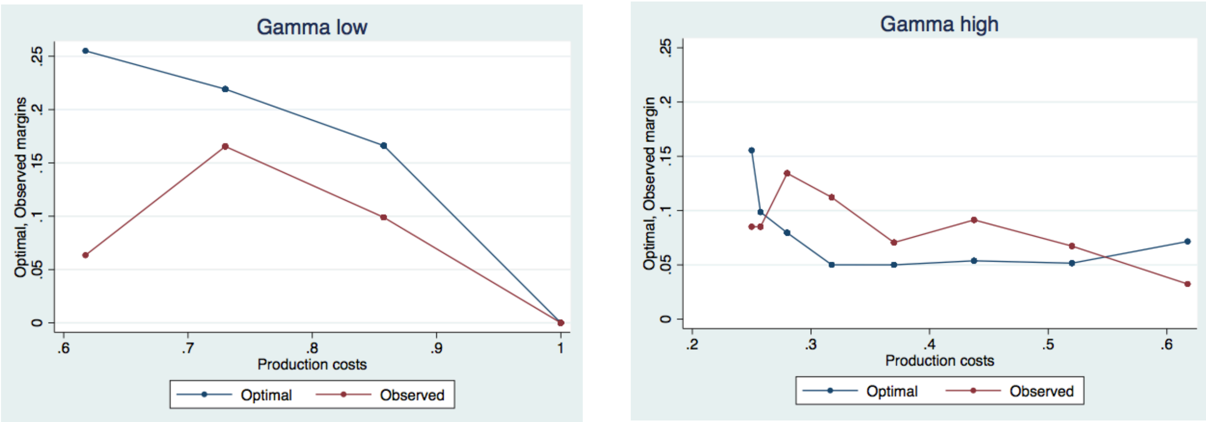


Figure 13: Optimal vs Observed margins. Exogenous quality provision. γ high/low

The first observation to be made regards the two markups obtained by playing the equilibrium strategies. As *Figure 13* shows, optimal margins differ in the shape according to the parameter γ. Specifically, they are always deacreasing in production costs when γ is low, while the same relation is not always true in the setting in which γ is high. In the latter case in fact optimal margins strictly decrease with costs when they are quite low – i.e till 0.3 – while are constant and sometimes increasing too for higher production costs. As regards the comparison between observed and equilibrium margins, the same relation discovered for the other procurement game is consistent even in this game. Subjects on average obtained markup always lower than that pronosticated by the equilibrium strategies when γ is low, while almost always higher when γ is instead high – except for extreme values of production costs. Again, when subjects expect to gain margins close to zero, they prefer to increase their price offer so to return with acceptable payoff in case of winning the auction. In this setting, differently from the experimental game in which the quality is endogenous treated, subjects managed to obtain higher margins but still decreasing with production costs.

SECTION 5

5. Conclusions

This article has studied in detail how subjects behave in two different experimental procurements with scoring auctions awarding criterion. One in which subjects formulate both price and quality bids and the other in which they are free only to bid on the price dimension. Margins extracted from the trade by winning subjects were analyzed. We first studied them separately and then compared them with each other. The main question of this paper was indeed that one of understanding, from a behavioral perspective, which was the setting in which partecipants were able to obtain higher markup. The first observation to me made regards the selection rule. The latter was more pronounced in the exogenous quality experiment rather than in the other, simply because of the more degrees of freedom each subject has in choosing the production costs he wants to bear in the latter case. When technical merits of the tender have more weight in the buyer's utility function – i.e. the scoring rule – only subjects with very high quality type managed to win the auction in the exogenous quality environment. On the other hand, a significant difference across margins in the two experiments was discovered. When subjects leave out every consideration about the quality choice, they manage to perform better, resulting with higher margins. What is interesting is how these margins are influenced by the tender's characteristics the two experiments have in common. In the experiment in which the quality is endogenously determined, the mechanism designer seems to manage to influence less subjects' behaviour only by expressing his preferences in the overall scoring function. Albeit in both the experiments they return with lower markup in high competitive settings with respect to the others, these difference tends indeed to be considerably lower when quality is endogenously treated. The same happens when analyzing the impact of production costs. Whereas markup are heavily influenced by production costs in the exogenous quality environment – in the sense that an increase in costs induces a huge negative fall of the margin extracted -, the same relation is substantially mitigated when subjects can decide both the price and the quality of their tender. Moreover, the difference across experiment is emphasized when the buyer cares more about the technical specifications of the contract rather than its price. We could say that the higher action space subjects have in the endogenous quality setting makes them to be less affected by buyer's needs and so to be able to reach an acceptable margin regardless the costs they are suffering.

When compared against markup pronosticated by the theoretical models behind the experiments, a common evidence is discovered. While subjects return most of the times with margins lower than the optimal ones when competition is on the quality dimension, they sistematically do better than that when competition shifts on prices. In this case, monetary payoffs were too low in equilibrium and this obviously prevent optimal bids to be an interesting alternative. Subjects indeed preferred in both the experiments to reduce their probability of winning the auction in exchange for accepatable rewards in case of victory.

The investigation carried out in this paper was from the point of view of the partecipating subjects only. It would have been really interisting to understand how buyer's utily function and his costs savings obtained with the auction scheme could have been influenced by the exogeneity or endogeneity of the quality bid. Unfortunately, the structures of the two procurement games here considered was so different so to prevent us every possible considerations about other variables, which result to be not directly comparable.

References

Albano G. L., Ponti G., Di Paolo R., and Sparro M. (2015). Absolute vs. Relative scoring in Experimental Procurement . Universidad de Alicante & LUISS Guido Carli, mimeo.

Asker, J., & Cantillon, E. (2005). Optimal procurement when both price and quality matter.

Asker, J., & Cantillon, E. (2008). Properties of scoring auctions. The RAND Journal of Economics, 39(1), 69-85.

Battalion, Kogut e Meyer. Learning in common value auctions, Experimental Games and Market.

Bergman, M., & Lundberg, S. (2011). Tender Evaluation and Award Methodologies in Public Procurement. Umeå Economic Studies, (821).

Branco, F. (1997). The design of multidimensional auctions. The RAND Journal of Economics, 63-81.

Chang, W. S., Chen, B., & Salmon, T. C. (2014). An investigation of the average bid mechanism for procurement auctions. Management Science.

Che, Y. K. (1993). Design competition through multidimensional auctions. The RAND Journal of Economics, 668-680.

Chen-Ritzo, C. H., Harrison, T. P., Kwasnica, A. M., & Thomas, D. J. (2005). Better, faster, cheaper: An experimental analysis of a multiattribute reverse auction mechanism with restricted information feedback. Management Science, 51(12), 1753-1762.

Cox, J. C., Smith, V. L., & Walker, J. M. (1982). Auction market theory of heterogeneous bidders. Economics Letters, 9(4), 319-325.

Cripps, M., & Ireland, N. (1994). The design of auctions and tenders with quality thresholds: the symmetric case. The Economic Journal, 316-326.

Dastidar, K.G. (2013). *Scoring auctions with non-quasilinear scoring rules*. Mimeo. Japan: ISER, Osaka University.

Dastidar, K. G. (2014). Scoring Auctions A Brief Survey. *Studies in Microeconomics*, *2*(1), 35-48.

Davis, D. D., & Holt, C. A. (1993). Experimental economics. Princeton university press.

Dini F, Pacini R and Valletti T (2006). "Scoring Rules", in N., Dimitri, G., Piga, G. & Spagnolo (eds.), *Handbook of Procurement*, Cambridge, UK, Cambridge University Press, 293–321.

Fischbacher, U. (2007). z-Tree: Zurich toolbox for ready-made economic experiments. Experimental economics, 10(2), 171-178.

Greiner, B. (2004). The online recruitment system orsee 2.0-a guide for the organization of experiments in economics. University of Cologne, Working paper series in economics, 10(23), 63-104.

Grimm, V., Kovarik, J., & Ponti, G. (2008). Fixed price plus rationing: an experiment. Experimental Economics, 11(4), 402-422.

Hanazono M., Nakabayashi J. & Tsuruoka M. (2011). A theoretical analysis of scoring auctions with a general form of scoring rules. mimeo.

Harsanyi, J. C. (2004). Games with Incomplete Information Played by "Bayesian" Players, I–III: Part I. The Basic Model&. Management science, 50(12_supplement), 1804-1817.

Kagel, J. H., & Levin, D. (1990). Auctions: a survey of experimental research, 1995–2008. Auctions Handbook, 2.

Kagel, J. H., & Roth, A. E. (1995). The handbook of experimental economics (pp. 501-85). Princeton, NJ: Princeton university press.

Klemperer, P. (2004). What Really Matters in Auction Design, from Auctions: Theory and Practice.

M. Bichler (2000). An experimental analysis of multi-attribute auctions. Decision Support System, 29:249-268

Rosi M. (2015). Procurement Auctions with endogenous investment: a theoretical and empirical investigation. LUISS Guido Carli, Rome.

Milgrom, P. R. (2004). Putting auction theory to work. Cambridge University Press.

Nishimura, T. (2012). Scoring Auction by an Informed Principal (No. gd11-224). Institute of Economic Research, Hitotsubashi University.

Onderstal, S., & Van de Meerendonk, A. (2008). How (not) to design procurement mechanisms: A laboratory experiment. *3rd International Public Procurement Conference Proceedings* (pp. 391–421). 28–30 August 2008. Amsterdam, the Netherlands.

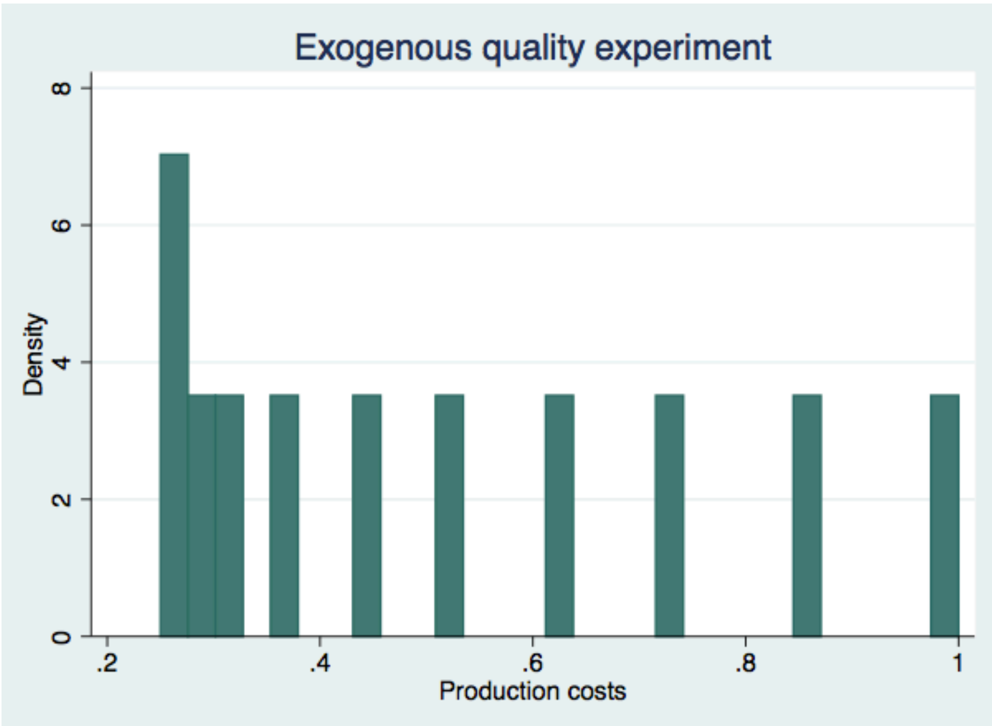
Riley, J. G., & Samuelson, W. F. (1981). Optimal auctions. The American Economic Review, 381-392.

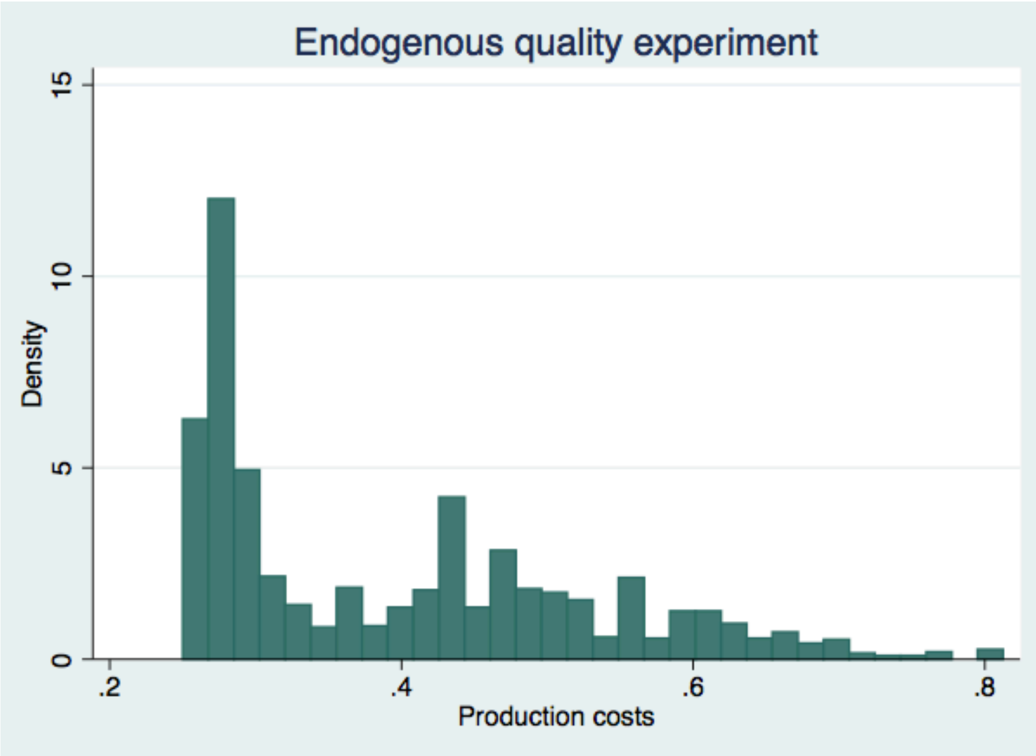
Ufficio Studi di Consip S.p.a. (2011), "L'attribuzione del punteggio economico nelle gare all'offerta economicamente più vantaggiosa: Formule a confronto", annex to the Green Book of Patrimoni PA Net "Analisi, indirizzi e proposte per il mercato dei servizi di gestione dei patrimoni pubblici".

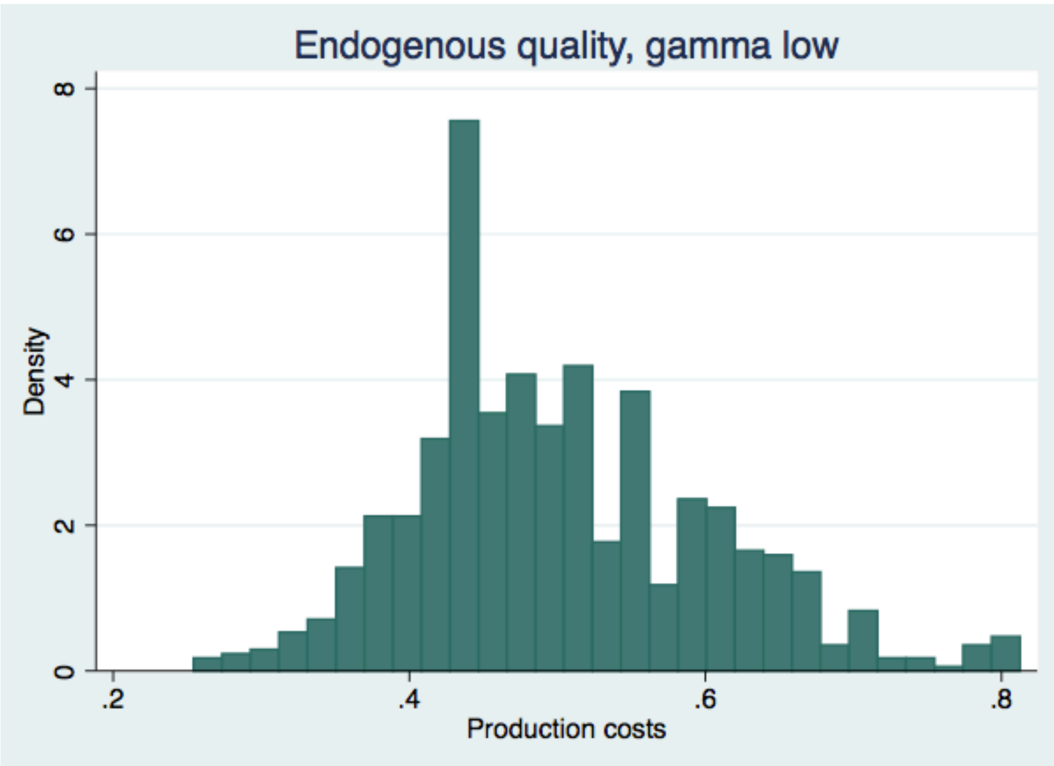
Wilson, R. (1977). A bidding model of perfect competition. The Review of Economic Studies, 511-518.

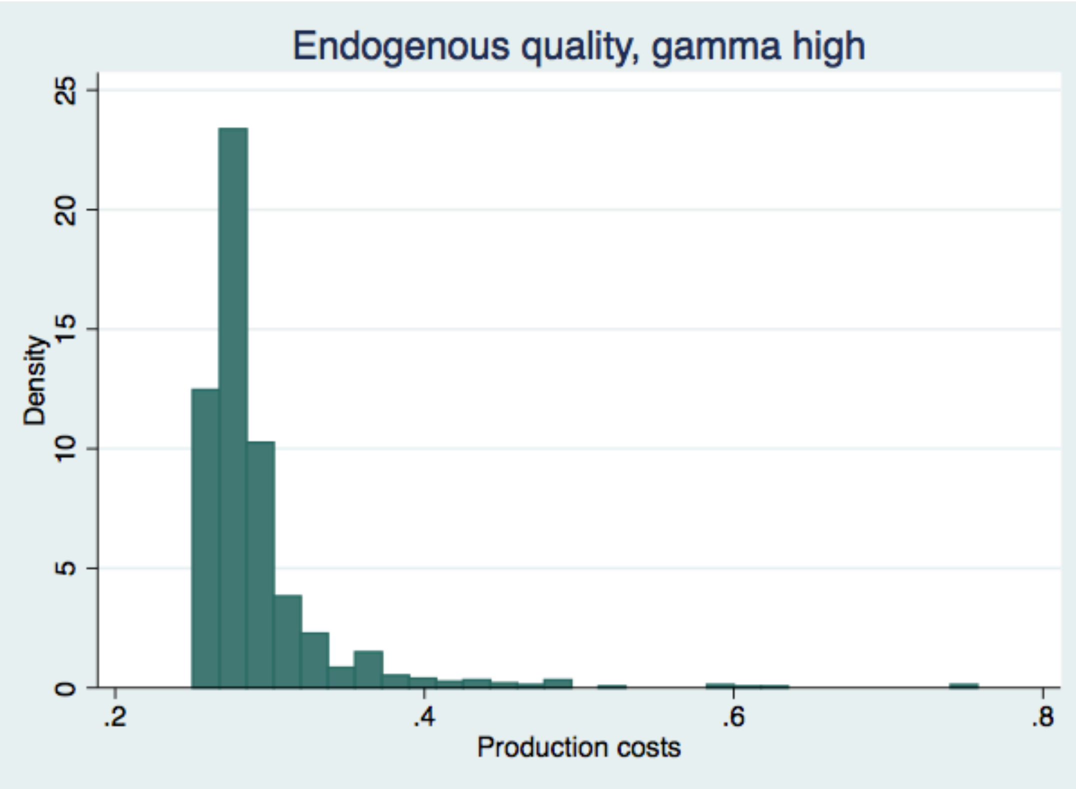
Appendix A

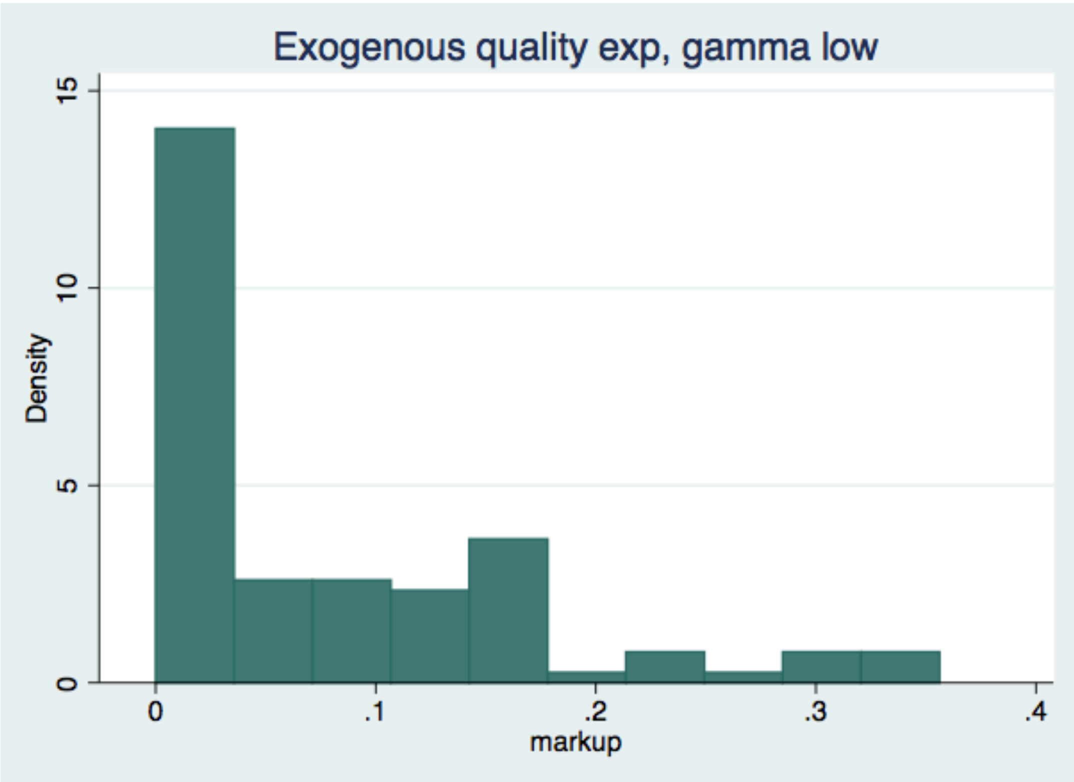
Some figures here reported should be considered to understand the results provided by this paper. In particular they could be really useful to control for the different structures of the two procurement games. Cost structure and margins obtained are below described.

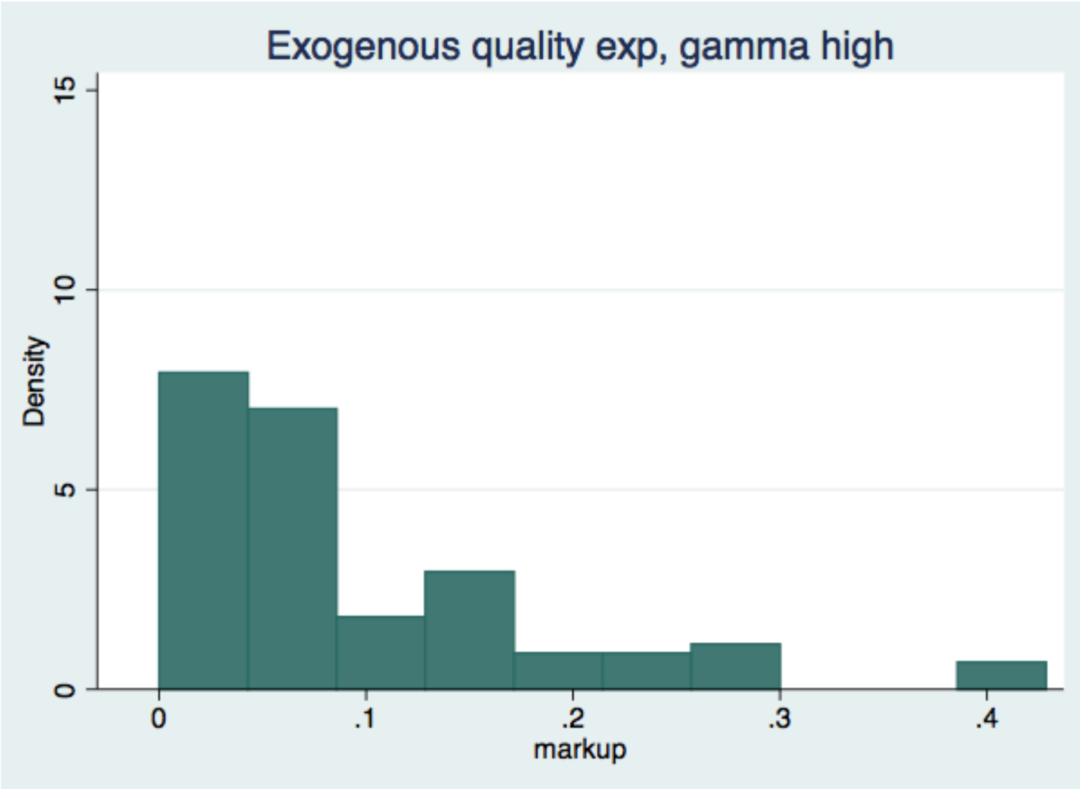



Endogenous quality experiment
Density
0
5
10
15
.2
.4
.6
.8
Production costs

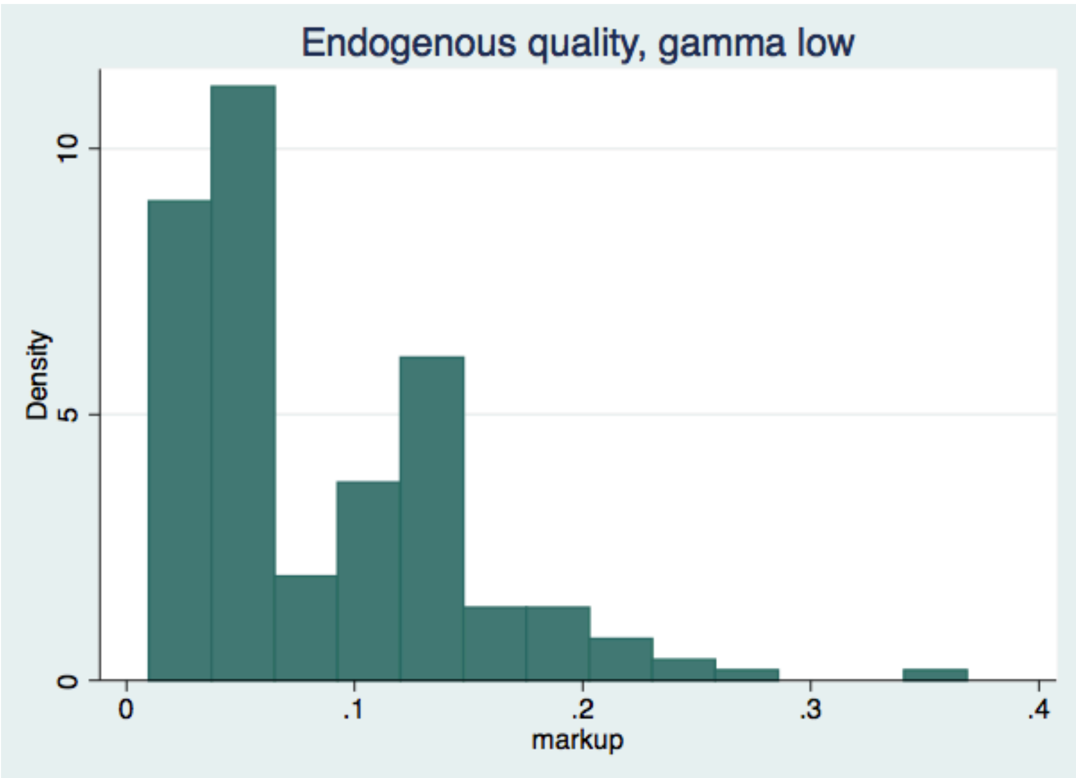

Endogenous quality, gamma low
Density
0
2
4
6
8
.2
.4
.6
.8
Production costs


Endogenous quality, gamma high
Density
0
5
10
15
20
25
.2
.4
.6
.8
Production costs

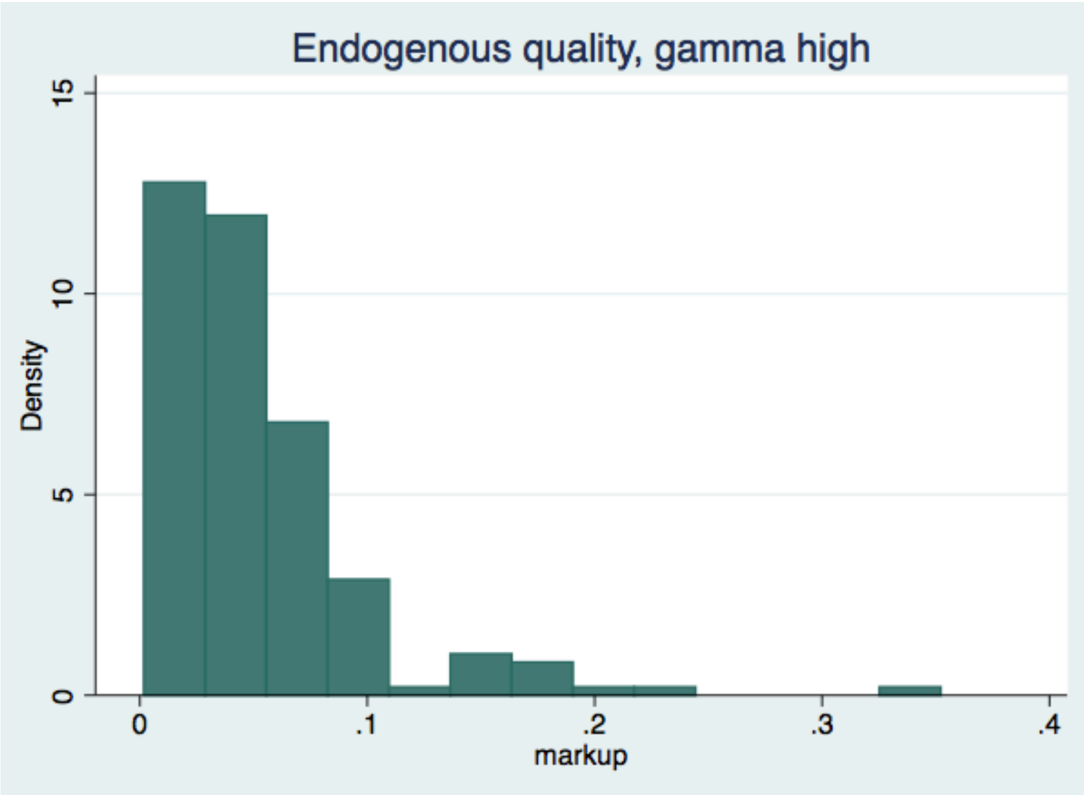

Exogenous quality exp, gamma low
Density
0
5
10
15
0
.1
.2
.3
.4
markup



Exogenous quality exp, gamma high
Density
0
5
10
15
0
.1
.2
.3
.4
markup



Endogenous quality, gamma low
Density
0
5
10
0
.1
.2
.3
.4
markup


Endogenous quality, gamma high
Density
0
5
10
15
0
.1
.2
.3
.4
markup

Appendix B

Equilibrium in the highest-score auction with exogenous quality. Here we mimic the procedure adopted by Che (1993).

Let

$$s(\alpha) \equiv max\left\{\frac{1-\gamma}{\gamma}\cdot\alpha - p(\alpha)\right\} = \frac{1-\gamma}{\gamma}\cdot\alpha - c(\alpha),$$

and

$$\sigma(\alpha) = \frac{1-\gamma}{\gamma}\cdot\alpha - p(\alpha).$$

Consider now the following change of variable

$$v \equiv s(\alpha); H(v) \equiv F\left(s^{-1}(v)\right).$$

The maximization problem of type-α seller then writes

$$max_b\{(v-b)[H(b^{-1}(b))]^{N-1}\}.$$

Applying the standard results in first-price auction yields

$$b^*(v) = v - \int_0^v \left[\frac{H(t)}{H(v)}\right]^{N-1} dt,$$

which is equivalent to

$$p^*(\alpha) = c(\alpha) + \int_0^\alpha \left[\frac{F(x)}{F(\alpha)}\right]^{N-1} \left[\frac{1-\gamma}{\gamma} - c'(x)\right] dx.$$

Notice that we can mimick Che's procedure only when $\gamma = 1/3$, as in this case $s(\alpha)$ is monotonic in its argument and thus invertible.

SUMMARY

Abstract

This paper reports evidences of two first score experimental procurements with scoring auction awarding criterion; one in which subjects need to endogenously determine both the price and the technical attributes of their tender, and the other in which they are only free to bid on the price dimension, while the quality is them exogenously assigned before bidding. The experimental design of both the games consists of two different treatments of interest: one in which quality matters more for the selection of the winning subject and the other in which the price bid instead is the determinant factor. Margins extracted over production costs are here analyzed and compared, from a behavioral perspective, across experiments and against equilibrium margins provided by the theoretical models. Main findings are the following. Higher margins were obtained in exogenously quality environments but they here even resulted more heavily influenced by tender's characteristics choosed by the buyer. This directly depends on the selection each mechanism is able to carry out. When compared against optimal margins both the experiments provide the same results. Subjects almost always obtained lower margins when competition is on the quality dimension, while always higher when competition transfers on the price dimension.

1. Introduction

A huge amount of resources is allocated every year by the mean of public procurement.

The latter, which is commonly defined as the process of purchasing by Governments, state-owned entrerprises or local entitites, of goods, services and work, accounts for a large amount of the public expenditures, ranging from 18-20% in countries like Portugal and Greece, to 35-40% as in Japan or Korea. An increasing number of sectors, such as waste management, construction, energy, telecommunications, transport, social and sanitary protection, and the provision of financial and education services, are indeed characterized by having public authorities acting as the standard buyers. Whilst OECD countries register almost 10-16% of their GDP spent in public procurement, the so called developing countries manage to do even better, reaching the threshold of 20-22%.Thus, it not only means that government procurement produces a substantial effect on the aggregated demand of any domestic economy, but also that it accounts for a sizable portion of the taxpayers' money. For this reason it's a process which needs particular attention by policymakers and researchers, who are continuosly asked to look for optimal and efficient mechanisms to be able to safeguard the public interest and maximize the social welfare. Although for years the literature of auction theory described a procurement contract as a game of incomplete information *based on price competition only*, leaving every tender to comply with some minimal quality requisites, technical specifications of the goods or services provided are becoming more and more important for the final decision of the buyers. In the contract for national defense acquisitions or for the allocation of public works for instance, the price to be paid is just one of the multiple attributes to be considered. In other words, we are progressively passing from the better known "*Lowest Price*" evaluation criterion (LP), according to which only the price bid makes a difference in the selection of the winning tender, to the "*Most Economically Advantageous Tender*" evaluation criterion (MEAT), which asks the buyer to assess and rank the presented bids in relation to multiple attributes, such as technical characteristics, delivery terms, after-sale services, sustainable aspects, innovative and environmental characteristics, and so on. Towards this aim, the most commonly used procedures in real practices include: (i) menù auctions, (ii) scoring auctions, (iii) beauty constests and (iv) bargaining, according to which the buyer selects some potential sellers and negotiate with them on all possible dimensions to be evaluated. The technique we are focusing on in this paper is that of a scoring auction, which according to Milgrom (2004) is particularly interesting since, under some circumstances, may let procurers to obtain more valuable contracts at the same time promoting partecipation by more bidders and not lowering sellers' expected profits, with so increasing value for both parties. In his opinion, bidders must always prefer scoring auctions procedure rather than price-only auctions, since in thie latter case they do not manage to exploit their peculiar advantages and characteristics. The robustness of these findings was confirmed by Asker and Cantillon(2008), who showed how scoring auctions let both buyers and sellers to achieve higher expected payoffs. Other several theoretical works focused recently their attention on multidimensional scoring auctions. The latter, which is commonly considered a two-stage evaluation process consisting of a *technical evaluation* - the technical merits of the tenders - and of a *financial evaluation* - which relates to the price only - let the contracting agency to award the contract to the tender that reached the highest overall score, which is just defined as a weighted average of both technical and fincancial criteria, with weights well defined before bidding.

Albeit Thiel (1988), Cripps and Ireland (1994) firstly approached the problem of multidimensionality by assuming respectively (i) that procurers do not value any savings and choose on the basis of preset budget – publicly communicated to all bidders – and (ii) quality thresholds to be passed to enter the auction, the first work which really came close to find a solution for the design of multidimensional scoring auctions was that one realized by

Che(1993). The author, who considered a two-dimensional model in which every bid consists of a pair of price and quality assessed according to a scoring rule converting each pair into a single number, discovered how first and second score auctions are able to implement an optimal and efficient mechanism when the scoring function is quasi-linear in price. In his simple model, each bidder differs from the others just for his marginal costs of improving quality, which, following Harsanyi(1967), is randomly drawn by the "Nature" from a well-defined ex-ante probability distribution, which is common knowledge among all bidders. In other words, each firm is assumed to bear a cost which is independent from all the others. The latter is one of the most criticised point of Che's findings, since in real procurement practices it's reasonable to assume that the costs of the partecipating firms must have some in common and so cannot be completely independent from each other. Branco(1997), for example, tried to extend Che's model by analyzing the impact of costs correlation on the design of multidimensional mechanisms. His findings were surpising, since none of the properties regarding the mechanisms studied by Che resulted to be valid in his setting. In particular, when costs are correlated, any mechanism based on a single stage auction - and so even that one discovered by Che – is not optimal anymore. Only two-stage mechanisms where in the first stage bids are evaluated according to a scoring rule and in the second the winner of the first round bargains with the contracting authority, seem to be able to reach optimality.

On the other hand, other works contested the mono dimensionality of each bidder's type of Che's model, since in reality firms can differ not only in the marginal costs, but even in their fixed costs. Asker and Cantillon (2010), on the basis of several previous works on the same topic, extended the analysis of optimal procurement mechanisms to the more elaborate environment in which private information is multidimensional, although independent among bidders. The latter assumption, togheter with a scoring function quasi-linear in price, results here to be crucial to reduce the dimensionality of the relevant private information to one – i.e. each bidder's "pseudotype"- which is necessary to characterize the equilibrium of this particular scoring auction. In addition, the authors managed to find, as already done by Che in his simplified environment, an extension of the famous revenue equivalence theorem, showing that buyers are indifferent between first score, second score and ascending or descending scoring auctions when suppliers are symmetric in their pseudotypes. Although it's really hard to characterize the optimal mechanism when private information is multidimensional, the authors managed to show that scoring auctions yield the same performance of the optimal one, obtained by numerical simulation. Nishimura (2012) extended the results of Che (1993) to an environment in which private information is mono-dimensional but the quality is here represented by a vector of multiple attributes and found that an optimal mechanism is achievable even in this case, as well as the buyer describe scoring rules which must be additively separable in some or all quality attributes. All of these works relate to procurement contract where quasi-linear scoring rules - as for instance the weighted criteria - have been adopted. Hanazono et al (2011) tried to remove this strong assumption by assuming "price-quality ratio" scoring rules (PQR) and M-dimensional private information for each supplier. According to the authors, the intense competition provided by this setting induces undercutting the provision of technical attributes, which turns out to reduce the price asked in equilibrium. None of the properties regarding optimality and revenue equivalence resulted here to be valid. Unfortunately, although interesting and well structured, none of the above mentioned theoretical works manage to capture all the difficulties encountered by practitioners in real procurement practices and to provide them useful and easily implementable paths to follow. This is mainly addressed to the impossibility for partecipating firms to understand the excessive sophisticated optimal strategies discovered by theorists and to the procurers' difficulties in having full knowledge of the environment they deal with. As a consequence, laboratory experiments are progressively becoming always more useful tools to interpret the effects that a mechanism or another could have. Bichler (2000) and Chen-Ritzo et al (2005), for instance, explored experimentally a multiattribute reverse auction in which price and two technical attributes of the goods to be sold have been considered. Despite the fact that the laboratory setting still remains a simpler and less sophisticated environment than real world, and even if lots of doubts regarding the generalization of laboratory results are recurring among practitioners, could sometimes reveal very useful to study in the lab a new procedure before introducing it in the field. The aim of this paper follows exactly this intuition and consists of providing empirical evidence regarding the bidding behaviour of subjects involved in a public procurement tender with scoring auctions procedure. Specifically, we compared two different experimental works - whereof our investigation is an extention –, one in which suppliers, endowed with iid quality before offering, bid only in the price dimension, and the other in which instead they need to endogenously determine both the quality and the price to be assigned to the good or service they want to sell throught the procurement auction. In particular, we are interested in understanding, form a behavioral perspective, how the subjects' behavior changes when they are asked to deal with more tasks with respect to the case in which they have to provide a bid in the price dimension only. Margins obtained by winning subjects are here compared.

2. Background

The framework upon which my research is based, mainly consists of two different experimental papers on scoring auctions in public procurement – that one of Albano, Ponti et al (2015) and the Master degree Thesis of M. Rosi (2015), a LUISS Guido Carli student – whose data I exploited to better understand the main characteristics of the empirical behavior of subjects involved across the two different settings. For this reason, a brief description of their structures, – illustrated in detail in *Section 3* – their research questions and their main findings is reported in this section.

3. Experimental design

The experimental structure of both the two different projects is in this section analyzed. Both the two experimental games were fully computerized and programmed using the well-know software Z-tree, and consist of four different sessions. Whilst the game in which the quality has been made exogenous was carried out at the Centro di Economia Sperimentale a Roma Est (CESARE), at Luiss University in Rome, the game in which the quality bid has to be determined by partecipating subjects was conducted at the University of Alicante, Spain. Subjects – 80 in one project and 90 in the other - were randomly divided in different groups, each of which represents an isolated "world" from the remainings. Thus, different groups never interact with each other towards the whole session. Whilst the exogenous quality procurement game aims at simulating first-score procurement auctions with adverse selection only, the other game results to be more sophisticated, since, as in real practices happens, both adverse selection and moral hazard are here contemplated. After bids have been submitted, the game outcome is determined by a scoring function representing the true preferences of the mechanism designer. Depending on the sessions considered, more weight could be given to the price offered or to the technical merits of the tender. Denoting as n^* the number of subjects who managed to obtain the higher score, each player's payoff, writes :

$$\pi_i(\alpha, p, \gamma) = \begin{cases} \dfrac{p - c(\alpha_i)}{n^*} & if\ s_i(.) = max_j(s_j(.)) \\ 0\ otherwise. & \end{cases}$$

In other words, if player i wins the auction, he will be rewarded with a payoff corresponding to its profit obtained through the trade, while if he does not he will recieve a payoff of zero. The payoff above is true for exogenous quality experiment, while, albeit defined in a similar way, it depends on an additional parameter when quality is endogenous. Although constructed almost in the same way, the costs structure in fact differs significantly across the experiments, depending on α only in one case while on both α and θ in the other.

4. Experimental results

In this Section the main conclusions of our experimental work are reported. In particular, before comparing the two different projects from the partecipants' point of view, it is worth exploring wheter the theoretical predictions regarding the quality bid characterized by Che(1993) – obviously only in the case in which it's endogenously determined by the players - are compatible with the empirical counterparts. The author, as already mentioned, showed how the optimal strategy in the quality dimension results to be completely unrelated to that of the score, since just obtained by maximizing the consumers' surplus with respect to the quality and imposing a price which exactly covers the cost incurred. If empirical quality bids happen to follow this intuition, it means that the outcomes of the two experiments, although suffering the relevant discrepancies in terms of their structures, should be at least consistent.

4.1 Empirical quality choices and deviations from optimality

With the help of the following graphs we put in comparison empirical quality choices and their equilibrium counterparts, obviously dealing only with the experimental game charcterized by an endogenous quality provision. Each figure plots optimal and observed quality bid against the cost paramenter affecting the marginal cost of increasing quality – i.e each player's "type", with which suppliers are endowed at the beginning of each round before bidding.

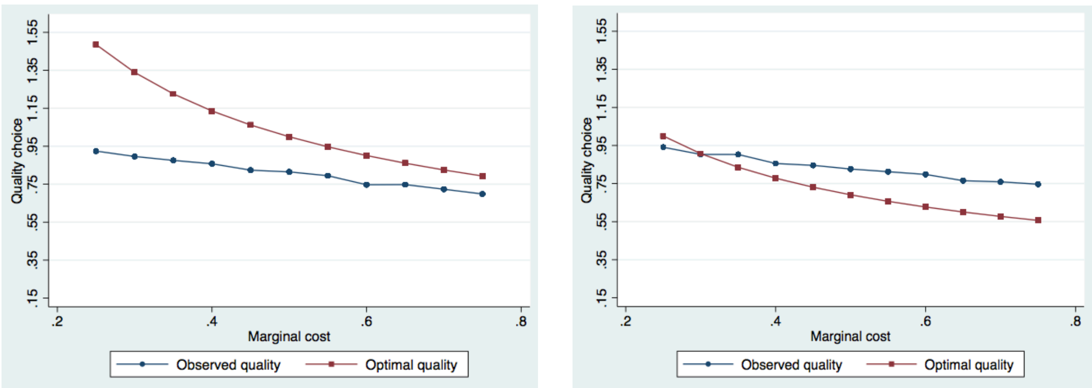


Figure 1: *Optimal vs Observed quality choice. Gamma low.*

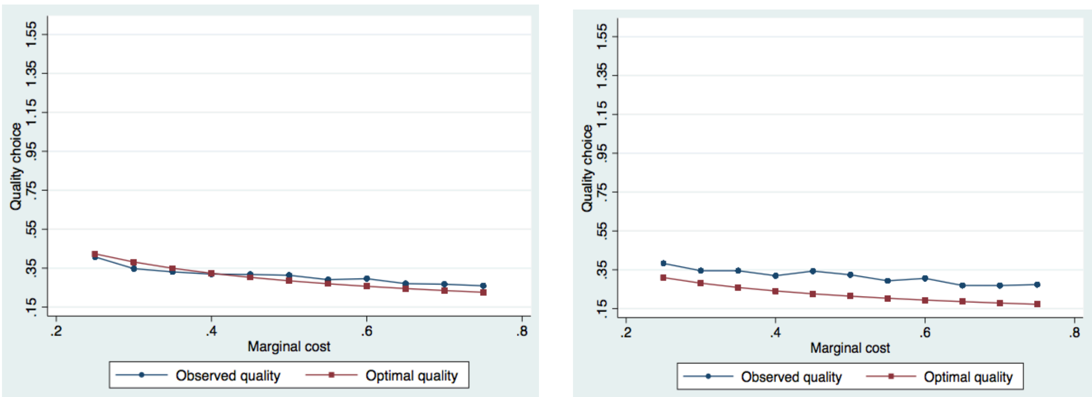


Figure 2: *Optimal vs Observed quality choice. Gamma high.*

Figure 1 reports average quality bid of suppliers with respect to that ones pronosticated by the Che's model, when the final overall score is more influenced by the technical one (i.e. $\gamma = \frac{1}{4}$). The left panel differs from the right one just in the specification of the technical score, which exhibits higher concavity in the latter panel. *Figure 2* instead, while representing the same comparison of the others, deal with a scoring function in which the financial score, which in the experiment is linear in price, has been assigned an higher weight (i.e. $\gamma = \frac{3}{4}$). In this case too, the right panel is characterized by having a more concave technical score than the left one. As we can see from the graphs, albeit subjects under or overbid most of the time with respect to their optimal strategies, they manage to follow the theoretical predictions in all the four settings here considered. More in the specific, the quality offered is always decreasing in suppliers' type – i.e. their marginal cost of increasing a given level of quality. This is in accordance with the theory and basic intuitions, since it's quite obvious that less efficient suppliers – i.e. firms whit higher marginal costs – are less able to offer very sophisticated goods or services unless by asking a very huge price to cover their costs. Even the distance between the two plots seems not to be large, expecially when price matters more than quality in the final overall score. Moreover, it becomes immediately evident the huge downward shift in the quality bid - under both two specifications of the technical score contempleted - when competition transfers on price, i.e when the financial score has an higher weight in the final overall outcome. In the latter case in fact, an higher level of quality not only does not increase the probability of winning the tender, but can even lower it by forcing suppliers to increase the price asked, so to avoid to make very low profits.

4.2 How does markup relate with each player's type?

In order to enable the comparison between the two experiments, the markup is here defined as the difference between the selling price and the cost price, computed as a percentage of the latter. Recall in fact that the cost space is the only one subjects have in common across the two procurement games. Although the experiment in which the quality is endogenously treated exhibit marginal costs always lower than that provided by the other experiment, – mainly because of the presence of the cost parameter, θ, which is smaller than one – the cost functions were indeed constructed exactly in the same way. Therefore, for the purpose of our investigation, it is interesting to understand which results to be the experiment in which subjects extracted, for each value of the costs borne, higher margins. Particularly appealing is the case of winners only. In this situation, because of the different cost structures could happen that winners are characterized by different costs according to which experimental game is considered. As already mentioned in *Section 3*, costs depend on the quality only, in the case in which the latter is exogenously treated, while both on quality and theta in the other case. Thus, before comparing the experiments, could be here useful to consider them separately and understand, for the case of winners only, how the markup relates to the quality level and the cost parameter θ, so to be able to predict which could be the winners' overall production costs in both the games. Observing that, we will be able to capture the behavior of the winners across the two different settings, controlling for what could be just explained by the differences in terms of structure.

4.2.1 Experimental game with endogenous quality provision

The following figure plots average margins against θ. Observed values are here analyzed for every specification of gamma - the weight of the financial score in the final overall score – and technical functions. For the sake of simplicity and to avoid that the notation used can be misleading, we defined T1 as the situation in which $\gamma = \frac{1}{4}$ and $n = 4$; T2 when $\gamma = \frac{1}{4}$ and $n = 8$; T3 when $\gamma = \frac{3}{4}$ and $n = 4$ and T4 when $\gamma = \frac{3}{4}$ and $n = 8$.

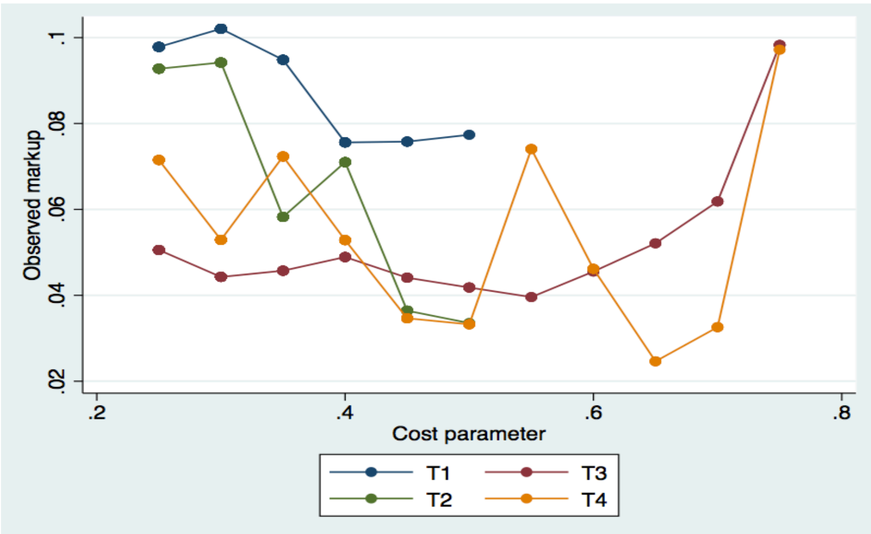


Figure 6: Average margins against cost parameter (θ). Sessions comparison.

As the figure shows, when quality matters more than price for the final decision of the buyer (T1, T2), the winners' average markup is always decreasing in the cost parameter. When θ is higher costs become higher as well, with so reducing the markup for every level of quality choice. The trend is the one expected, more efficient firms extracting higher margins, decreasing as the cost parameter gets bigger. Moreover, there is a threshold beyond which none of the partecipating subjects manage to ask a price higher than the cost incurred at the same time winning the auction. When the cost parameter is higher than 0.5 in fact, the graphs are interrupted since no firms manage to win. As soon as competition shifts on prices instead, we can immediatelly notice the downward shift of the subjects' average markup with respect to the previous case. In this setting, firms are forced to lower their price bid in order to win the auction, which immediately translates in lower margins. The trend, although still decreasing, is not monotonic anymore, meaning that less efficient subjects managed here to win more often with respect to before. In T3 or T4 for instance, even subjects with high cost parameters were able to extract huge margins when winning the auction. In some sense, when the weight assigned to the financial score in the overall score is higher than that of the technical score, all subjects, even the less efficient, have a chance of victory. Only efficient firms instead results to be able to stay in the market in the first above mentioned setting. Consequently, we expect that winning firms will exhibit lower costs when technical score has an higher weight than when fincial score is more relevant for the final decision of the buyer.

4.2.2 Experimental game with exogenous quality provision

The figure below plots the average markup this time against the quality level, which is here exogenously treated and randomly assigned to subjects before bidding. We used the quality since it's the only variable affecting here the cost of partecipating firms. As we did for the previous experiment, we denoted as T1 the environment in which the final score is more affected by the choice of quality and T2 that one in which the financial score has been assigned an higher weight.

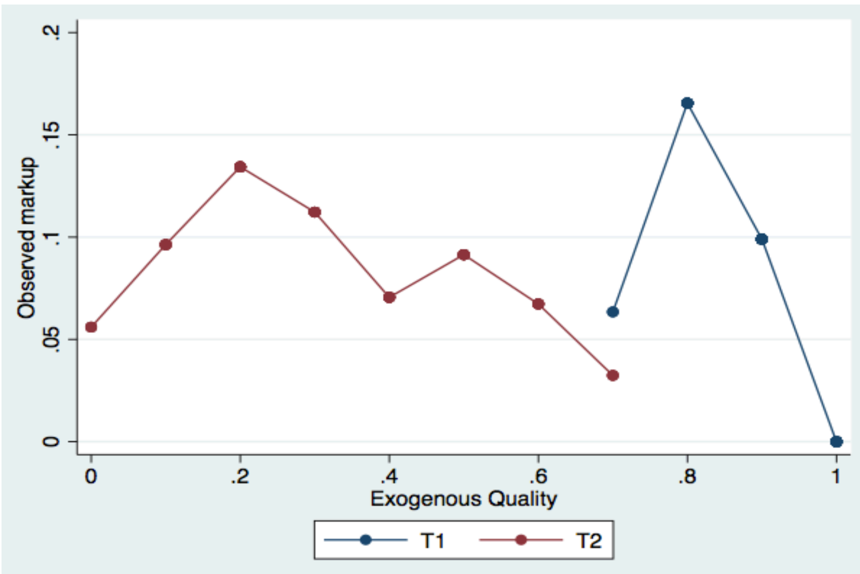


Figure 7: Average margins against cost parameter (γ). Sessions comparison.

The results provided by the figure above are interesting and consistent with basic intuitions. In particular, trends are always decreasing in quality - since here an high quality level immediately translets into higher costs, which negatively enter our markup function - except for subjects endowed with very low quality type in T2. In the latter case in fact, subjects with type smaller than 0.2 exhibit an inverted trend in the margin extracted, meaning that even if endowed with higher costs, they were able to ask higher prices as well still winning the auction. Moreover, average margins significantly differ according to which criterion has been given an higher weight in the final scoring rule. Specifically, both the graphs are interrupted after or from a given threshold of quality. In the first treatment (T1), none of the subjects with type smaller than 0.7 managed to win the auction and so to obtain a positive payoff. This has a clear explanation. When quality matters more than price in buyer's preferences – here summarized in the overall scoring rule adopted – we expect that only subjects endowed with very high technical attributes manage to win the auction, exploiting their peculiar advantages. When price matters more than quality instead (T2), we can easily notice how subjects with type higher than 0.7 are completely out of the market. In this case in fact, even if more subjects resulted having the contract awarded, the harsh price competition arised forces partecipants to lower their price to win the auction. As a result, subjects enriched with quality higher than the above threshold are obliged to offer prices higher than their competitors to cover costs they have to bear, but this prevents them to win the tender. Different from the experiment in which the quality has been endogenized, winning subjects managed here to obtain almost the same margin either when financial or techincal score has an higher weight in the overall outcome. This obviously depends on the higher costs suffered by subjects in T1 with respect to that of the other treatment, which directly reduse the markup for every characterization of the price bid.

4.3 Regression analysis

In this paragraph, margins obtained are analyzed more in detail. The effects of some independent variables on the markup extracted from the trade are deepened. Variables considered include: (i) tender characteristics, even interacted with other variables; (ii) production costs; (iii) some exogenous parameters; (iv) individual characteristics, obtained by the answers of questionnaire provided at the end of each experiment. To reach this aim, we use three different econometric approaches which give us useful information about the robustness of results. Recall our dependent variable – the markup – is defined for the winners only, thus we cannot exploit the advantages of the Panel analysis simply since one subject may have won in one round and nevermore, which makes the Panel strongly unbalanced. We start the analysis by running the simple **OLS regression**, which is just our standard benchmark. However, some problems could arise when using OLS. The sample of observed margins is not random anymore, but it has been selected according to a specific rule – i.e. for each period, only subjects who reached the highest overal score of their group return with positive margins. In other words we recognize that some coefficients of OLS regression could suffer a distortion because of the selection of the sample. As a consequence we improve the previous test by running for the same variables the **Heckam's two step regression** in which the selection equation expresses the ex-post probability of winning as explained by (i) the probability of winning the auction at equilibrium; (ii) some individual characteristics, which could influence the behavior of partecipants; (iii) tender's characteristics and their interactions with the others. In this way, we can analyze the sample of winners observed controlling for the ex-post probability that a each subject had to win. Results confirm what OLS already showed but each coefficient has been adjusted. The only thing we can do at this point to keep on improving our results is working on the selection equation of the Heckman regression. The latter in fact, estimates the effects of the independent variables on the probability of winning the auction, but treat each observation as an independent one, without understanding that each subject played for more than one rounds. So, we use **Two-stage least square regression** where at the first stage we control for the sample selection by regressing the ex-post probability of winning on the same variables used in Heckman, but considering the observations as panel data, so taking the adavantages that these models garantee. The markup is year expressed as dependent on the same variables used in OLS and Heckman, and the prediction of the probability of winning obtained at the first stage. After a first analysis in which margins were analyzed for each experiment separately, we studied them togheter to see which is the setting in which subjects managed to obtain higher markup and how the latter are influenced by the tender's characteristics across experiments.

4.3.1 Margins' comparison

After having understood which are the main drivers of the margin obtained by subjects in the two different games, it's interesting now to see which experimental setting let to extract higher markup and how tender's caracteristics are percieved across the experiments. For this purpose, we run a two-stage regression in which our dependent variable is the aggregate margin for both the games. For the first-stage regression we used predicted values obtained in the previous paragraph, when we dealt with each experiment separately. We then generated a dummy variable which discriminates among experiments and which was even interacted with all tender's characteristics the two procurement games have in common. Before discussing results, we report a short description of the variables we never met so far and the equation of the model used.

(i) *D_exp*, is the dummy taking value 1 if the endogenous quality experiment is considered, 0 otherwise

(ii) *D_exp*cost*, is the interaction between D_exp and costs. It gives us the effect of a raise in cost only for the endogenous quality game.

(iii) *D_exp*gamma,* is the interaction between gamma and the project. It's equal to 1 if gamma is high and we are in the endogenous quality environment.

(iv) *D_exp*cost*gamma,* is a dummy which discriminates not only among experiments, but even among gamma (high/low).

(v) *Individual_control,* is just a set of individual characteristics we used as control. They exactlu correspond to those used in previous analysis.

(vi) *Gamma,* is a dummy variable taking value 1 if price have more weight in the buyer's utility function and 0 otherwise

$$
\begin{aligned}
Markup = \beta_0 + \beta_1 D\_exp + \beta_2 cost + \beta_3 D\_exp\_cost + \beta_4 gamma + \beta_5 D\_exp\_gamma \\
+ \beta_6 cost\_gamma + \beta_7 D\_exp\_cost\_gamma + \beta_8 impulsive + \beta_9 impulsive\_gamma \\
+ \beta_{10} winner\_hat + \varphi\ individual\_control
\end{aligned}
$$

When subjects are asked to bid on both the price and quality dimension margins obtained from the trade are substantially lower than when they have to formulate an offer on the price dimension only. The coefficient of *D_exp* is indeed negative and significant at 1% level of confidence, as well as that of *Gamma.* As already noticed in the previous subsection, when gamma is high, subjects return on average with lower margins because of the strong price competition. What it's interesting here is that this effect is significantly mitigated in the endogenous quality experiment with respect to the other. When adding the effect of *D_exp*gamma* in fact, margins continue to be lower in high gamma settings but the difference tends to be substantially smaller than that observed in the other experiment. Another interesting result comes from the different marginal effects that a raise in production costs have on the margin extracted. The following table summarizes these effects discriminating across experiments and low/ high gamma settings.

Costs' marginal effect	**Gamma low**	**Gamma high**
Endogenous quality experiment	$\beta_2+\beta_3$ (- 0.2161)	$\beta_2+\beta_3+\beta_6+\beta_7$ (- 0.1516)
Exogenous quality experiment	β_2 (- 0.5949)	$\beta_2+\beta_6$ (- 0.2620)

Table1 : Marginal effects of a rise in production costs on the markup extracted.

Table1 reports the importance of production costs in the determination of margins obtained for each specification of the parameter γ. As can easily be noticed, although markup are always negatively related with costs, their impact considerably changes according to which experimental game has been played. Starting from the game in which the quality is exogenously treated, the marginal effect of a raise in production costs is β_2 (- 0.5949) in low gamma settings, while β_2 +β_6 (- 0.2620) when gamma is high. Thinking about the latter game's structure and the selection rule of the winning firm previously discussed, this result is normal and consistent with our expectations. When gamma is low, only subjects endowed with very high technical attributes – and so high costs as well - manage to win the auction. This obviously lower the action space of each subject and makes them more dependent on costs they are suffering.[1] What is interesting here is that this relation between gamma, production costs and markup is not confirmed in the experiment in which the qualidy bid is endogenously determined by partecipants. In the latter situation, the marginal effect of a rise in production costs is equal to β_2+β_3 (- 0.2161) when gamma is low, while β_2+β_3+β_6+β_7 (- 0.1516) when gamma is high, and these two effects result to be not significantly different from each other. In other words when subjects decide both the price and the technical attributes of the good or service they want to sell within the auction, they seem not to be particularly constrained by buyer's preferences. Albeit on average subjects obtained lower margins in high gamma setting, either when gamma is low or when it's high the marginal effect that a raise in costs have on the margin extracted is almost the same. This result obviously depends on the different selection of the winning firm across experiments, which works better in the exogenous quality environment rather than in the other, mainly because of the more degree of freedom of partecipants.

4.4 Deviations from optimality

In this Section deviations from equilibrium strategies are analyzed. Equilibrium margins are here computed using optimal price bid strategies and, for the experiment in which subjects bid even on the quality dimension, optimal costs. The latters, for every player's type, depend on the optimal quality bid only. For the other experiment instead, the players' type is the production costs – they depend on quality only, which here is exogenously assigned to subjects – and the optimal price bid is just computed according to the latter formula:

$$optimal\ price\ bid = 1 - optimal\ rebate\ bid$$

For the sake of transparency, as common practice in this paper, we prefer to analyze each procurement game separately.

[1] Recall in fact that subjects cannot formulate a price bid lower than costs they have to bear.

When analyzed against production costs, deviations from equilibrium margins strongly depend on which setting of the game has been played. In particular, albeit in both the experiments observed margins follow the optimal ones at least in the shape, observed margins where almost always lower than that pronosticated by the models when γ is low, while are sistematically higher when competition transfers on the price dimension. In the latter case, althoug subjects' margins are strictly lower than when γ is low – due to the harsh competition on the price dimension – the equilibrium ones result to be even lower. This result is valid both for exogeneous and endogenous quality environment. One possible reason of this behavior could be addressed to one common cognitive bias. When γ is high, subjects percieve that even if they win the auction they probably return with a payoff close to zero, expecially if they suffer high production costs. Therefore, they prefer to bid higher prices than expected, which, albeit lowering their probability of winning, can garantee them acceptable payoffs in case of victory.

5. Conclusions

This article has studied in detail how subjects behave in two different experimental procurements with scoring auctions awarding criterion. One in which subjects formulate both price and quality bids and the other in which they are free only to bid on the price dimension. Margins extracted from the trade by winning subjects were analyzed. We first studied them separately and then compared them with each other. The main question of this paper was indeed that one of understanding, from a behavioral perspective, which was the setting in which partecipants were able to obtain higher markup. The first observation to me made regards the selection rule. The latter was more pronounced in the exogenous quality experiment rather than in the other, simply because of the more degrees of freedom each subject has in choosing the production costs he wants to bear in the latter case. When technical merits of the tender have more weight in the buyer's utility function – i.e. the scoring rule – only subjects with very high quality type managed to win the auction in the exogenous quality environment. On the other hand, a significant difference across margins in the two experiments was discovered. When subjects leave out every consideration about the quality choice, they manage to perform better, resulting with higher margins. What is interesting is how these margins are influenced by the tender's characteristics the two experiments have in common. In the experiment in which the quality is endogenously determined, the mechanism designer seems to manage to influence less subjects' behaviour only by expressing his preferences in the overall scoring function. Albeit in both the experiments they return with lower markup in high competitive settings with respect to the others, these difference tends indeed to be considerably lower when quality is endogenously treated. The same happens when analyzing the impact of production costs.

Whereas markup are heavily influenced by production costs in the exogenous quality environment – in the sense that an increase in costs induces a huge negative fall of the margin extracted -, the same relation is substantially mitigated when subjects can decide both the price and the quality of their tender. Moreover, the difference across experiment is emphasized when the buyer cares more about the technical specifications of the contract rather than its price. We could say that the higher action space subjects have in the endogenous quality setting makes them to be less affected by buyer's needs and so to be able to reach an acceptable margin regardless the costs they are suffering.

When compared against markup pronosticated by the theoretical models behind the experiments, a common evidence is discovered. While subjects return most of the times with margins lower than the optimal ones when competition is on the quality dimension, they sistematically do better than that when competition shifts on prices. In this case, monetary payoffs were too low in equilibrium and this obviously prevent optimal bids to be an interesting alternative. Subjects indeed preferred in both the experiments to reduce their probability of winning the auction in exchange for accepatable rewards in case of victory.

The investigation carried out in this paper was from the point of view of the partecipating subjects only. It would have been really interesting to understand how buyer's utily function and his costs savings obtained with the auction scheme could have been influenced by the exogeneity or endogeneity of the quality bid. Unfortunately, the structures of the two procurement games here considered was so different so to prevent us every possible considerations about other variables, which result to be not directly comparable.